Mega-Enhancer Bodies Organize Neuronal Long Genes in the Cerebellum

Ziyu Zhao[1,2,#], Omar Payán Parra[1,2,#], Francesco Muscella[3,4], Nicolas Scrutton-Alvarado[1,2], Shin-ichiro Fujita[1], Frank Alber[3,4], Yue Yang[1] , Tomoko Yamada[1,*]

[1] Department of Neurobiology, Northwestern University, Evanston, IL, USA

[2] Program in Interdisciplinary Biological Sciences, Northwestern University, Evanston, IL, USA

[3] Institute for Quantitative and Computational Biosciences and [4] Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA, USA

[#] These authors contributed equally

[*]To whom correspondence should be addressed. Email: tomoko.yamada@northwestern.edu

Abstract

Dynamic regulation of gene expression plays a key role in establishing the diverse neuronal cell types in the brain. Recent findings in genome biology suggest that three-dimensional (3D) genome organization has important, but mechanistically poorly understood functions in gene transcription. Beyond local genomic interactions between promoters and enhancers, we find that cerebellar granule neurons undergoing differentiation *in vivo* exhibit striking increases in long-distance genomic interactions between transcriptionally active genomic loci, which are separated by tens of megabases within a chromosome or located on different chromosomes. Among these interactions, we identify a nuclear subcompartment enriched for near-megabase long enhancers and their associated neuronal long genes encoding synaptic or signaling proteins. Neuronal long genes are differentially recruited to this enhancer-dense subcompartment to help shape the transcriptional identities of granule neuron subtypes in the cerebellum. SPRITE analyses of higher-order genomic interactions, together with IGM-based 3D genome modeling and imaging approaches, reveal that the enhancer-dense subcompartment forms prominent nuclear structures, which we term mega-enhancer bodies. These novel nuclear bodies reside in the nuclear periphery, away from other transcriptionally active structures, including nuclear speckles located in the nuclear interior. Together, our findings define additional layers of higher-order 3D genome organization closely linked to neuronal maturation and identity in the brain.

Introduction

Transcriptional programs drive neuronal specification and the integration of neurons into specialized neural circuits[1-3]. Transcription factors and chromatin enzymes control these gene expression programs by recruiting transcription machinery to gene promoters and enhancers[4-6]. Beyond local regulatory mechanisms, recent advances in chromatin biology have revealed important roles for the three-dimensional organization of the genome in coordinating developmental programs of gene expression, which are specific to cell types[7-10].

Among principles of genome organization, interactions between cis-regulatory enhancers and gene promoters have well characterized roles in facilitating transcription[11]. These interactions are mostly confined within insulated domains called topologically associated domains (TADs) that span up to 2-3Mb[12,13]. At genomic scales beyond TADs, the genome may be partitioned into a transcriptionally active A-compartment and a repressive B-compartment in eukaryotic cells. The A-compartment may partly reflect the recruitment of active genes to nuclear speckles, which are nuclear bodies enriched with RNA splicing machinery[14]. Nuclear speckles are often found in the nuclear interior, localized away from heterochromatin regions anchored to the nuclear lamina[15]. Besides nuclear speckles and lamina-associated heterochromatin, recent studies have revealed additional transcriptionally active and repressive nuclear compartments, which may be unique to brain cells[10,15-20]. However, our understanding of nuclear structures in the brain, from their regulation to their biological roles, remains limited and extends to only a few cell types.

Neurons in the mammalian brain express some of the longest genes encoded in the genome, spanning up to several megabases of DNA[21]. These neuronal long genes encode synaptic and transmembrane proteins, which help establish specialized brain circuit architectures that support cognitive and sensorimotor functions in animals[21,22]. Moreover, the selective dysregulation of long

gene expression is often observed in human neurodevelopmental disorders[22]. With the evolution of long genes in vertebrates, there is also a genome-wide expansion of *cis*-regulatory enhancer elements that are often found along gene bodies and are functionally active in neurons[23,24]. Although these regulatory enhancers may form extensive interactions within TADs[13,19,25], how these enhancers and their associated long genes are organized over genomic scales beyond TADs was previously unknown.

In this study, we investigated the higher-order genome organization of granule neurons in the cerebellum during development and adulthood. Interestingly, we found massive increases in genomic interactions between transcriptionally active loci, separated by tens to hundreds of megabases or across different chromosomes, during granule neuron differentiation. A group of these long-distance interactions gave rise to a nuclear subcompartment that was enriched for near-megabase-long regulatory enhancers and their associated neuronal long genes, which encoded for synaptic organizers and signaling proteins. Neuronal long genes may be differentially recruited to this enhancer-dense subcompartment to control transcriptional programs that shape the identity of cerebellar granule neurons. Using split-pool recognition of interactions by tag extension (SPRITE), data-driven genome modeling, and *in situ* imaging approaches, we also discovered that the enhancer-dense subcompartment occupied a distinct microenvironment near the nuclear periphery, well separated from other transcriptionally active subcompartments in the nuclear interior. Our findings collectively define a novel nuclear subcompartment that plays crucial roles in the 3D organization of long genes required for neuronal specification and integration into neural circuits.

Results

Massive increases in megaloops during cerebellar development

We first determined the three-dimensional genome organization of the cerebellum, a key brain region necessary for motor behavior and cognition[26,27], during development when cerebellar circuits are assembled. We employed *in situ* chromosome conformation capture with high-throughput sequencing (Hi-C) using the cerebellum of postnatal day 6 (P6) mouse pups, when the majority of the cerebellum comprises granule cell progenitors and immature postmitotic granule neurons, or the cerebellum of postnatal day 22 (P22) mice, when granule neurons have matured and integrated into neural circuits (Fig. 1a,b). Strikingly, Hi-C analyses revealed that long-distance intra-chromosomal interactions spanning 3Mb to 200Mb dramatically increased with cerebellar development (Fig. 1c). We also observed additional increases in long-distance interactions in Hi-C analyses of the cerebellum from postnatal day 56 (P56) adult mice (Fig. 1c).

We further characterized the genomic loci that formed short-distance and long-distance interactions in the cerebellum. During cerebellar development, we observed nearly equal numbers of gained and lost short-distance interactions, which were confined within loop domains marked by CTCF binding at boundaries[12] and spanned up to ~2.6Mb (Extended Data Fig. 1a-d). While short-distance interactions within loop domains were found at all stages of cerebellar development, long-distance interactions that spanned longer over 3Mb, and therefore beyond a loop domain, were nearly absent in early development and dramatically increased with brain maturation (Fig. 1d). An example was an inter-domain interaction that formed across nearly the entirety of chromosome 1 and was found in P22, but not in P6 cerebellum (Fig. 1e). To distinguish these long-distance interactions from short-distance interactions, which were confined within loop domains, we termed the interactions that extended over 3Mb as "megaloops".

Examination of megaloop anchors revealed that most were associated with the transcriptionally active A compartment (Fig. 1f). Megaloops straddled by A compartment anchors (A-A anchors) significantly increased during cerebellar maturation (Fig. 1f,g). Intriguingly, A-A anchors formed intra- and inter-chromosomal cliques, where each anchor interacted with multiple other anchors (Fig. 1h and Extended Data Fig. 1e,f). For example, A-A anchors on chromosome 1 and 12 displayed both long-distance intra-chromosomal and inter-chromosomal interactions (Fig. 1h). In other analyses, we observed a modest net decrease in interactions between a small set of megaloop anchors marked by trimethylated H3K27 (H3K27me3) in P22 compared to P6 cerebellum (Extended Data Fig. 1g). Together, these results suggest that the majority of megaloops gained during cerebellar development are formed between transcriptionally active genomic loci.

Distinct A subcompartments in the cerebellum

We next determined additional genomic features of transcriptionally active megaloops in the cerebellum by performing dimensionality reduction on the interaction matrix of A-A megaloop anchors (Extended Data Fig. 1h). Using principal component analysis (PCA), we found that the first principal component (PC1) correlated with the density of transcriptionally active genes and enhancers at anchors, which is the hallmark of the A compartment (Fig. 1f,i,j). Interestingly, the second principal component (PC2) correlated with the density of genes minus the density of enhancers at A-A anchors. An example of a PC2-negative genomic region is the *rims1* locus, which harbors dense stretches of enhancers located within its 519Kb gene body (Fig. 1j, top). This contrasts with a PC2-positive genomic locus, which harbors dense stretches of genes (Fig. 1j, bottom). Thus, A-A megaloop anchors may be separated into at least two subcompartments comprising "enhancer-dense" regions and "gene-dense" regions, and these regions formed robust

long-distance interactions with regions in the same subcompartment, but not with regions in a different subcompartment (Fig. 1k). These observations were also supported by non-negative matrix factorization (NMF)-based clustering (Extended Data Fig. 1h). Together, these results reveal the presence of distinct A subcompartments in mature cerebellar granule neurons.

The identification of A subcompartments led us next to examine the genes located at their anchors. The enhancer-dense subcompartment was enriched for genes that spanned over 100Kb and encoded synaptic and membrane-targeted signaling proteins (Fig. 1l and Extended Data Fig. 1i). On the other hand, the gene-dense subcompartment was enriched for shorter genes encoding zinc finger proteins and major histocompatibility complex proteins (Fig. 1l and Extended Data Fig. 1i). Together, these results suggest that each subcompartment contains different types of transcriptionally active genes in the adult brain.

Long genes are differentially recruited to the enhancer-dense subcompartment in granule neurons

We next asked whether the genomic subcompartments found in granule neurons were associated with the transcriptional regulation of these neurons across different cerebellar regions. The cerebellar vermis is composed of 10 lobules and each lobule is necessary for distinct sensorimotor and cognitive functions (Fig. 2a)[28,29]. We performed RNA-Seq analyses on lobules 4 to 9 along the dorsal cerebellar vermis in adult mice and found distinct gene expression patterns that distinguished these lobules (Extended Data Fig. 2a and Extended Data Table 1). Consistent with these observations, single nucleus RNA-Seq analyses of adult cerebellar granule neurons revealed distinct subtypes of granule neurons located in the anterior versus posterior cerebellar lobules (Extended Data Fig. 2b)[30]. Interestingly, genes differentially expressed across cerebellar lobules were typically longer than the average of all expressed genes in the cerebellum (Fig. 2b).

These results suggest that differential expression of long genes may help to define the identities of granule neurons across cerebellar lobules.

To characterize the roles of genome organization in the regulation of these lobule-specific long genes, we performed Hi-C analyses using lobule 4/5 or lobule 9 from the adult cerebellum and intersected these results with our RNA-Seq analyses. The lobule 4/5-specific gene *ptprt*, which spanned 1.1Mb on chromosome 2 and encoded for a synaptic adhesion protein[31], showed marked changes in long-distance genomic interactions between lobules 4/5 and 9 (Fig. 2c,d). Regions with increased interactions with *ptprt* in lobule 4/5 consisted of transcriptionally active loci such as the 272Kb *plcb4* gene (Fig. 2d,e) located 26Mb away from *ptprt* on chromosome 2. We visualized these long-distance genomic interactions by performing 3D DNA-fluorescent *in situ* hybridization (FISH) analyses using probes targeting the *ptprt* and *plcb4* genes and quantified their interaction frequency in the granule layer of lobule 4/5 or lobule 9 in the adult cerebellum (Fig. 2f,g). The co-localization of the *ptprt* and *plcb4* genes was significantly higher in granule neurons in lobule 4/5 compared to lobule 9, which is consistent with our Hi-C analyses (Fig. 2e-g).

We next determined if the lobule specificity of the long-distance intra-chromosomal interactions formed by the *ptprt* locus also extended to its inter-chromosomal interactions with different transcriptionally active subcompartments in the cerebellum. We found that the *ptprt* locus formed stronger inter-chromosomal interactions with the enhancer-dense subcompartment in lobule 4/5 compared to lobule 9 (Fig. 2h,i). By contrast, the *plcb4* locus interacted with the enhancer-dense subcompartment in both lobule 4/5 and lobule 9, which reflected the ubiquitous expression of *plcb4* mRNA across cerebellar lobules (Fig. 2h and Extended Data Table 1). The *ptprt* locus preferentially formed inter-chromosomal interactions with the enhancer-dense subcompartment rather than the gene-dense subcompartment (Fig. 2j). Similar observations were

made for the entire group of long genes enriched in either lobule 4/5 or lobule 9 (Fig. 2j). Each group of lobule-specific long genes formed robust inter-chromosomal interactions with enhancer-dense anchors, but not gene-dense anchors, in the corresponding lobule with higher long gene expression. Together, these results suggest that different groups of neuronal long genes are recruited to the enhancer-dense subcompartment to help establish the distinct transcriptional profiles of granule neurons across cerebellar lobules.

Mega-enhancers coalesce into nuclear bodies in adult cerebellar granule neurons

The discovery of an enhancer-dense subcompartment linked to the transcriptional identity of granule neuron subtypes in the cerebellum led us to further characterize this genomic subcompartment. Remarkably, H3K27ac-marked enhancers that spanned hundreds of kilobases emerged during cerebellar maturation, with the longest of these enhancers reaching over 500Kb in the adult cerebellum (Fig. 3a and Extended Data Fig. 3a). Thus, we termed these "mega-enhancers" due to their considerably greater length compared to super-enhancers, which have been reported in proliferating cells to span ~10Kb[32]. Similar observations were found in humans, where the brain harbored some of the longest active enhancers compared to other tissues (Extended Data Fig. 3b)[33]. We next asked if mega-enhancers shared characteristics with super-enhancers, such as a close association with genes that define cell identity[32]. Genes overlapping with active mega-enhancers in granule neurons showed higher levels of expression in these neurons compared to other cerebellar cell types (Extended Data Fig. 3c), suggesting that mega-enhancers may drive the expression of cell identity genes. In contrast, genes within the gene-dense subcompartment exhibited broad expression across various cell types in the cerebellum (Extended Data Fig. 3c).

We next examined the relationship between the activity of enhancers and the strength of their interactions with the enhancer-dense subcompartment. We categorized enhancers based on H3K27ac-enrichment, which reflects enhancer activity[32], and analyzed the strength of their inter-chromosomal interactions with the enhancer-dense subcompartment (Fig. 3b,c and Extended Data Fig. 3d). The top 1% and 5% ranked enhancers, which include mega-enhancers, exhibited robust interactions with the enhancer-dense subcompartment, but not with the gene-dense subcompartment (Fig. 3c and Extended Data Fig. 3d). However, enhancers in quartile 1, with median span of ~18Kb and comparable in size to super-enhancers found in non-neuronal cells, only exhibited mild interactions with the enhancer-dense subcompartment (Fig. 3c and Extended Data Fig. 3d). Mega-enhancers were highly enriched for the transcriptional regulator Brd4, which may regulate the expression of cell identity genes (Fig. 3d,e)[33]. These results suggest that mega-enhancers assemble into a transcriptionally active hub to control cell type-specific gene expression programs in neurons.

Recent microscopy approaches have revealed that the nucleus may be parcellated into zones which are enriched for distinct combinations of chromatin marks[15,34]. These zones may reflect the presence of nuclear bodies such as chromocenters or nuclear speckles, with the latter characterized by transcriptionally active chromatin modifications and functioning as hubs to recruit genomic loci with high gene densities from different chromosomes[14,15,34,35]. Having identified the enhancer-dense subcompartment in granule neurons, we next used microscopy to visualize this subcompartment and determine its interactions with known nuclear structures. Immunohistochemical analyses of the H3K27ac mark, a key feature of mega-enhancers, revealed prominent puncta in granule neurons in the adult cerebellum (Fig. 3f). We next tested whether genomic loci within the enhancer-dense subcompartment were co-localized with H3K27ac-

enriched puncta. We combined immunohistochemical analyses with DNA FISH analyses using probes targeting the *ptprt* locus, which interacts with the enhancer-dense subcompartment in lobule 4/5, but not lobule 9 in the cerebellum (Figs. 2h-j and 3g). Consistent with our Hi-C analyses, H3K27ac immunofluorescence showed increased levels near the *ptprt* locus in granule neurons from lobule 4/5 compared to lobule 9 (Fig. 3h,i). The transcriptionally active mark, serine 5 phosphorylated RNA polymerase II (Pol2 Ser5p), also exhibited increased levels near the *ptprt* locus in lobule 4/5 compared to lobule 9 (Fig. 3h,i). These data suggest that mega-enhancers and active transcription machinery associate to form discernable nuclear bodies in granule neurons.

We next analyzed the co-localization of the *ptprt* locus with markers of well-characterized nuclear bodies including nuclear speckles, Polycomb bodies, and chromocenters in granule neurons in the adult mouse cerebellum. There were little or no differences in the fluorescence levels of immunolabeled sf3a66 or the Hoechst dye, which marked nuclear speckles or chromocenters, respectively, near the *ptprt* locus in lobule 4/5 compared to lobule 9 in the cerebellum (Fig. 3h,i). By contrast, higher sf3a66 immunofluorescence levels were observed near a genomic locus known to associate with nuclear speckles[14,15] compared to the *ptprt* locus in granule neurons (Extended Data Fig. 3e). In other experiments, we found that H3K27me3 immunofluorescence levels were higher at the *ptprt* locus in lobule 9 compared to lobule 4/5, suggesting that the *ptprt* gene may be silenced by the Polycomb complex and excluded from the enhancer-dense subcompartment in the posterior cerebellum (Fig. 3h,i). Together, these results suggest that the enhancer-dense subcompartment, which includes the transcriptionally active *ptprt* locus in lobule 4/5, occupies a subnuclear locale that is separated from known transcriptionally active and repressive nuclear structures.

We also characterized the gene-dense subcompartment, which was identified in the adult cerebellum along with the enhancer-dense subcompartment (Fig. 1i-l). Genomic loci in the gene-dense subcompartment exhibited high levels of the H3K4me3 mark found at gene promoters, but were also adjacent to broad H3K9me3-marked heterochromatin regions corresponding to the transcriptionally repressive B-compartment (Extended Data Fig. 3f,g). We used imaging to confirm that a genomic locus in the gene-dense subcompartment was closer to chromocenters marked by Hoechst dye or H3K9me3 compared to the *rims1* locus located within the enhancer-dense subcompartment (Extended Data Fig. 3h). In other analyses, we found that anchors within the gene-dense subcompartment were concentrated on chromosomes 7 and 17, and these chromosomes also exhibited increased H3K9me3 levels with cerebellar development (Extended Data Fig. 3i,j). Recent reports that chromosomes 7 and 17 strengthen their inter-chromosomal interactions during neuronal differentiation in the olfactory bulb, hippocampus, and cortex are consistent with our findings in the cerebellum[15,18,19]. Together, these results reveal multiple mechanisms that regulate long-distance interactions between transcriptionally active loci in cerebellar granule neurons.

Small groups of mega-enhancers simultaneously interact within clusters in granule neurons

Having identified multiple nuclear subcompartments for active gene transcription in granule neurons, we next characterized the higher-order assembly of genomic loci within these subcompartments. However, a limitation of the Hi-C technique is that it only captures pairwise interactions between genomic loci within close spatial proximity[36]. Therefore, we took advantage of the split-pool recognition of interactions by tag extension (SPRITE) approach which allows for the interrogation of simultaneously interacting genomic loci within DNA-protein complexes of

varying sizes (Fig 4a)[14]. SPRITE contact maps of P22 cerebellum showed increased resolution of long-distance genomic interactions compared to Hi-C analyses (Fig. 4b), which has also observed in other cell types[14].

To assess how genomic loci were organized within nuclear subcompartments using SPRITE, we divided the genome into 500Kb bins and calculated the number of unique bins found within each DNA cluster. We also stratified SPRITE clusters based on the number of bins to determine how genome organization relates to the size of DNA-protein complexes (Fig. 4c,d). Interestingly, the *rims1* locus, which belongs to the enhancer-dense subcompartment, formed multi-way interactions within small SPRITE clusters at frequencies significantly above background, the latter determined using genomic regions that were not involved in long-distance interactions from Hi-C analyses (Figs. 1 and 4c). We further divided these clusters based on the distance between a genomic locus and other bins within the cluster. We found that the *rims1* locus was enriched in subclusters where other bins were over 20Mb away or located on different chromosomes, suggesting that *rims1* formed multi-way interactions over long genomic distances (Extended Data Fig. 4a). Extending these analyses to all regions within the enhancer-dense or gene-dense subcompartment, we found that the enhancer-dense subcompartment engaged in inter-chromosomal interactions significantly above background levels (Fig. 4d). Notably, the enhancer-dense subcompartment was enriched in smaller SPRITE subclusters compared to the gene-dense subcompartment (Fig. 4d). These results suggest that both the enhancer-dense and gene-dense subcompartments form higher-order genomic interactions between different chromosomes, but the extent of these multi-way interactions differs between the subcompartments.

To further characterize multi-way genomic interactions in granule neurons, we performed unsupervised dimensionality reduction for all genomic regions within SPRITE clusters. This

allowed us to visualize the similarity in inter-chromosomal interactions of genomic loci, as depicted by their proximity on UMAP plots. We found that the anchors of megaloops can be distinguished from non-megaloop regions across a wide range of SPRITE subcluster sizes (Extended Data Fig. 4b,c). Therefore, we next directed our analyses to consider only megaloop anchors. We observed that within smaller SPRITE clusters, a group of genomic loci from different chromosomes shared similar inter-chromosomal interaction preferences and were predominantly part of the enhancer-dense subcompartment (Extended Data Fig. 4d-f). These results support our findings that genomic regions within the enhancer-dense subcompartment show strong tendencies to engage in multi-way inter-chromosomal interactions within small genomic clusters.

Transcriptionally active nuclear subcompartments occupy distinct radial positions in neurons

Our insights into the higher-order assembly of regions within each transcriptionally active subcompartment in granule neurons led us to further investigate how these subcompartments may orient within the nuclear topography, including with respect to nuclear landmarks. We used our integrative genome modeling platform (IGM)[37,38] to generate a population of single cell 3D genome structures of granule neurons that were statistically consistent with the Hi-C datasets from P22 cerebellum (Fig. 4e and Methods). We previously showed that such structures can predict the folding of chromosomes relative to nuclear speckles, nucleoli, and the lamina compartments with good agreement with experimental measurements[37,38]. We first studied the radial positions of chromatin in the nucleus for each active subcompartment. Interestingly, we found that enhancer-dense anchors were located peripherally in the nucleus (Fig. 4e,f), clearly distinct from the more interior locations of gene-dense anchors (Fig. 4e,f and Extended Data Fig. 5a,b). By contrast, genomic loci associated with nuclear speckles showed the most interior nuclear locations in our

models. Moreover, in single cell models, each active subcompartment formed distinct spatial partitions, with enhancer-dense anchors forming the smallest spatial clusters among all active subcompartments, consistent with observations from our SPRITE analyses (Fig. 4d,e). We next characterized each genomic locus using 14 topographic markers, which represent the nuclear microenvironment of a genomic locus defined by the mean distances and association frequencies to nuclear speckles, nucleoli, and lamina and local chromatin fiber properties (Extended Data Fig. 5a-c)[37,38]. Enhancer-dense anchors showed distinct enrichment patterns of these topographic markers in comparison to those of gene-dense anchors and nuclear speckle-associated regions (Extended Data Fig. 5a-c). For instance, enhancer-dense anchors showed significantly longer distances to nuclear speckles (SpD) and nucleoli (NuD) among the transcriptionally active subcompartments (Extended Data Fig. 5a,b).

We next incorporated the radial positions from IGM into our SPRITE analyses to assess the relationship between higher-order genome organization and radial positioning. We focused on larger SPRITE clusters, which better represented global genome structure and the interactions between different subcompartments. We observed with UMAP that the gene-dense subcompartment, nuclear speckles, and other genomic loci predicted to localize in the nuclear interior were partially overlapping (Fig. 4g). Consistent with these findings, our imaging approach also revealed partial co-localization of the gene-dense subcompartment with nuclear speckles in granule neurons (Extended Data Fig. 3e). Conversely, the enhancer-dense subcompartment and genomic loci predicted to be peripherally located were found to be distributed away from the gene-dense subcompartment and nuclear speckles in the UMAP visualization of large SPRITE clusters (Fig. 4g). Together, these results support our model that transcriptionally active subcompartments are spatially partitioned between the nuclear interior and periphery in granule neurons.

Next, to test the predictions of the nuclear localization of subcompartments from IGM and SPRITE analyses, we performed DNA FISH and immunohistochemical analyses to visualize the enhancer-dense subcompartment, gene-dense subcompartment, and nuclear speckles, and assessed their radial positioning in granule neurons in the adult cerebellum. The radial positions of genomic loci measured with imaging aligned with the predictions made by IGM modeling, and confirmed that the enhancer-dense subcompartment was localized to the outer half of the nucleus in granule neurons (Fig. 4h and Extended Data Fig. 5d,e). We also observed outward nuclear localization for the H3K27ac mark, which was associated with mega-enhancers (Fig. 4h and Extended Data Fig. 5d,e). In contrast, the inner half of the nucleus was concentrated for genomic loci from the gene-dense subcompartment and nuclear speckles, along with their associated protein markers (Fig. 4h and Extended Data Figs. 3g and 5d,e). Collectively, our findings uncover multiple transcriptionally active subcompartments in granule neurons, including the newly identified mega-enhancer bodies. The unique microenvironment formed by these mega-enhancer bodies likely regulates key gene expression programs that shape neuronal identity.

Discussion

Our study unveils new principles for higher-order genome organization in defining neuronal cell types in the brain. We report that long-distance genomic interactions spanning across TADs and chromosomes are formed during postnatal cerebellar development. These interactions, which we termed as megaloops, encompass several distinct transcriptionally active subcompartments including an enhancer-dense subcompartment and a gene-dense subcompartment adjacent to heterochromatin. Notably, the enhancer-dense subcompartment is enriched for neuronal long genes that encode synaptic and membrane-targeted signaling proteins. Distinct sets of these neuronal long genes are recruited to the enhancer-dense subcompartment to help shape the transcriptional identity of granule neuron subtypes found across the cerebellum. By combining molecular genomics, genome modeling, and imaging, we also reveal that the enhancer-dense subcompartment is formed from numerous smaller DNA clusters situated in a unique microenvironment near the nuclear periphery, spatially separated from other transcriptionally active subcompartments in the nuclear interior. These findings shed light on novel nuclear mechanisms that control precise gene expression patterns underlying neuronal identity in the brain.

The prevalence of “mega-enhancers" in mammalian brain neurons plays a significant role in the establishment of the enhancer-dense subcompartment. These mega-enhancers are characterized by remarkably long stretches of the H3K27ac mark spanning hundreds of kilobases, which are over an order magnitude larger than the super-enhancers found in non-neuronal cell types[32]. Super-enhancers are reported to form three-dimensional (3D) interactions with neighboring gene promoters or enhancers, but these interactions are primarily confined within TADs[39,40]. A notable exception is in olfactory neurons, where Greek Island enhancers which are comparable to super-enhancers, interact over long genomic distances to stabilize expression of a

chosen olfactory receptor gene[17]. Moreover, under certain genetic perturbations, such as the depletion of the cohesin subunit Rad21, super-enhancers have been observed to form robust inter-chromosomal interactions[8,41]. In contrast, mega-enhancers coalesce across different chromosomes to form a distinct subcompartment in neurons even under physiological conditions. It remains to be determined how mega-enhancers in neurons establish robust long-distance interactions. Despite their differences, mega-enhancers share certain characteristics with super-enhancers, including an association with genes involved in cellular identity and the recruitment of the transcriptional regulator Brd4. Nuclear condensates are formed between super-enhancers, transcription regulators, and RNA through phase separation[42]. However, the mechanisms that drive mega-enhancers to form the enhancer-dense subcompartment are unknown. Exploring potential roles for both coding and non-coding RNAs, along with other chromatin mechanisms, in the regulation of the enhancer-dense subcompartment will be important for understanding the unique cellular biology of neurons in the mammalian brain.

The spatial distribution of nuclear bodies such as nuclear speckles or the nucleolus plays crucial functions in organizing the 3D genome and in gene transcription in cells[34,43]. Through the integration of experimental and computational approaches, we found that the enhancer-dense subcompartment forms an independent microenvironment in the outer-half of the nucleus. In contrast, the gene-dense subcompartment is concentrated in the nuclear interior alongside nuclear speckles. These findings raise questions on why mega-enhancers within the enhancer-dense subcompartment are positioned in a distinct locale in neurons. One possibility is that the peripheral localization of mega-enhancers could improve sensitivity in detecting signals triggered by receptor activation in neurons[1]. For example, once protein signals translocate through the nuclear pore, they may more readily activate mega-enhancers and their associated neuronal long genes, which are

located near the nuclear envelope. Another possibility is that because long genes have been reported to physically expand during transcription[43], the nuclear periphery may provide additional space to accommodate the transcription of the numerous long genes found in neurons in the mammalian brain. Given the larger volume of the nuclear periphery compared to the nuclear interior, mega-enhancer bodies may also be important for recruiting transcriptional regulators to sustain the expression of this large group of long genes.

Recent studies on mouse chromosome 7 and 17 in excitatory forebrain neurons suggest that significant parts of these chromosomes move towards the nuclear interior during neuronal maturation[15,18,19]. This is consistent with our findings on the gene-dense subcompartment, which is concentrated on these chromosomes in granule neurons and is enriched within the nuclear interior in adult granule neurons. These observations suggest that there may be shared mechanisms between neuronal cells to re-localize these genomic loci upon neuronal differentiation. The positioning of genes within the nuclear interior and particularly in proximity to nuclear speckles may increase the efficiency of mRNA processing[44]. However, these observations also suggest that the more peripherally-located neuronal long genes, which also harbor exceptionally long introns, may employ distinct mechanisms for mRNA processing compared to that of short genes situated near nuclear speckles.

Together, our study reveals novel nuclear structures, including mega-enhancer bodies, that serve as the architectural backbone for organizing various classes of genes in neurons. These specialized nuclear compartments likely perform critical biological functions in neurons, such as orchestrating gene expression programs that give rise to distinct neuronal cell types in the brain. In the cerebellum, anterior versus posterior granule neurons exhibit different intrinsic properties and synaptic plasticity mechanisms, which support regional processing of sensory, motor, and

cognitive signals[45-47]. Our discovery that enhancer-dense subcompartment is tightly linked to the identity of anterior and posterior granule neurons unveils new mechanisms to establish functional differences between regions in the brain. Finally, neuronal long genes encoding synaptic proteins have been implicated as mutation hotspots for autism spectrum disorders and other neurodevelopmental diseases[48]. Elucidating how higher-order genome organization regulates these genes will advance our understanding of the mechanisms that ensure the proper assembly and patterning of neural circuits in humans.

Acknowledgments

We thank members of the Yang Lab for helpful discussions and the Biological Imaging Facility at Northwestern University (RRID:SCR_017767) for imaging. This project was supported by NIH 4DN grants U01DA053691 (T.Y.). and UM1HG011593 (F.A.), and R01NS123285 (Y.Y.).

Authors contributions

Z.Z., O.P., Y.Y. and T.Y. designed the study and wrote the manuscript. Z.Z., O.P., N.S., S.F., Y.Y. and T.Y. performed RNA-Seq, ChIP-Seq, Hi-C, SPRITE, DNA FISH, immunohistochemistry, and bioinformatics analyses. F.M. and F.A. generated and analyzed the IGM models and described the results.

Author Information

The authors declare no conflicts of interest.

Figure Legends

Figure 1

(a) Experimental design. The cerebellum from postnatal day 6 or 22 mice was dissected and subjected to Hi-C analyses. (b) Hi-C map of the cerebellum at P22 versus P6. Fold changes in interactions are shown. (c) A histogram of intra-chromosomal contact frequencies stratified by genomic distances between regions (n = 3 biological samples). Data show mean and shading denote s.e.m. (d) Distribution of intra-chromosomal loops at various genomic distances between loop anchors. The median and maximum sizes of all detected loops from Hi-C analyses of P6 and P22 cerebellum are indicated. Long-distance "megaloops" spanning more than 3Mb significantly increase with cerebellar development. (e) An example of a megaloop formed between genomic loci located ~165Mb apart on chromosome 1 during cerebellar development. (f) A/B compartment identity of megaloop anchors. With neuronal maturation, megaloops were primarily formed between genomic loci in the active A compartment. (g) Heatmap of significantly differentially interacting A-A megaloops during cerebellar development (two-sided *P* value from negative binomial distribution with Benjamini–Hochberg post hoc test, n = 3 biological samples, false discovery rate (FDR) < 0.05, $\log_2$ mean centered). (h) An example from Hi-C analyses of a clique of long-distance genomic interactions spanning across chromosomes 1 and 12, and aligned with UCSC genome browser tracks of the active histone marks H3K4me3 and H3K27ac in the P22 cerebellum. Arrows denote A-A megaloops and yellow highlights denote megaloop anchors. (i) Left, principal component analysis (PCA) of megaloop anchor interactions. Right, Pearson correlation of principal components 1 (PC1) and 2 (PC2) with various features. PC1 correlated with the density of transcriptionally active genes and enhancers. PC2 correlated with the density of enhancers versus the density of genes. (j) UCSC genome browser tracks depicting examples of

enhancer-dense and gene-dense anchors, which belong to distinct A subcompartments. (k) Interaction frequencies within and between subcompartments. Enhancer-dense anchors interacted with each other but failed to interact with gene-dense anchors ($P = 0$, Kruskal-Wallis test with Dunn's post hoc test). Box plots show median, quartiles (box) and range (whiskers). (l) Gene ontology analyses of the A subcompartments. Enhancer-dense anchors contained genes encoding for synaptic and signaling proteins.

Figure 2

(a) Sagittal view of the mouse cerebellar vermis. Dorsal lobules (L) 4/5, 6, 7, 8, and 9 are denoted. (b) The length of genes differentially expressed (DE) across the dorsal cerebellar lobules compared to the length of all expressed genes in the cerebellum ($P = 5.1 \times 10^{-89}$, two-sided Mann-Whitney-Wilcoxon test, $n = 10324$, 2498 for expr genes, DE L4-9). (c) RNA-Seq tracks of the dorsal cerebellar lobules at the *ptprt* locus. *Ptprt* mRNA was enriched in the anterior cerebellum, including in lobule 4/5. (d) Hi-C interaction map of chromosome 2 showing the changes in genomic interactions between lobule 4/5 and lobule 9. The *ptprt* locus exhibited pronounced changes in long-distance interactions between these lobules. (e) A zoomed in view from (d) illustrating changes in long-distance interactions between the *ptprt* and *plcb4* loci across lobules 4/5 and 9. (f) Representative images of DNA FISH using probes targeting the *ptprt* (red) locus and the *plcb4* (green) locus. Left, 3D projection of the granule layer in cerebellar lobule 4/5. Right, 2D cross section of a single nucleus. The dashed line indicates the boundary of the nucleus. (g) The co-localization frequency of *ptprt* and *plcb4* in lobules 4/5 and 9 ($P = 0.012$, two-tailed t-test, n = 3 mice). Data show mean and error bars denote s.e.m. (h) Hi-C maps depicting the aggregate inter-chromosomal interactions between all enhancer-dense anchors and the *ptprt* locus or *plcb4* locus

in lobule 4/5 or 9. (i) Hi-C maps depicting inter-chromosomal interactions between the *ptprt* locus on chromosome 2 and regions on chromosome 1 harboring A-A anchors in lobules 4/5 and 9. The *ptprt* locus selectively formed inter-chromosomal interactions with transcriptionally active regions in lobule 4/5, but not in lobule 9, as indicated with the dashed lines. (j) Hi-C analyses of inter-chromosomal interactions between lobule-enriched long genes and the enhancer-dense or gene-dense subcompartment in lobule 4/5 or lobule 9. Long genes preferentially interacted with enhancer-dense anchors in a lobule-specific and gene expression-dependent manner (*$P < 0.05$, ***$P < 0.001$, Friedman test with Nemenyi's post hoc test). (b, j), box plots show median, quartiles (box) and range (whiskers).

Figure 3

(a) The length of stitched enhancers and the number of stitched enhancers spanning over 100Kb in postnatal day 6 or 22 cerebellum. Robust increases in enhancers spanning over 100Kb were observed during cerebellar development. (b) The fraction of stitched enhancers overlapping with gene bodies, categorized by H3K27ac levels at enhancers. The median length of stitched enhancers for each group is indicated. (c) The interaction frequency of stitched enhancers, grouped as in (b), with the enhancer-dense or gene-dense subcompartment. The top 1% or 5% of enhancers, which include mega-enhancers, formed robust interactions with megaloop anchors within the enhancer-dense subcompartment. By contrast, the gene-dense subcompartment formed little or no inter-chromosomal interactions with the enhancer-dense subcompartment. Box plots show median, quartiles (box) and range (whiskers). (d) UCSC genome browser tracks showing H3K27ac and Brd4 levels at the mega-enhancer (ME) region of the *rims2* locus. (e) The enrichment of H3K27ac and Brd4 at stitched enhancers grouped by H3K27ac levels as in (b). (f) The cerebellar cortex

labeled with antibodies targeting H3K27ac (green) and the granule neuron marker NeuN (red) together with a Hoechst DNA dye (blue). H3K27ac puncta were observed in granule neuron nuclei and were distinguishable from heterochromatic chromocenters. Purkinje cell layer (PCL), molecular layer (ML), and internal granule layer (IGL) are denoted. The dashed line indicates the boundary of a nucleus. Scale bar, 25 μm (left), 1 μm (right). (g) An example of H3K27ac immunolabeling (green) combined with DNA FISH using a probe targeting the *ptprt* locus (red) and Hoechst DNA dye labeling (blue) in cerebellar lobule 4/5. 3D projection of labeled neurons (left) and a cross section of a single neuron (right) in the IGL of lobule 4/5. The arrow indicates the *ptprt* allele observed within this cross section and the dashed line indicates the boundary of the nucleus. (h, i) 2D intensity maps (h) or bar graphs (i) depicting normalized fluorescence signals of the indicated features surrounding the *ptprt* locus in granule neurons from lobule 4/5 or 9 (n = 3 mice). (e, i), data show mean and shading or error bars denote s.e.m.

Figure 4

(a) Schematic of the SPRITE technique. (b) Long-distance interactions between regions in chromosome 2 from Hi-C (middle) or SPRITE (right) analyses together with H3K4me3 ChIP-Seq tracks (left and bottom) of P22 cerebellum. Pairwise contacts from SPRITE were downweighted based on cluster size[14]. Arrows indicate megaloops from Hi-C. (c) Top, UCSC genome browser track of chromatin mRNA from the adult mouse cerebellum[13]. *Rims1* was highly transcribed, while little or no *kcnq5* mRNA was detected in the adult cerebellum. Bottom left, The relative observed frequencies of the *rims1* locus, the *kcnq5* locus, and all non-megaloop control (ctrl) regions within SPRITE clusters of various sizes. Bottom right, the observed frequencies of the *rims1* or *kcnq5* locus within SPRITE clusters normalized to that of control regions. The *rims1* locus was observed

at significantly higher frequencies within SPRITE clusters of sizes 4 to 32 compared to control regions (n = 2 biological replicates). (d) Normalized frequencies of enhancer-dense (left) or gene-dense (right) subcompartment anchors appearing in SPRITE subclusters. Subclusters were selected based on the genomic distances of loci from either the same chromosome (<5Mb or >20Mb, intra-chrom) or from different chromosomes (inter-chrom) in relation to enhancer-dense or gene-dense subcompartment anchors. (e) Top, the structure of an individual granule neuron nucleus generated with IGM. Bottom, IGM models highlighting the position of enhancer-dense (red), gene-dense (blue), and nuclear speckle (yellow) anchors, showing their center of mass. (f) IGM predicted radial distributions of enhancer-dense (red), gene-dense (blue), and nuclear speckle (yellow) associated genomic anchors. A single structure (left) and population average (right) are shown. (g) UMAP plot representing the inter-chromosomal interaction preferences of megaloop anchors within large SPRITE subclusters. Top, anchors within the enhancer-dense subcompartment (red), gene-dense subcompartment (blue), or nuclear speckles (gold) are highlighted. Bottom left, the chromosomes forming a distinct grouping on the UMAP plot and harboring anchors from nuclear speckles and the gene-dense subcompartment (top right). Bottom right, radial positions of megaloop anchors calculated from IGM and projected onto the UMAP plot of large SPRITE subclusters. (h) Top, examples of granule neurons from adult cerebellum labeled using sf3a66 (white) or H3K27ac (red) antibodies, DNA FISH probes targeting genomic loci associated with the enhancer-dense (red) and gene-dense (cyan) subcompartments, and the Hoechst DNA dye (blue). The dashed lines indicate the boundaries of nuclei. Bottom, radial positioning of genomic loci and nuclear markers. DNA FISH probes targeting *rims1* (r1), *plcb4* (r2), and *ptprt* (r3) were used to identify anchors within the enhancer-dense subcompartment. H3K27ac antibodies were sourced from rabbit (rb) and mouse (ms). In all panels, data show mean

and shading or error bars denote s.e.m.

Methods

Animals

C57BL/6 mice were purchased from Charles River Laboratories or maintained under pathogen-free conditions. All animal experiments were done according to protocols approved by the Animal Studies Committee of Northwestern University in accordance with the National Institutes of Health guidelines.

Antibodies

Antibodies to H3K4me3 (Abcam, ab8580), CTCF (Millipore, 07-729), Rad21 (Abcam, ab992), H3K27ac (Active Motif, 39085, lot#17019020; 39134, lot#06921014), phosphorylated pol II Ser5 (Abcam, ab5408), H3K27me3 (Millipore, 07-449, lot#2919706), sf3a66 (Abcam, ab77800), NeuN (Abcam, ab177487), Brd4 (Bethyl Laboratories, A301-985A, lot#7), and H3K9me3 (Abcam, ab8898) for immunohistochemistry or ChIP-Seq experiments were purchased.

Statistics

Statistical analyses were performed using Matlab, Microsoft Excel, or R. Normality of data was assessed using the Shapiro–Wilk test. For experiments in which only two groups were analyzed, the t-test was used for normal distributions. For two groups with non-normal distributions, the Mann-Whitney-Wilcoxon test (unpaired data) or Wilcoxon signed rank test (paired data) was used. Comparisons within multiple groups were performed using analysis of variance (ANOVA) for normal distributions, Kruskal-Wallis followed by Dunn's post hoc testing for unpaired non-normal distributions, or Friedman test followed by Nemenyi's post hoc testing for paired non-normal distributions, and p-values were adjusted with the Benjamini-Hochberg procedure.

RNA-Seq

Total RNA was extracted from the cerebellum of postnatal day 6 and 22 mice or cerebellar lobules 4/5, 6, 7, 8, and 9 microdissected from postnatal day 56 mice using Trizol (Thermo Fisher Scientific) according to the manufacturer's instructions. ~100 ng of RNA was treated with NEBNext rRNA Depletion Kit (New England Biolabs). RNA-Seq was performed using libraries prepared with NEBNext Ultra™ Directional RNA Library Prep Kit for Illumina (New England Biolabs) and sequenced on the Illumina NextSeq 500 platform to obtain 36bp paired-end reads. Four biological RNA-Seq replicates were performed in all experiments.

ChIP-Seq

ChIP-Seq assays were performed as described with modifications[49,50]. Cerebella from postnatal day 6 and 22 mice were fixed with 1.1% formaldehyde solution. Immunoprecipitation was performed in RIPA buffer (10 mM Tris-HCl pH 8.0, 140 mM NaCl, 0.1% SDS, 1% Triton-X, 0.1% DOC, 1 mM EDTA, 0.5 mM EGTA) using the indicated antibodies with dynabeads protein G (Thermo Fisher Scientific).

For the Brd4 antibody, dynabeads protein G were preincubated with the antibody for several hours and washed with RIPA buffer before immunoprecipitation. Lysates were added to the antibody-coupled beads and incubated at 4C overnight. After extensive washes of beads with RIPA buffer three times, high salt RIPA buffer (10 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.1% SDS, 1% Triton-X, 0.1% DOC, 1 mM EDTA, 0.5 mM EGTA) three times, LiCl buffer (10 mM Tris-HCl pH 8.0, 150 mM LiCl, 1% Triton-X, 1% DOC, 1 mM EDTA, 1 mM EGTA) and TE (10 mM Tris-HCl, 1mM EDTA) twice, tagmentation reactions were performed as described[51]. Briefly,

immunoprecipitated DNA fragments on the beads were digested with Tn5 (Illumina), free DNA fragments were washed away by RIPA buffer and TE. The DNA fragments on the beads were treated with proteinase K (Thermo Fisher Scientific) for 1hr at 37C and de-crosslinked at 65C overnight. DNA fragments were purified with PCR purification kit (Qiagen) and amplified with primers containing indexes for sequencing using Q5 High-Fidelity 2xMaster mix (New England Biolabs).

For the H3K27ac, H3K4me3, H3K27me3, H3K9me3 or Rad21 antibody, immunoprecipitation was performed overnight at 4C. For the CTCF antibody, Protein G Sepharose (GE Healthcare) was added and incubated at 4C for 1 hr. For all other antibodies, dynabeads protein G were added and incubated at 4C for 1 hr. After extensive washes of beads with RIPA buffer three times, high salt RIPA buffer three times, and TE three times, DNA fragments were eluted in elution buffer (100 mM NaHCO3, 1% SDS, 10mM DTT at room temperature for Sepharose beads; 10 mM Tris-HCl pH 8.0, 350mM NaCl, 1% SDS, 0.1 mM EDTA at 65C for dynabeads) for 30 min, treated with proteinase K for 1hr at 37C, and de-crosslinked at 65C overnight. DNA fragments were purified with PCR purification kit (Qiagen). ChIP-Seq was performed using libraries prepared with the NEBNext Ultra™ II DNA Library Prep Kit for Illumina (New England Biolabs) as per the manufacturer's instructions.

All libraries were sequenced on an Illumina NextSeq 500 or 550 platform to obtain 36-37bp paired-end reads. Two to four biological ChIP-Seq replicates were performed in all experiments.

Hi-C

Hi-C assays were performed as described[12,13]. Cerebella from mice at postnatal day 6 and 22 or microdissected cerebellar lobule 4/5 or 9 from postnatal day 56 mice were fixed with 1%

formaldehyde. Hi-C libraries were generated and sequenced on the Illumina NextSeq 500 (P6 and P22) or 550 (lobule 4/5 and lobule 9) platform to obtain 36-37bp paired-end reads. Three biological Hi-C replicates were performed in all experiments.

SPRITE

SPRITE experiments were performed as described with modifications[14]. Briefly, cerebella from postnatal day 22 mice were fixed with 1% formaldehyde in PBS. ~10 billion molecules of DNA-protein complexes (~6000 cells) were coupled with 1.7 ml of NHS-Activated Magnetic Beads (Pierce) per experiment. After five rounds of split-barcoding reactions (DPM, odd, even, odd, and terminal odd), SPRITE libraries were generated and sequenced on the Illumina HiSeq X or NovaSeq platform (Admera Health) to obtain 150bp paired-end reads. Two biological SPRITE replicates were performed.

DNA FISH

DNA FISH assays were performed as described with modifications[13,52]. Cerebella prepared from 6-8 week-old mice were fixed with 4% PFA and 4% sucrose by transcardiac perfusion. Sagittal cryosections (12 μm thickness) were prepared using a cryostat (Leica Biosystems or Epredia) and cerebellar sections containing the vermis area were used for further analyses. Sections were treated with sodium citrate solution (10mM sodium citrate, pH 6.0) for 10 min at 70C and cooled down to room temperature. Sections were dehydrated with sequential treatment of ice-cold 70% EtOH for 2 min, ice-cold 90% EtOH for 2 min, and ice-cold 100% EtOH for 2 min, and air-dried for 5 min. For hybridization, genomic DNA was denatured by incubating glass plates with buffer containing 2x SSC and 70% formamide for 2.5 min at 73C, followed by incubation with 2x SSC

and 50% formamide for 1 min. All probes were made from BAC clones (BACPAC Resources Center) including RP24-337G9 (*ptprt* locus), RP24-209O19 (*plcb4* locus), RP24-120L14 (*rims1* locus), RP24-153C13 (gene-dense subcompartment locus), and RP24-358P20 (nuclear speckle locus) by nick translation with chemical coupling using an Alexa Fluor succinimidyl ester (Alexa555 for *ptprt*; Alexa647 for *plcb4*; Alexa488 for *rims1*; Alexa555 for gene-dense locus; Alexa647 for nuclear speckle locus) (Thermo Fisher Scientific). The hybridization solution contained ~70-80 ng of each labeled probe, 6 μg of mouse Cot-1 DNA (Thermo Fisher Scientific), and 10 μg of sheared salmon-sperm DNA (Thermo Fisher Scientific) in hybridization buffer (10% dextran sulfate, 50% formamide, 2xSSC). The probe mixtures were denatured at 73C for 5 min before use. Denatured sections and probes were sealed and incubated in an incubator at 37C overnight. On the next day, coverslips were removed and washed once in 2x SSC and 50% formamide solution for 15 min at 45C and three times in 2x SSC for 5 minutes at 45C with gentle agitation. Sections were washed once with PBS and stained with Hoechst dye. After washing with PBS, sections were mounted with Fluoromount-G mounting medium (Southern Biotech). Images were acquired at the Biological Imaging Facility at Northwestern University on a confocal laser scanning microscope (Leica Microsystems, TCS SP8 confocal) using a 63x objective lens (NA 1.40). A stack of 0.3 μm thick optical sections was acquired for each field of view in the UV, red, and far-red channels. Two to three biological DNA FISH replicates were performed in all experiments.

Immunohistochemistry

Cerebella sections from 6-8 week-old mice were prepared and labeled with the relevant antibodies as previously described[49]. Secondary antibodies included anti-mouse or rabbit IgG Alexa 488

(Thermo Fisher Scientific, A11029 or A11034) or anti-rabbit Alexa647 (Abcam, ab150075).

DNA FISH combined with immunohistochemistry

Cerebella sections from 6-8 week-old mice were prepared and treated with sodium citrate solution as described above. Sections were then labeled with the relevant primary and secondary antibodies. After extensive washing of secondary antibodies using PBS with 0.3% Tween 20, sections were fixed again with 4% PFA in PBS for 10 min at room temperature, washed with PBS for 5 min twice and with 2xSSC once for 5 min, and treated with ice-cold HCl solution (0.1M HCl, 0.7% Triton-X) for 15 min at room temperature. Subsequently, DNA FISH procedures including prehybridization, hybridization, and washing steps, and imaging were performed as described above. Three biological replicates were performed in all experiments.

RNA-Seq analyses

For bulk RNA-Seq analyses, sequenced reads were aligned to the mm10 reference genome with HISAT2 using the public server at https://usegalaxy.org/ and normalized by library size. Gene annotations were derived using GENCODE version M11. Differential gene expression using pair-wise negative binomial tests with edgeR[53] and the false discovery rate (FDR) were calculated for all genes.

Single nucleus RNA-Seq (snRNA-Seq) datasets from mouse cerebellar lobules and cell type annotations were obtained[30] and analyzed using Seurat 4[54]. Briefly, granule neurons or all cells from lobule 4/5 were extracted for downstream analysis. Read counts were normalized and variance stabilized using sctransform and subjected to dimensionality reduction using PCA preprocessing followed by UMAP embedding[54]. Module scores were calculated from the average

expression of genes within the module, subtracted by the levels of similarly expressed control genes[54,55].

ChIP-Seq analyses

ChIP-seq reads were aligned to the mm10 reference genome with Bowtie2 using the public server at https://usegalaxy.org/ and normalized by library size. Peak calling was performed using MACS2[56]. H3K27ac peaks >2 kb distal to H3K4me3-bound TSSs were considered as enhancers. Enhancer ranks were calculated as previously described[32]. H3K27ac peaks within 12.5Kb were stitched together. H3K27ac reads within each enhancer stretch was normalized by library size, input subtracted, and ordered by rank.

To calculate the average ChIP-Seq profile at enhancer stretches, the centers of enhancer stretches were aligned and the read density was calculated as a function of the distance to center. The standard error of the mean was calculated using two to four biological replicates of the ChIP-seq condition.

Hi-C analyses

Hi-C reads were aligned to the mm10 reference genome with Bowtie2 and HiC-Pro as described[57]. Uniquely mapped paired-end reads were assigned to MboI restriction fragments and valid pairs with a minimum genomic distance of 1Kb were filtered for PCR duplicates. Contact frequency as a function of genomic distance was calculated as described with modifications[58]. The frequency of valid pairs was calculated as the sum of observed contacts per $\log_2$ distance bin divided by all observed contacts. Interaction matrices were normalized using Knight-Ruiz (KR) matrix balancing and observed divided by expected (o/e) transformation and visualized using Juicebox[59].

HiCCUPS[12] was used at 5Kb, 10Kb, and 25Kb resolutions to identify intra-domain loops and at 25Kb, 50Kb, and 100Kb resolutions to identify megaloops. Arrowhead was used at 5Kb, 10Kb, and 25Kb resolutions to identify domain boundaries and intersected with loops to identify loop domain anchors enriched for convergent CTCF binding sites as described[12]. A/B compartment strength was derived from the eigenvector (λ) of the Hi-C correlation matrix[12] using 25Kb bins. Genomic loci enriched in the active histone modification H3K27ac and de-enriched in the repressive histone modification H3K9me3 were assigned a positive λ and denoted as the active A compartment, while genomic loci with a negative λ were denoted as the repressive B compartment. Aggregate peak analysis (APA)[12] was performed using lists of megaloops formed by interactions within the active A compartment.

Differential interactions between loop domain anchors or megaloop anchors across cerebellar development were calculated from biological replicates using the edgeR package[53]. Raw interaction counts were normalized using the offsets parameter by KR balancing and o/e transformation, and a generalized linear model-based test was used to identify significant differential interactions.

Dimensionality reduction was performed on megaloop anchors in the active A compartment using Hi-C datasets from P22 cerebellum at 500Kb resolution. The intra-chromosomal and inter-chromosomal interaction matrix was normalized by KR balancing and o/e transformation. Principal component analysis (PCA) was computed using the covariance matrix of normalized contacts, and non-negative matrix factorization (NMF) was computed directly from the normalized contact matrix in Matlab. The NMF rank (k) was approximated by computing at each k the linear regression of lower and higher k residuals and minimizing the joint error sum of squares from the linear regressions[60]. PCA separated the gene-dense and enhancer-dense subcompartments into

positive and negative scores of PC2, while NMF separated the gene-dense and enhancer-dense subcompartments into two factors. Comparison of PCA and NMF approaches revealed that NMF factors exhibited better correlation with identifiable chromatin features. Thus, the top 100 bins at 500Kb resolution with the highest NMF factor scores for the gene-dense subcompartment or the enhancer-dense subcompartment were selected for downstream analyses.

To identify at higher resolution the genomic loci associated with transcriptionally active subcompartments, the genome was divided into 5Kb bins and their normalized inter-chromosomal interactions with the top 100 500Kb bins of the gene-dense and enhancer-dense subcompartments were calculated. Genomic loci that belonged to the repressive B-compartment were excluded. 5Kb bins with >3-fold enrichment for the gene-dense versus enhancer-dense subcompartment were assigned to the corresponding enriched subcompartment.

Gene features including gene length, exons, and transcriptional start and termination sites were obtained from GENCODE mouse version M11. Gene ontology analyses for genes overlapping with transcriptionally active subcompartments were performed using DAVID Bioinformatics Resources 6.8[61] with a background of all expressed genes in the cerebellum.

DNA FISH and immunohistochemical analyses

For each cell, DNA FISH signals were subjected to automated identification using custom macros in ImageJ as previously described[13,62]. The centroid of DNA FISH signals was detected using a 3D gaussian filter to reduce noise followed by 3D spot segmentation. For co-localization between pairs of DNA FISH probes, the Euclidian distance to nearest alternate DNA FISH probe was calculated and distances < 350nm were considered as co-localized.

The enrichment of immunohistochemical signals surrounding DNA FISH probe signals was analyzed using the z-sections containing the centroid of each DNA FISH probe signal. The average pixel intensity of the immunohistochemical signal in a 300nm x 300nm area in the x-y plane surrounding the centroid was background subtracted using the bottom 20th percentile of signal in the nucleus and normalized to the top 90th percentile of signal in the nucleus. For visualization using aggregate pixel analysis, a $1\mu m^2$ area surrounding the DNA FISH centroid was obtained.

Radial positioning of DNA FISH or immunohistochemical signals was analyzed using a single z-section in the middle third (z-axis) of a granule neuron nucleus to ensure a maximal x-y area for calculations. For DNA FISH analyses, the z-section containing the centroid of the DNA FISH probe signal was used. Loci failing to meet both criteria were excluded from analysis. For immunohistochemical analyses, the top 1% pixels by intensity for each neuron were identified. The radial distances from the centroid of the nucleus to the DNA FISH probe signal or immunohistochemical signals were then calculated, where 0 represents the nuclear center and 1 represents the nuclear membrane.

Distances to chromocenters were analysed via a custom ImageJ macro. DNA FISH probe centroids were detected using a 3D Gaussian Blur followed by 3D Object Counter. Extracted centroids were overlaid on the chromocenter channel, and a line was draw between each DNA FISH probe and their nearest chromocenter border to extract the Euclidean distance. DNA FISH probes for which the closest chromocenter was found on a different z-section were excluded from analysis.

SPRITE analyses

SPRITE demultiplexing and read count normalization were performed as described with modifications[14]. The DPM sequence was identified from the beginning of read 1 and the odd, even,

odd, and terminal odd sequences were identified from read 2 using the Guttman Lab SPRITE pipeline (https://github.com/GuttmanLab/sprite-pipeline) and together this sequence combination represented the barcode sequence. After trimming the DPM sequence from read 1, the read was aligned to the mm10 reference genome. Mapped reads with MAPQ scores > 30 and a fully identified barcode sequence were kept. Identically mapped reads for a particular barcode sequence were considered PCR duplicates and filtered. After pre-processing, all mapped genomic coordinates with the same barcode sequence were grouped together as a SPRITE cluster. To generate pairwise contact maps, each pairwise interaction was downweighted by 2/n, where n is the total number of reads within a SPRITE cluster. For comparisons with Hi-C analyses, pairwise contacts < 1Kb were removed. Downweighted contacts were then normalized with KR matrix balancing using Juicer[59].

To analyze long-distance interactions, genomic loci within SPRITE clusters were assigned to 500Kb bins. Within each SPRITE cluster, overlapping reads in a bin were counted only once, focusing downstream computations on long-distance genome structure[14,63]. Clusters containing a minimum of 2 bins and a maximum of 1024 bins were analyzed. For each genomic bin, its observed frequency distribution over a range of cluster sizes was calculated. For a group of genomic bins, the observed frequency distributions were averaged. Genomic loci that did not form megaloops in Hi-C analyses were used to generate control frequency distributions. SPRITE subclusters were generated based on the genomic distances of loci from either the same chromosome (<5Mb or >20Mb, intra-chromosomal contacts) or from different chromosomes (inter-chromosomal contacts) in relation to a reference genomic bin. Genomic bins with low coverage, defined as those observed in fewer than 1000 clusters or 100 subclusters, were excluded from frequency distribution analyses.

To calculate interaction frequencies with A subcompartments, SPRITE subclusters harboring anchors located on different chromosomes relative to a reference genomic bin were considered. The fraction of bins, excluding the reference genomic bin, that overlapped with anchors within the enhancer-dense or gene-dense subcompartment was calculated for each SPRITE subcluster. The mean fractional overlap for each group of reference genomic bins was then computed across a range of SPRITE subcluster sizes and defined as the observed inter-chromosomal interaction frequency. Genomic bins containing non-megaloop regions were used to generate control interaction frequencies for normalization and genomic bins with low coverage were removed as described above.

Dimensionality reduction was performed on the DNA cluster matrix from SPRITE. SPRITE subclusters harboring bins located on different chromosomes with respect to each reference genomic bin were generated. SPRITE subclusters were divided based on size ranges (2, 3-32, or 33-1024 bins) to assess the role of DNA cluster size on genomic interaction preferences. Reference genomic bins observed in less than 200 subclusters were excluded. In a subset of analyses, reference genomic bins overlapping with megaloop anchors were selected. A bin ID x subcluster ID matrix was generated with each entry of the matrix containing a 1/n weight, where n represented the size of the subcluster, to ensure that a bin's weight is proportional to the number of unique bins within the subcluster. Each bin in the matrix was then normalized by the sum of weights for the bin. The bin x subcluster matrix was next subjected PCA preprocessing followed by UMAP embedding using the Monocle 3 package[64]. When considering UMAP parameters, we noted that inter-chromosomal genomic interactions between any two regions within a nuclear subcompartment only occurred in a few percent of cells, as measured in DNA FISH analyses. Imaging also revealed numerous DNA-protein clusters for each subcompartment within granule

neuron nuclei and suggest stochastic and possibly transient association between genomic loci within a subcompartment. In contrast, the chromosome identity of genomic loci strongly influenced their inter-chromosomal interaction preferences due to the packing of chromosomes within territories. To preserve global genome structure, including the sparser interactions between chromosomes, we used high n neighbors (~20% of bins) and high minimum distance (0.9) for UMAP to also preserve global topological structure for visualization. Louvain community detection was used to identify partitions, which represent groups of bins with similar genomic interaction preferences in UMAP-reduced space. The chromosome mixing index was calculated by determining the number of bins in the chromosome with the most bins in the partition, dividing this by the overall number of bins in the partition, and then transforming the ratio to a negative natural log scale. A higher chromosome mixing index reflected a broader assortment of chromosomes represented within the partition.

IGM genome structure-modeling and analysis

We applied our Integrative Genome Modeling (IGM) platform to generate a population of 1000 single cell 3D genome structures, in which the accumulated genome-wide chromatin contacts across the population are statistically consistent with the HiC experiments. We previously demonstrated that IGM structures from Hi-C data can predict with high correlation independent experimental data that map the nuclear topography of genomic regions, including data from SON and Lamin B1 TSA-seq data, lamin B1 DamID data, GPseq data, and association frequencies of genomic regions to nuclear speckles, lamina and nucleoli from DNA MERFISH data[37,38,65,66].

Structure representation. The diploid genome is segmented into chromatin regions of 100Kb DNA sequence length, each represented by a spherical domain with radius = 50.6 nm and a genome

volume occupancy of 40%, following a procedure as described in[37]. Autosomes are represented by two homologous chromatin copies, leading to a total of N = 51,907 chromatin domains in the diploid genome forming a total of 40 chromosome chains. The mouse cerebellum cell nucleus is modelled as an ellipsoid with semi-axes a = 3050 nm, b = 2350 nm, and c = 2350 nm, based on the results of our microscopy measurements.

Optimization procedure. The structure optimization is formulated as a maximum likelihood estimation problem using an iterative optimization scheme[37,65,67]. Briefly, given the HiC based contact probability matrix **A,** we reconstruct the coordinates **X** of all 1000 3D genome structures, together with a latent indicator tensor **W**, which specifies the contacts of homologous genomic regions in each individual structure of the population. We jointly optimize the structure population **X** and the contact tensor **W** by maximizing the log-likelihood of the probability $\log P(\mathbf{X}|\mathbf{A},\mathbf{W}) = \log P(\mathbf{A},\mathbf{W}|\mathbf{X})$ following our established population-based modeling approach[37]. Optimizations are initiated with random chromosome configurations.

HiC data pre-processing. We used the P22 HiC data as input information and employed our standard pipeline for preprocessing HiC data[37,38,65]. This process involved removal of regions with low Hi-C coverage (<3%), matrix balancing through KR normalization[68], conversion of normalized count to contact probabilities, and outlier removal as described in[37].

Prediction of speckle locations in single cell models. We predict the nuclear locations of nuclear speckles in each single cell model as described before[37,38]. In each single cell structure, a chromatin interaction network (CIN) is calculated for all chromatin regions ***S*** known to have a high propensity to be speckle associated. Each vertex of CIN represents a 100-kb chromatin region in ***S***. An edge between two vertices *i, j* is drawn if the corresponding chromatin regions are in physical contact in the structure. We then used the Markov Clustering Algorithm (MCL) to identify highly

connected subgraphs of the CIN for each genome structure by using the *mcl* tool in the *MCL-edge* software[69]. Speckle locations in each genome structure are then identified as the geometric centers of all chromatin regions of each subgraph containing at least 3 nodes. We previously demonstrated that this strategy produces approximate speckle locations, which reproduce experimental SON-TSA-seq data with high correlation[37,38].

Model assessment by 3D-FISH experiments. To assess the models, we compared the distributions of radial positions for 5 genomic loci from enhancer-dense, gene-dense and speckle g-DNA regions measured by 3D FISH experiments with those predicted by our models (Extended Data Fig. 5d). To measure the radial positions in the models we applied the same procedure as for the analysis of the 3D FISH imaging data. First, for each single cell structure, we projected the positions of each genomic region into the X-Y plane. Second, we measured the normalized radial position in the X-Y plane. To account for the elliptical asymmetry of the nucleus, the radial position for a genomic region *i* in a given cell is calculated as $r_i = \sqrt{\frac{x_i^2}{a^2} + \frac{y_i^2}{b^2}}$, with a, b as the semi-axis along the X-axis and Y-axis, respectively.

Structure feature calculations.

RAD: The radial position of the nucleus is calculated as the normalized distance with respect to the nuclear center (0, 0, 0). To account for the elliptical asymmetry of the nucleus, the radial position for a genomic region *i* is calculated as $r_i = \sqrt{\frac{x_i^2}{a^2} + \frac{y_i^2}{b^2} + \frac{z_i^2}{c^2}}$, with a, b and c as the semiaxes of the ellipsoid and x_i, y_i and z_i as the coordinates of genomic region *I* in a given genome structure. RAD is the average radial positions for a given genomic region over and all homologous copies and all structures in the population. RAD-SD is the standard deviation of the mean radial position across the population of genome structures.

ILF (Interior Localization Frequency): The interior localization frequency (ILF) of a genomic region *I* is calculated as $ILF_I = \frac{n_{r<0.5}}{M=1000}$, where $n_{r<0.5}$ is the number of structures where either one of the two homologous copies of *I* has a radial position lower than 0.5,; *M* is the total number of genome structures in the population.

LAF (Lamina association frequency): The lamina association frequency of a genomic region *I* is calculated as $LAF_I = \frac{n_{r_i>r_t}+n_{r_{i'}>r_t}}{2M}$, where $M = 1000$ is the number of structures in the population; $n_{r_i>r_t}$ and $n_{r_{i'}>r_t}$ are the number of structures, in which region i and its homologous copy i' have a radial position r_i larger than the association threshold, $r_t = 0.65xR_{nuc}$.

SpD (Mean distance to the nearest speckle): For a genomic region, SpD is the distance to the closest nuclear speckle, averaged over the population of all cells.

SpD-SD: is the standard deviation of SpD across the population of cells.

SAF (Speckle association frequency): The speckle association frequency of a genomic region *I* is calculated as $SAF_I = \frac{n_{d_i<d_t}+n_{d_{i'}<d_t}}{2M}$, where $M = 1000$ is the number of structures in the population; $n_{d_i<d_t}$ and $n_{d_{i'}<d_t}$ are the number of structures, in which region i and its homologous copy i' have a bead surface – nearest speckle distance smaller than the threshold, $d_t = 500$ nm.

NuD (mean distance to the nearest nucleolus): NuD is the distance of a genomic region to the center of the closest nucleolus, averaged over the population of cells. NuD-SD is the standard deviation of the mean nucleolus distance):

NAF (Nucleolus association frequency): The nucleolus association frequency of a genomic region *I* is calculated as $NAF_I = \frac{n_{d_i<d_t}+n_{d_{i'}<d_t}}{2M}$, where $M = 1000$ is the number of structures in the population; $n_{d_i<d_t}$ ($n_{d_{i'}<d_t}$) is the number of structures, in which region i (its homologous copy i') have a surface distance to the nearest nucleolus smaller than the threshold, $d_t = 1000$ nm.

RG: (mean radius of gyration of local chromatin fiber). The local compaction of the chromatin fiber at a given locus is calculated by the radius of gyration (RG) for a 1 Mb region centered at the locus *i* (i.e. +500kb up- and 500 kb downstream of *i*), defined as $RG_i = \frac{\sum_{m=1}^{M}\sum_{j=1}^{N} {d_j}^2}{M}$, with N as the number of chromatin regions in the 1-Mb window, and d_j is the distance between the chromatin region *j* to the center of mass of the 1-Mb region, *M*=1000 is structures *m* in the population. RG-SD is the standard deviation of the mean RG across the population of genome structures.

IPP (mean inter-chromosomal proximity probability): For each genomic region i, we define $E_i = \{j: j \neq i, d_{ij} < 500\ nm\}$ as the set of genomic regions $\{j\}$ with center-to-center distances to *i* to be within 500 nm. $n_i^{intra}(m)$ and $n_i^{inter}(m)$ are the number of intra- and inter-chromosomal regions in the set E_i in structure m. IPP is then calculated as: $IPP_I = \frac{1}{2M}\sum_{m=1}^{M}\frac{n_i^{inter}(m)}{n_i^{inter}(m)+n_i^{intra}(m)} + \frac{n_{i'}^{inter}(m)}{n_{i'}^{inter}(m)+n_{i'}^{intra}(m)}$, where *M* =1000 is the number of structures, *i* and *i*' are the two homologous copies of genomic region *I*.

TransAB: For a genomic region I, the trans A/B ratio is defined as $transAB_I = \frac{1}{2M}\sum_{m=1}^{M}\frac{n_i^A}{n_i^B} + \frac{n_{i'}^A}{n_{i'}^B}$ with n_i^A and n_i^B ($n_{i'}^A$ and $n_{i'}^B$) as the number genomic regions in the A or B compartment in the trans neighborhood set of genomic region *i* (and its homologous copy *i*') defined as $E_i^t = \{j: chrom_i \neq chrom_j, d_{ij} < 500\ nm\}$. E_i^t is the set of inter-chromosomal genomic regions within a distance of 500 nm to genomic region *i*. The mean of the transA/B ratios for a region is then calculated as the average of transA/B ratios from both homologous copies over all structures of the population. The values are rescaled to range between 0 – 1.

Fold change analysis of structure features. For each genomic region in each active subcompartment (enh-d, gene-d, speckle), we calculate all 14 structural features. We then calculate the log2-fold change of feature values in each subcompartment group with those from a corresponding randomly selected set of genomic regions (The fold change is the average value resulting from 100 sets of randomly selected genomic regions).

Visualization of 3D structures was performed with Chimera Chimera[70].

1 Yap, E. L. & Greenberg, M. E. Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. *Neuron* **100**, 330-348 (2018). https://doi.org:10.1016/j.neuron.2018.10.013
2 Shibata, M., Gulden, F. O. & Sestan, N. From trans to cis: transcriptional regulatory networks in neocortical development. *Trends Genet* **31**, 77-87 (2015). https://doi.org:10.1016/j.tig.2014.12.004
3 Farini, D., Marazziti, D., Geloso, M. C. & Sette, C. Transcriptome programs involved in the development and structure of the cerebellum. *Cell Mol Life Sci* **78**, 6431-6451 (2021). https://doi.org:10.1007/s00018-021-03911-w
4 Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490-503 (2013). https://doi.org:10.1016/j.cell.2013.07.011
5 Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet* **15**, 69-81 (2014). https://doi.org:10.1038/nrg3623
6 Weake, V. M. & Workman, J. L. Inducible gene expression: diverse regulatory mechanisms. *Nat Rev Genet* **11**, 426-437 (2010). https://doi.org:10.1038/nrg2781
7 Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403-413 (2006). https://doi.org:10.1016/j.cell.2006.06.035
8 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324 (2017). https://doi.org:10.1016/j.cell.2017.09.026
9 Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519-524 (2017). https://doi.org:10.1038/nature21411
10 Monahan, K. *et al.* Cooperative interactions enable singular olfactory receptor expression in mouse olfactory neurons. *Elife* **6** (2017). https://doi.org:10.7554/eLife.28620
11 Zhang, S., Ubelmesser, N., Barbieri, M. & Papantonis, A. Enhancer-promoter contact formation requires RNAPII and antagonizes loop extrusion. *Nature genetics* **55**, 832-840 (2023). https://doi.org:10.1038/s41588-023-01364-4
12 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014). https://doi.org:10.1016/j.cell.2014.11.021
13 Yamada, T. *et al.* Sensory experience remodels genome architecture in neural circuit to drive motor learning. *Nature* **569**, 708-713 (2019). https://doi.org:10.1038/s41586-019-1190-7
14 Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757 e724 (2018). https://doi.org:10.1016/j.cell.2018.05.024
15 Takei, Y. *et al.* Single-cell nuclear architecture across cell types in the mouse brain. *Science* **374**, 586-594 (2021). https://doi.org:10.1126/science.abj1966
16 Tan, L., Xing, D., Daley, N. & Xie, X. S. Three-dimensional genome structures of single sensory neurons in mouse visual and olfactory systems. *Nat Struct Mol Biol* **26**, 297-307 (2019). https://doi.org:10.1038/s41594-019-0205-2

17 Monahan, K., Horta, A. & Lomvardas, S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* **565**, 448-453 (2019). https://doi.org:10.1038/s41586-018-0845-0
18 Tan, L. *et al.* Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* **184**, 741-758 e717 (2021). https://doi.org:10.1016/j.cell.2020.12.032
19 Winick-Ng, W. *et al.* Cell-type specialization is encoded by specific chromatin topologies. *Nature* **599**, 684-691 (2021). https://doi.org:10.1038/s41586-021-04081-2
20 Li, C. H. *et al.* MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature* **586**, 440-444 (2020). https://doi.org:10.1038/s41586-020-2574-4
21 Zylka, M. J., Simon, J. M. & Philpot, B. D. Gene length matters in neurons. *Neuron* **86**, 353-355 (2015). https://doi.org:10.1016/j.neuron.2015.03.059
22 Gabel, H. W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89-93 (2015). https://doi.org:10.1038/nature14319
23 Closser, M. *et al.* An expansion of the non-coding genome and its regulatory potential underlies vertebrate neuronal diversity. *Neuron* **110**, 70-85 e76 (2022). https://doi.org:10.1016/j.neuron.2021.10.014
24 Nord, A. S. & West, A. E. Neurobiological functions of transcriptional enhancers. *Nat Neurosci* **23**, 5-14 (2020). https://doi.org:10.1038/s41593-019-0538-5
25 Beagan, J. A. *et al.* Three-dimensional genome restructuring across timescales of activity-induced neuronal gene expression. *Nat Neurosci* **23**, 707-717 (2020). https://doi.org:10.1038/s41593-020-0634-6
26 Wagner, M. J. & Luo, L. Neocortex-Cerebellum Circuits for Cognitive Processing. *Trends in neurosciences* **43**, 42-54 (2020). https://doi.org:10.1016/j.tins.2019.11.002
27 Hull, C. & Regehr, W. G. The Cerebellar Cortex. *Annual review of neuroscience* **45**, 151-175 (2022). https://doi.org:10.1146/annurev-neuro-091421-125115
28 King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B. & Diedrichsen, J. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nat Neurosci* **22**, 1371-1378 (2019). https://doi.org:10.1038/s41593-019-0436-x
29 Witter, L. & De Zeeuw, C. I. Regional functionality of the cerebellum. *Curr Opin Neurobiol* **33**, 150-155 (2015). https://doi.org:10.1016/j.conb.2015.03.017
30 Kozareva, V. *et al.* A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. *Nature* **598**, 214-219 (2021). https://doi.org:10.1038/s41586-021-03220-z
31 Lee, J. R. Protein tyrosine phosphatase PTPRT as a regulator of synaptic formation and neuronal development. *BMB Rep* **48**, 249-255 (2015). https://doi.org:10.5483/bmbrep.2015.48.5.037
32 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319 (2013). https://doi.org:10.1016/j.cell.2013.03.035
33 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947 (2013). https://doi.org:10.1016/j.cell.2013.09.053
34 Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344-350 (2021). https://doi.org:10.1038/s41586-020-03126-2

35 Galganski, L., Urbanek, M. O. & Krzyzosiak, W. J. Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res* **45**, 10350-10368 (2017). https://doi.org:10.1093/nar/gkx759
36 Fiorillo, L. *et al.* Comparison of the Hi-C, GAM and SPRITE methods using polymer models of chromatin. *Nat Methods* **18**, 482-490 (2021). https://doi.org:10.1038/s41592-021-01135-1
37 Boninsegna, L. *et al.* Integrative genome modeling platform reveals essentiality of rare contact events in 3D genome organizations. *Nat Methods* **19**, 938-949 (2022). https://doi.org:10.1038/s41592-022-01527-x
38 Yildirim, A., Hua, N., Boninsegna, L., Polles, G., Gong, K., Hao, S., Li, W., Zhou, X. J., Alber, F. Population-based structure modeling reveals key roles of nuclear microenviroment in gene functions. *bioRxiv* (2022). https://doi.org:10.1101/2021.07.11.451976
39 Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361** (2018). https://doi.org:10.1126/science.aar3958
40 Oudelaar, A. M. & Higgs, D. R. The relationship between genome structure and function. *Nat Rev Genet* **22**, 154-168 (2021). https://doi.org:10.1038/s41576-020-00303-x
41 Xie, L. *et al.* BRD2 compartmentalizes the accessible genome. *Nature genetics* **54**, 481-491 (2022). https://doi.org:10.1038/s41588-022-01044-9
42 Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855 e1816 (2018). https://doi.org:10.1016/j.cell.2018.10.042
43 Leidescher, S. *et al.* Spatial organization of transcribed eukaryotic genes. *Nat Cell Biol* **24**, 327-339 (2022). https://doi.org:10.1038/s41556-022-00847-6
44 Bhat, P. *et al.* 3D genome organization around nuclear speckles drives mRNA splicing efficiency. *bioRxiv* (2023). https://doi.org:10.1101/2023.01.04.522632
45 Rizwan, A. P., Zhan, X., Zamponi, G. W. & Turner, R. W. Long-Term Potentiation at the Mossy Fiber-Granule Cell Relay Invokes Postsynaptic Second-Messenger Regulation of Kv4 Channels. *J Neurosci* **36**, 11196-11207 (2016). https://doi.org:10.1523/JNEUROSCI.2051-16.2016
46 Heath, N. C. *et al.* The expression pattern of a Cav3-Kv4 complex differentially regulates spike output in cerebellar granule cells. *J Neurosci* **34**, 8800-8812 (2014). https://doi.org:10.1523/JNEUROSCI.0981-14.2014
47 Witter, L. & De Zeeuw, C. I. In vivo differences in inputs and spiking between neurons in lobules VI/VII of neocerebellum and lobule X of archaeocerebellum. *Cerebellum* **14**, 506-515 (2015). https://doi.org:10.1007/s12311-015-0654-z
48 Geschwind, D. H. & State, M. W. Gene hunting in autism spectrum disorder: on the path to precision medicine. *Lancet Neurol* **14**, 1109-1120 (2015). https://doi.org:10.1016/S1474-4422(15)00044-7
49 Yamada, T. *et al.* Promoter decommissioning by the NuRD chromatin remodeling complex triggers synaptic connectivity in the mammalian brain. *Neuron* **83**, 122-134 (2014). https://doi.org:S0896-6273(14)00483-8 [pii]
10.1016/j.neuron.2014.05.039
50 Yang, Y. *et al.* Chromatin remodeling inactivates activity genes and regulates neural coding. *Science* **353**, 300-305 (2016). https://doi.org:10.1126/science.aad4225

51 Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods* **12**, 963-965 (2015). https://doi.org:10.1038/nmeth.3542
52 Bolland, D. J., King, M. R., Reik, W., Corcoran, A. E. & Krueger, C. Robust 3D DNA FISH using directly labeled probes. *J Vis Exp* (2013). https://doi.org:10.3791/50587
53 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010). https://doi.org:10.1093/bioinformatics/btp616
54 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021). https://doi.org:10.1016/j.cell.2021.04.048
55 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196 (2016). https://doi.org:10.1126/science.aad0501
56 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008). https://doi.org:10.1186/gb-2008-9-9-r137
57 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259 (2015). https://doi.org:10.1186/s13059-015-0831-x
58 Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572 e524 (2017). https://doi.org:10.1016/j.cell.2017.09.043
59 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016). https://doi.org:10.1016/j.cels.2016.07.002
60 Saylor, J. E., Sundell, K. E. & Sharman, G. R. Characterizing sediment sources by non-negative matrix factorization of detrital geochronological data. *Earth Planet Sc Lett* **512**, 46-58 (2019). https://doi.org:10.1016/j.epsl.2019.01.044
61 Dennis, G., Jr. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).
62 Xu, J. *et al.* Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature* **611**, 387-398 (2022). https://doi.org:10.1038/s41586-022-05365-x
63 Zhang, R. & Ma, J. MATCHA: Probing multi-way chromatin interaction with hypergraph representation learning. *Cell Syst* **10**, 397-407 e395 (2020). https://doi.org:10.1016/j.cels.2020.04.004
64 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019). https://doi.org:10.1038/s41586-019-0969-x
65 Hua, N. *et al.* Producing genome structure populations with the dynamic and automated PGS software. *Nat Protoc* **13**, 915-926 (2018). https://doi.org:10.1038/nprot.2018.008
66 Zhu, Y. *et al.* Comprehensive characterization of neutrophil genome topology. *Genes Dev* **31**, 141-153 (2017). https://doi.org:10.1101/gad.293910.116
67 Tjong, H. *et al.* Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc Natl Acad Sci U S A* **113**, E1663-1672 (2016). https://doi.org:10.1073/pnas.1512577113
68 Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9**, 999-1003 (2012). https://doi.org:10.1038/nmeth.2148
69 Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**, 1575-1584 (2002). https://doi.org:10.1093/nar/30.7.1575
70 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004). https://doi.org:10.1002/jcc.20084

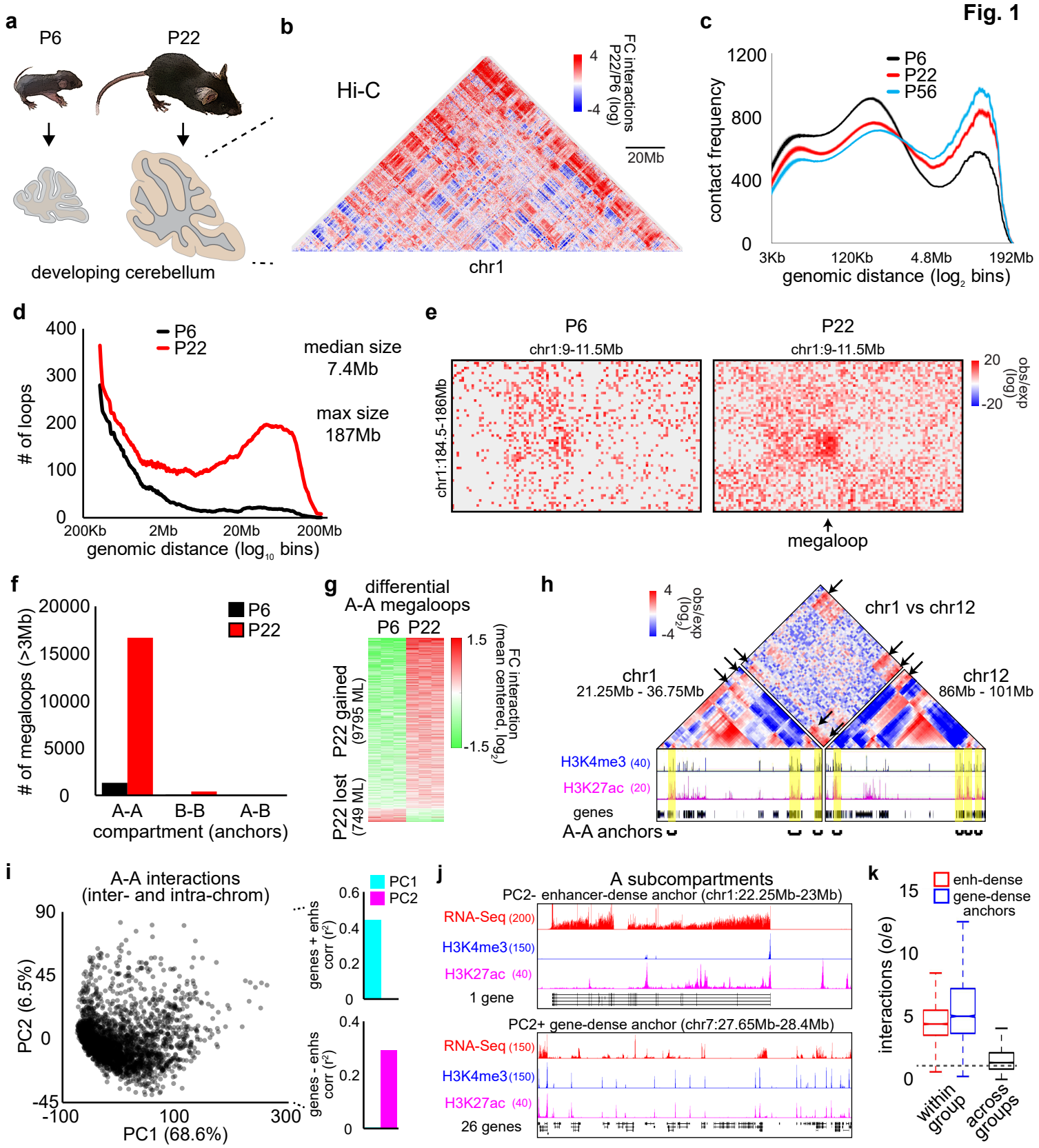
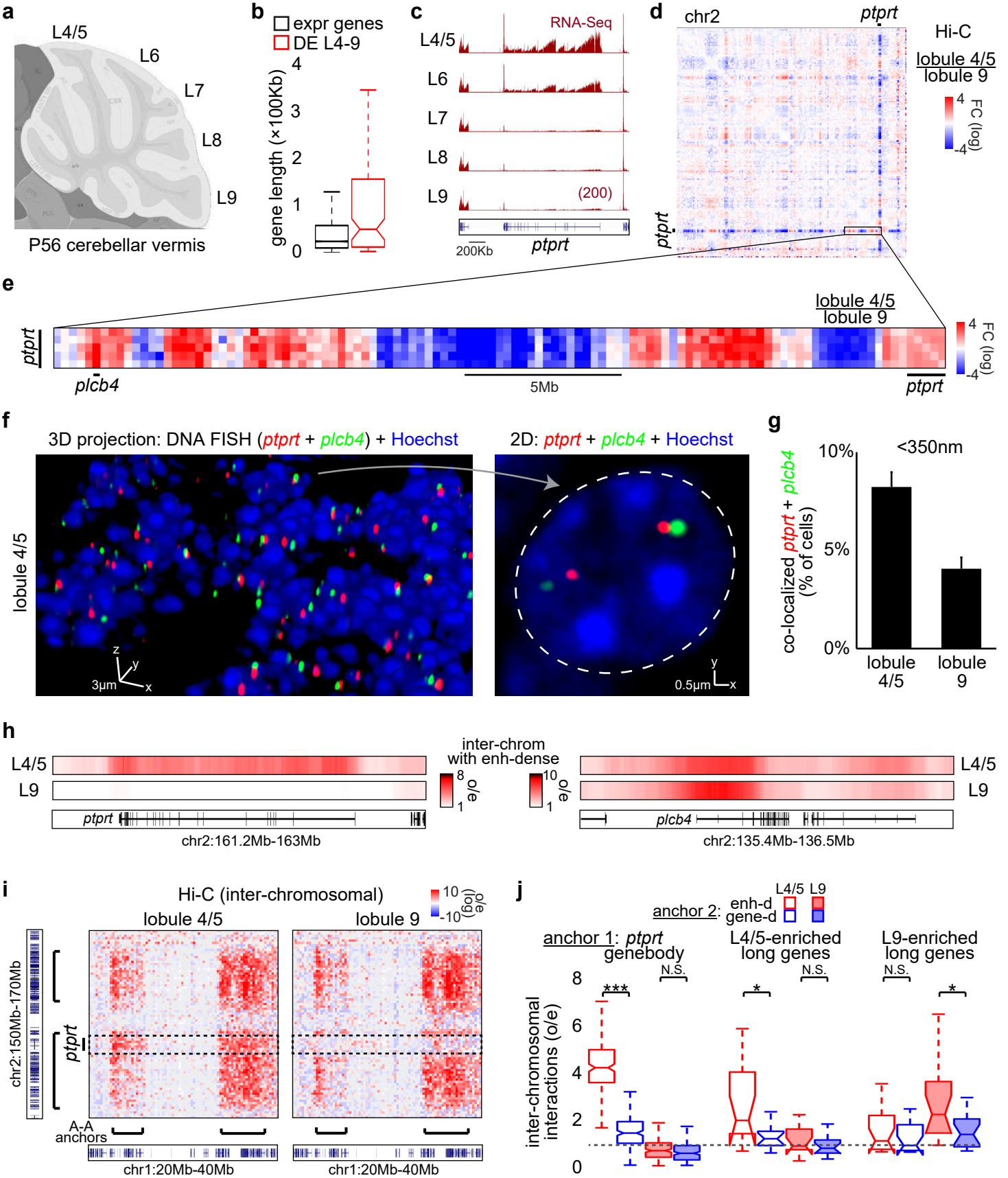


Fig. 1
a
P6
P22
developing cerebellum
b
Hi-C
FC interactions P22/P6 (log)
4
-4
20Mb
chr1
c
contact frequency
1200
800
400
0
P6
P22
P56
3Kb
120Kb
4.8Mb
192Mb
genomic distance (log2 bins)
d
# of loops
400
300
200
100
0
P6
P22
median size 7.4Mb
max size 187Mb
200Kb
2Mb
20Mb
200Mb
genomic distance (log10 bins)
e
P6
chr1:9-11.5Mb
P22
chr1:9-11.5Mb
chr1:184.5-186Mb
obs/exp (log2)
20
-20
megaloop
f
# of megaloops (>3Mb)
20000
15000
10000
5000
0
P6
P22
A-A
B-B
A-B
compartment (anchors)
g
differential A-A megaloops
P6 P22
P22 gained (9795 ML)
P22 lost (749 ML)
FC interaction (mean centered, log2)
1.5
-1.5
h
chr1 vs chr12
obs/exp (log2)
4
-4
chr1
21.25Mb - 36.75Mb
chr12
86Mb - 101Mb
H3K4me3 (40)
H3K27ac (20)
genes
A-A anchors
i
A-A interactions (inter- and intra-chrom)
PC2 (6.5%)
90
45
0
-45
-100
100
300
PC1 (68.6%)
PC1
PC2
genes + enhs corr (r2)
0.6
0.4
0.2
0
genes - enhs corr (r2)
0.4
0.2
0
j
A subcompartments
PC2- enhancer-dense anchor (chr1:22.25Mb-23Mb)
RNA-Seq (200)
H3K4me3 (150)
H3K27ac (40)
1 gene
PC2+ gene-dense anchor (chr7:27.65Mb-28.4Mb)
RNA-Seq (150)
H3K4me3 (150)
H3K27ac (40)
26 genes
k
enh-dense
gene-dense anchors
interactions (o/e)
15
10
5
0
within group
across groups
l
gene ontology
enh-dense anchors
score | term | Benjamini
4.74 | synapse | 4.87E-04
3.85 | pleckstrin homology-like domain | 8.05E-03
gene-dense anchors
score | term | Benjamini
8.65 | zinc finger DNA-binding domain | 9.07E-12
7.02 | MHC class I, alpha chain | 9.30E-11

Fig. 2


a
L4/5
L6
L7
L8
L9
P56 cerebellar vermis
b
expr genes
DE L4-9
gene length (×100Kb)
c
RNA-Seq
L4/5
L6
L7
L8
L9
(200)
200Kb
ptprt
d
chr2
ptprt
Hi-C
lobule 4/5
lobule 9
FC (log)
e
ptprt
plcb4
5Mb
f
3D projection: DNA FISH (ptprt + plcb4) + Hoechst
2D: ptprt + plcb4 + Hoechst
lobule 4/5
3μm
0.5μm
g
<350nm
co-localized ptprt + plcb4 (% of cells)
lobule 4/5
lobule 9
h
L4/5
L9
inter-chrom with enh-dense
o/e
ptprt
chr2:161.2Mb-163Mb
plcb4
chr2:135.4Mb-136.5Mb
i
Hi-C (inter-chromosomal)
lobule 4/5
lobule 9
chr2:150Mb-170Mb
A-A anchors
chr1:20Mb-40Mb
j
L4/5 L9
anchor 2: enh-d gene-d
anchor 1: ptprt genebody
L4/5-enriched long genes
L9-enriched long genes
N.S.
inter-chromosomal interactions (o/e)

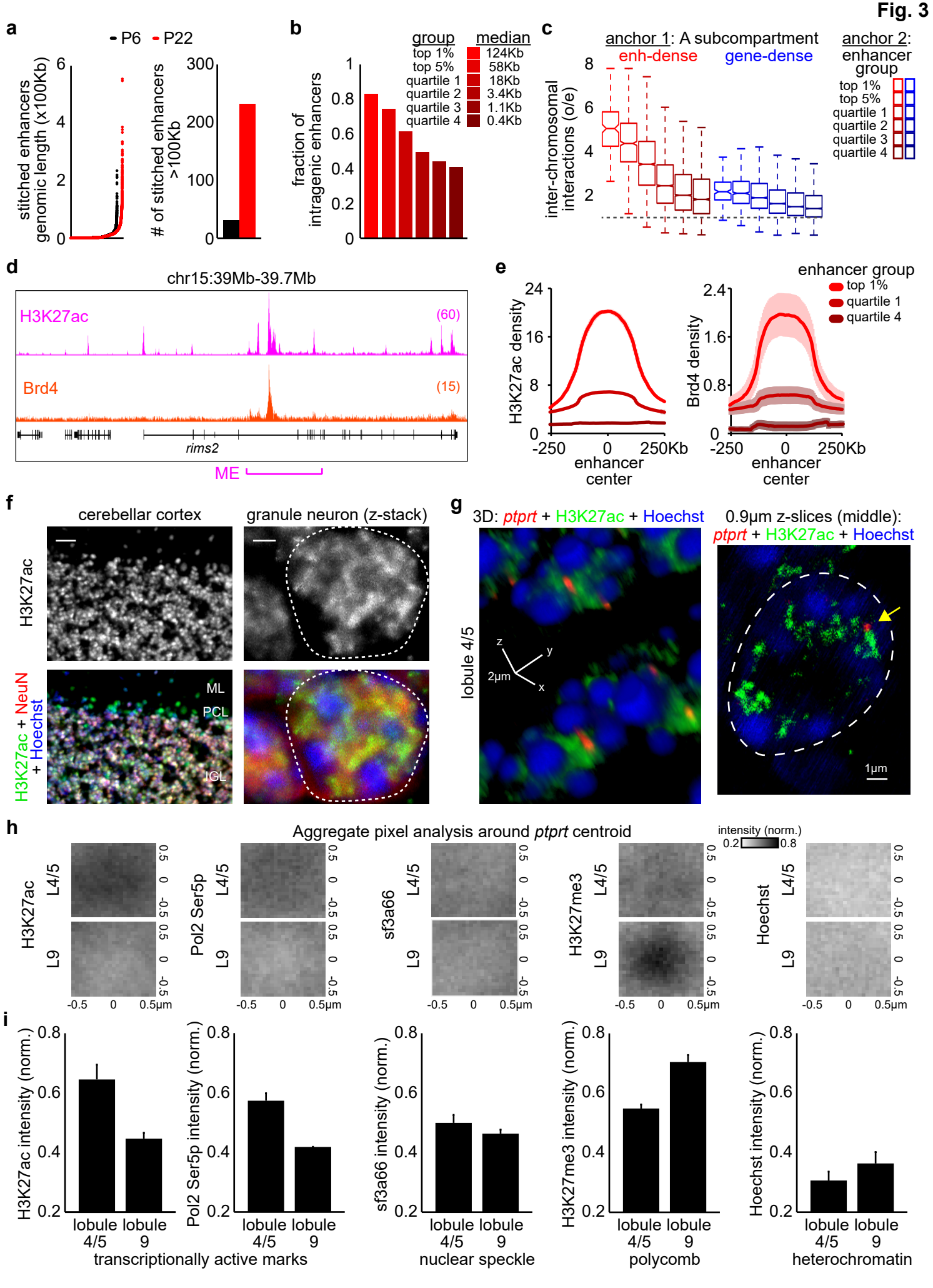


Fig. 3
a
P6
P22
stitched enhancers genomic length (x100Kb)
# of stitched enhancers >100Kb
b
group
median
top 1%
124Kb
top 5%
58Kb
quartile 1
18Kb
quartile 2
3.4Kb
quartile 3
1.1Kb
quartile 4
0.4Kb
fraction of intragenic enhancers
c
anchor 1: A subcompartment
enh-dense
gene-dense
anchor 2: enhancer group
top 1%
top 5%
quartile 1
quartile 2
quartile 3
quartile 4
inter-chromosomal interactions (o/e)
d
chr15:39Mb-39.7Mb
H3K27ac
(60)
Brd4
(15)
rims2
ME
e
enhancer group
top 1%
quartile 1
quartile 4
H3K27ac density
Brd4 density
enhancer center
-250
0
250Kb
f
cerebellar cortex
granule neuron (z-stack)
H3K27ac
H3K27ac + NeuN + Hoechst
ML
PCL
IGL
g
3D: ptprt + H3K27ac + Hoechst
0.9µm z-slices (middle): ptprt + H3K27ac + Hoechst
lobule 4/5
z
y
x
2µm
1µm
h
Aggregate pixel analysis around ptprt centroid
intensity (norm.)
0.2
0.8
H3K27ac
Pol2 Ser5p
sf3a66
H3K27me3
Hoechst
L4/5
L9
-0.5
0
0.5µm
i
H3K27ac intensity (norm.)
Pol2 Ser5p intensity (norm.)
sf3a66 intensity (norm.)
H3K27me3 intensity (norm.)
Hoechst intensity (norm.)
lobule 4/5
lobule 9
transcriptionally active marks
nuclear speckle
polycomb
heterochromatin

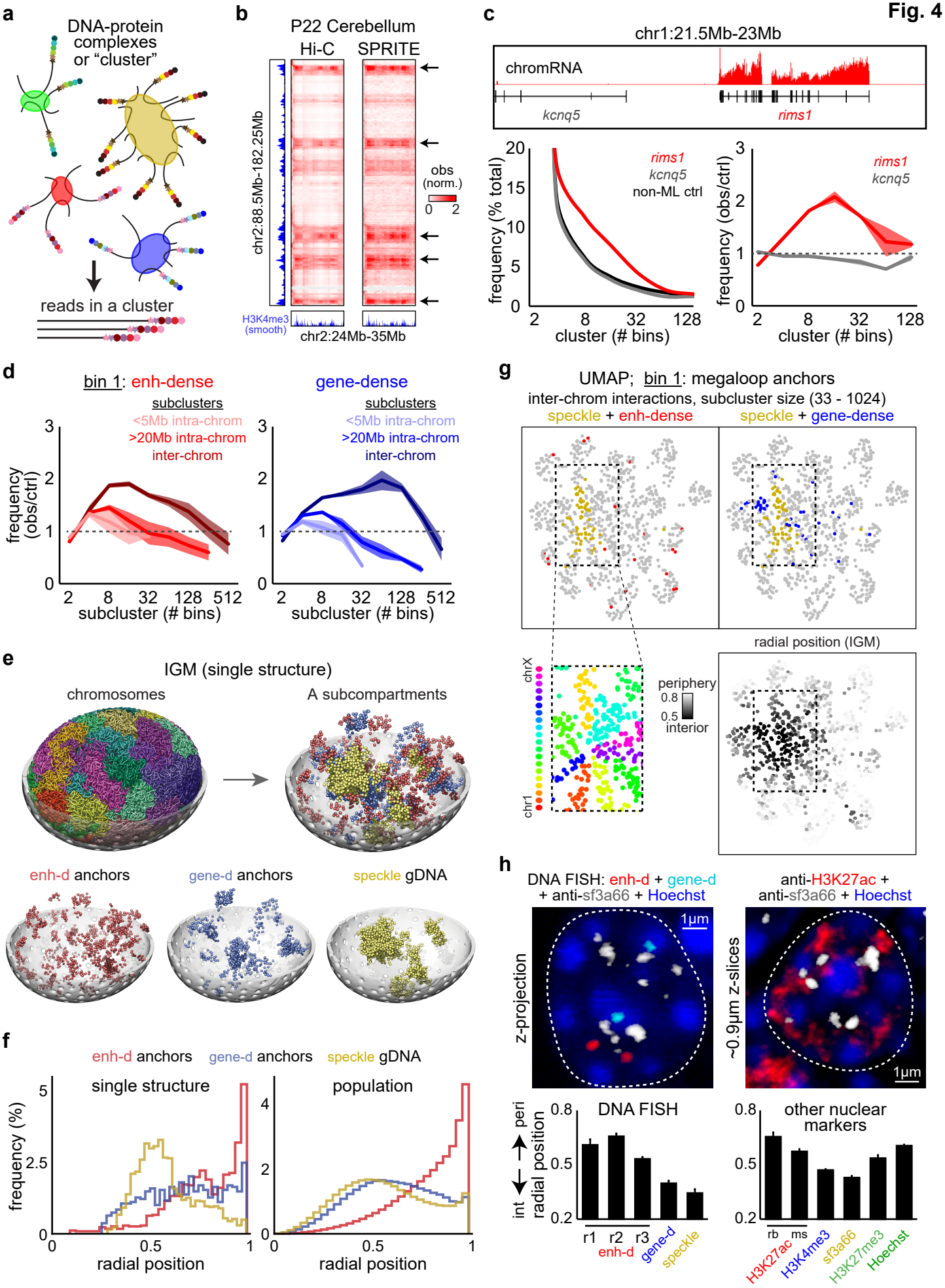


Fig. 4
a
DNA-protein complexes or "cluster"
reads in a cluster
b
P22 Cerebellum
Hi-C
SPRITE
chr2:88.5Mb-182.25Mb
obs (norm.)
0 2
H3K4me3 (smooth)
chr2:24Mb-35Mb
c
chr1:21.5Mb-23Mb
chromRNA
kcnq5
rims1
frequency (% total)
rims1
kcnq5
non-ML ctrl
cluster (# bins)
frequency (obs/ctrl)
rims1
kcnq5
cluster (# bins)
d
bin 1: enh-dense
subclusters
<5Mb intra-chrom
>20Mb intra-chrom
inter-chrom
frequency (obs/ctrl)
subcluster (# bins)
gene-dense
subclusters
<5Mb intra-chrom
>20Mb intra-chrom
inter-chrom
subcluster (# bins)
e
IGM (single structure)
chromosomes
A subcompartments
enh-d anchors
gene-d anchors
speckle gDNA
f
enh-d anchors
gene-d anchors
speckle gDNA
single structure
population
frequency (%)
radial position
radial position
g
UMAP; bin 1: megaloop anchors
inter-chrom interactions, subcluster size (33 - 1024)
speckle + enh-dense
speckle + gene-dense
chrX
chr1
periphery
0.8
0.5
interior
radial position (IGM)
h
DNA FISH: enh-d + gene-d + anti-sf3a66 + Hoechst
anti-H3K27ac + anti-sf3a66 + Hoechst
z-projection
~0.9µm z-slices
1µm
1µm
int
peri
radial position
DNA FISH
r1
r2
r3
enh-d
gene-d
speckle
other nuclear markers
rb
ms
H3K27ac
H3K4me3
sf3a66
H3K27me3
Hoechst