Exploring Cell Diversity in Complex Tissues through Spatial Genomics and Spatial Transcriptomics

Thesis by

Yujing Yang

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2025

(Defended May 23, 2025)

Yujing Yang
ORCID: 0000-0002-2338-6263

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Professor Long Cai. His high standards and incisive feedback pushed me to solve difficult problems and, through that process, helped me become a more independent and persistent scientist. It has been a privilege to work in this lab: I had the opportunity to apply cutting-edge spatial genomics and transcriptomics tools to turn abstract ideas into solid discoveries. The journey was full of failures, but those very challenges made each breakthrough more rewarding. One of the most valuable lessons I've learned in graduate school is that it's not failed experiments or analyses that hold you back, but the fear of failure itself.

I would also like to thank my committee, Professors Mitchell Guttman, David Van Valen, and Matt Thomson, for their constructive feedback throughout my graduate studies. Their questions pushed me to think more deeply, and their suggestions often led to creative, out-of-the-box solutions. I am especially grateful to Mitchell Guttman for his expertise in sequencing-based methods and for the valuable suggestions he provided during our manuscript revision.

My sincere thanks go to Professor Andrew McMahon, who taught me to approach research from a methods perspective while keeping the focus on asking meaningful biological questions. I am also grateful to Dr. Yodai Takei, from whom I learned the importance of thoughtful experimental design and rigorous validation. A special thanks goes to Professor Phillip Zamore, my first mentor, who encouraged me to stay genuine to myself and to follow my passion, especially during difficult times.

I've been fortunate to work with many talented colleagues in the Cai Lab. Katsuya "Lex" Colon, who developed the PyFISH pipeline and is the only reason I could decode all my experiments, and who was always ready to help and always brought brilliant ideas. I truly enjoyed late-night scientific conversations with my talented cohort, Arun Chakravorty, especially during long incubation hours. I'm grateful to Dr. Michal Polonsky for guiding me during my rotation, and to Drs. Carsten Tischbirek, Elsy Buitrago-Delgado, and Sheel Shah for our fun , often philosophical discussions about science and life. I also want to thank our

lab manager, Jina Yun, who kept the lab running smoothly, and kindly kept me fed with snacks. Thanks as well to Christopher Cronin, for maintaining the automation system and for saving me from many sleepless nights in the lab. I'm equally thankful to all current and former members of the Cai Lab, Konnor von Emster, Linjing (Lynn) Fang, Jonathan Fox, Chuqi Lu, Alexi McAdams, Jonathan White, Saori Lobbia, Lincoln Ombelets, Chee Huat (Linus) Eng, Anthony Linares, and Simone Schindler, for creating such a collaborative, kind, and inspiring research environment.

Graduate school was challenging, and I could not have made it through without the support of my friends, Yutian Li and Meichen Fang. We celebrated every success, supported each other through difficulties, and hunted for food and boba together. I also thank our neighbor Tracy and her golden retriever, Beacon, the world's cutest and most professional therapy dog, whose fluff and cuteness always warmed my heart.

Finally, and most importantly, I thank my family. Their unconditional love, understanding, and encouragement have shaped my values, taught me to be resilient, and reminded me to stay kind, work hard, and keep chasing my dreams.

ABSTRACT

The study of cellular diversity is a fundamental requirement for understanding how multicellular organisms function. During the development of multicellular organisms, cells differentiate into various cell types with different molecular compositions, exhibit different phenotypes, and show distinct morphologies. Each single cell occupies a specific spatial location within different tissues and organs and performs a unique function. A holistic understanding of cells requires the integration of multiple "omics" modalities, including genomics, epigenomics, transcriptomics, and proteomics. Current well-established single-cell sequencing methods have been used to build enormous single-cell transcriptomic atlases. While single-cell sequencing methods are now capable of multi-omic profiling, they all require cell dissociation, during which important spatial context information is lost. To study cellular diversity within its native spatial context, our lab has developed innovative spatial genomics and transcriptomics tools that enable multi-omics profiling at single-cell resolution while preserving intact tissue organization. This thesis presents two projects that leverage these tools to investigate cellular diversity in complex tissues across different biological scales, from subnuclear to tissue-level organization. In Chapter 2, we applied spatial multi-omics to the mouse cerebellum, achieving single-cell resolution profiling of 100,049 genomic loci, 17,856 nascent transcripts, 60 mature mRNAs, and 28 immunofluorescently labeled subnuclear structures. To achieve this, we developed innovative two-layer barcodes for DNA sequential fluorescence *in situ* hybridization (seqFISH). Combining cell-type information from nascent and mature transcriptomes, we captured the three-dimensional genomic architecture and its interactions with subnuclear compartments in a cell-type-specific manner. Our findings show that repressive chromatin compartments have greater cell-type specificity than active chromatin compartments in the mouse cerebellum. In Chapter 3, we integrated single-cell multiome sequencing, which profiles single-nucleus RNA and chromatin accessibility (ATAC) from the same cells, with seqFISH spatial transcriptomics. This approach was applied to the 17- to 18-week-old human fetal kidney, targeting 224 marker genes. By combining sequencing and spatial profiling data, we constructed a comprehensive developmental atlas of human kidney organogenesis, providing

new insights into the tissue organization and gene expression patterns during kidney development.

PUBLISHED CONTENT AND CONTRIBUTIONS

*Takei Y, *Yang Y, White J, Goronzy IN, Yun J, Prasad M, Ombelets LJ, Schindler S, Bhat P, Guttman M, Cai L. Spatial multi-omics reveals cell-type-specific nuclear compartments. Nature. 2025 Apr 9. doi: 10.1038/s41586-025-08838-x. *These authors contributed equally.

Y.Y. is a co-first author and contributed to data analysis, data visualization and sharing, validation experiments, and manuscript writing, all with input and guidance from L.C. This chapter is adapted from the published article.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 The need for single-cell multi-omics to understand cellular heterogeneity

XA cell is the fundamental functional unit of a living organism. In multicellular organisms, cells differentiate into diverse cell types that have different molecular compositions, exhibit distinct phenotypes, and have varied morphological structures. Each cell type occupies specific spatial locations and performs distinct functions (Zeng 2022; Arendt et al. 2016). Understanding this cellular heterogeneity is essential for understanding biological systems in both healthy and pathological states. The rapid development of single-cell sequencing technologies has revolutionized this field and generated enormous single-cell atlases of various organs and tissues under healthy and disease conditions (Kozareva et al. 2021; Hodge et al. 2019; Tabula Sapiens Consortium* et al. 2022; Lindström et al. 2018; Rozenblatt-Rosen et al. 2021; Tang et al. 2009; Shalek et al. 2013). However, most studies focus on single-cell transcriptomics, where a cell type is defined purely by gene expression patterns. It is important to develop a multi-omics method that can simultaneously profile the different molecular regulators in the gene regulatory network including the genomic DNA, epigenetic modifications, gene expression, and protein production (Badia-I-Mompel et al. 2023; Levine & Davidson 2005; Zhu et al. 2020). In addition, it is crucial to profile cells within their native tissue context to understand their responses to environmental stimuli. This need has motivated our laboratory to develop innovative spatial genomic and transcriptomic profiling tools.

In the introduction, I discuss recent advances in sequencing and spatial multi-omics technologies. In Chapter 2, I present our work on the detailed profiling of chromatin organization in the mouse cerebellum in a cell-type-specific manner. In Chapter 3, I present our study investigating the tissue organization of the human fetal kidney by integrating single-cell multiome sequencing and spatial transcriptomics.

1.2 Sequencing-based multi-omics technologies

Since researchers from Cambridge first performed single-cell RNA sequencing (scRNA-seq) using a next-generation sequencing platform (Tang et al. 2009), the field of single-cell sequencing has transformed how we study cell diversity. Single-cell sequencing technologies enable us to measure individual cells by their gene expression, revealing cell-type-specific features and exploring cellular heterogeneity in complex biological systems, such as brain function, organogenesis, oncogenesis, and other diseases (Lake et al. 2023; Nofech-Mozes et al. 2023; Cao et al. 2020; Braun et al. 2023; Mathys et al. 2019; Rood et al. 2025). By assigning each cell a unique barcode, researchers can track RNA, DNA and even proteins molecules from single cells using either droplet microfluidics (like 10x Genomics) (Zheng et al. 2017), or plate-based methods (SMART-seq) (Ramsköld et al. 2012; Hagemann-Jensen et al. 2020). Recent advance in this field enabled us to capture multiple types of molecules simultaneously from a single cell, acquiring multi-omic readouts. Those advances greatly enhanced our understanding of gene regulation, epigenetic modification, and protein expression dynamics. These approaches are particularly valuable in fields like lineage tracing, neuroscience, immunology, and oncology, where cellular heterogeneity plays a critical role (Baysoy et al. 2023; Nadalin et al. 2024; Xiong et al. 2024; Liu et al. 2023).

Techniques such as CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) (Stoeckius et al. 2017) and REAP-seq (RNA expression and protein sequencing assay) (Peterson et al. 2017) enable the simultaneous profiling of transcriptomic and protein expression at the single-cell level. Both methods utilize oligo-conjugated antibodies, assigning the same unique barcode to transcripts and antibody-bound oligos. These techniques have been applied in clinical and immunological research (Y. Su et al. 2020). However, both CITE-seq and REAP-seq rely on oligo conjugated antibodies, which can be limited by available antibodies and potential change of antibody specificity upon conjugation. To address these limitations, alternative methods like PHAGE-ATAC have been developed, using nanobody-presenting phages to detect epitopes, where the phage's hypervariable regions serve as barcodes (Fiskin et al. 2022).

In addition to profiling transcripts and proteins simultaneously, methods have been developed to jointly detect DNA and RNA molecules from single cells (Dey et al. 2015). The ability of sequencing DNA and RNA simultaneously from the same cells enables the integration of transcriptomics with epigenetic modifications, such as genomic DNA methylation (Lister et al. 2009; Flusberg et al. 2010), open and close chromatin status, or chromatin modifications (Angermueller et al. 2016). Methods like scNMT-seq (single-cell nucleosome, methylation, and transcription sequencing) allow simultaneous profiling of DNA methylation, nucleosome occupancy, and gene expression in single-cells (Clark et al. 2018; Cerrizuela et al. 2022). Other techniques focus on measuring chromatin accessibility together with transcriptomics. For example, SHARE-seq (Ma et al. 2020; Li et al. 2024), Paired-seq (Zhu et al. 2019), and 10X Genomics Multiome sequencing (Belhocine et al. 2021) allow jointly scATAC-seq (assay for transposase-accessible chromatin sequencing) and scRNA-seq from the same cell. Applied to hair follicle cells, SHARE-seq showed that chromatin accessibility at distal open regions of chromatin (DORCs) precedes target gene expression during differentiation, enabling the prediction of a cell's future transcriptional state and inferring its lineage trajectory from chromatin potential (Ma et al. 2020). Alternative approaches, such as Paired-Tag and epiDamID, profile histone modifications together with transcription, relating gene expression with histone mark association (Zhu et al. 2021; Rang et al. 2022). These methods collectively enhance our understanding of hierarchical structure of the gene regulation network and how epigenetic modifications regulate gene expression. Recent advancements have led to methods that integrate more than two modalities, such as DOGMA-seq and NEAT-seq, that can simultaneously profile gene expression, protein abundance, and chromatin accessibility (Mimitou et al. 2021), (A. F. Chen et al. 2022).

1.3 Spatial-based multi-omics technologies

Cells must be dissociated from their native tissue environment before going through single-cell sequencing protocols, during which spatial context information is lost (Longo et al. 2021). In contrast, imaging-based spatial technologies can perform single-cell multi-omics

profiling while preserving spatial information, as they typically involve repeated imaging of the same region from the same cell on the same slides, allowing sequential detection of different molecules (Marx 2021).

Imaging-based spatial technologies are built upon in situ hybridization (ISH), where complementary strand of nucleic acid probes are hybridized to a specific segment DNA or RNA target within a histologic section (Anon n.d.). To enable multiplexed target detection, instead of directly readout the target, barcodes attached on primary probes that hybridize to target RNA or DNA species were detected (Lubeck et al. 2014; Chen et al. 2015; Shah et al. 2017; Xia et al. 2019; J.-H. Su et al. 2020). The field has diverged into two directions for decoding barcodes. The first direction is based on Fluorescence In Situ Hybridization (FISH). Each primary probe contains overhangs with four or five barcode binding sites, each site is 15-20 nucleotides long. Fluorescent readouts hybridize to the barcode binding sites, are imaged, and are removed during sequential imaging rounds, utilizing multiple rounds of hybridizations to decode barcodes. Our laboratory developed sequential fluorescent *in situ* hybridization (seqFISH) and seqFISH+ that allow detection of 10,421 intron species (Shah et al. 2018), 10,000 mRNA transcripts from mouse brain (Eng & Others 2019), and 100,049 genomic DNA loci from mouse cerebellum (Takei et al. 2025). MERFISH, using similar strategy but different barcoding scheme, was able to barcode 10,050 transcripts in cell cultures (Xia et al. 2019). The second direction is based on *in situ* sequencing, where a short barcode sequence is attached to primary probes, and the barcode are sequenced by *in situ* sequencing such as FISSEQ (Lee et al. 2015) and STAR-map (Wang et al. 2018). These methods can be seamlessly integrated with immunofluorescence staining, to visualize both the expression levels and subcellular distribution of targeted proteins, whereas sequencing-based methods only measure protein abundance (Im et al. 2019).

A unique advantage of imaging-based spatial technologies is their ability to directly reconstruct the three-dimensional genome organization at the single-cell level by using multiplexed DNA FISH. For imaging-based multi-omics technologies, transcriptomics and genomics profiling can be easily integrated with protein measurement by sequentially apply

multiplexed RNA and DNA FISH and immunofluorescence staining. This provides us an unprecedent opportunity to understand how chromosomes are organized physically, and genome association with chromatin marks, in relation to gene expression at a cell-type-specific manner in intact complex tissues.

In 2018, our laboratory profiled 10,421 genes at their nascent transcription active sites using intron seqFISH, integrating this with chromosome painting to visualize entire chromosomes and map the transcriptionally active sites to their corresponding chromosomes (Shah et al. 2018). In 2019, Alistair's group utilized Optical Reconstruction of Chromatin Architecture (ORCA) to profile a 700-kb genomic region spanning the Bithorax Complex in Drosophila embryos at 10 kb–2 kb resolution, while simultaneously imaging 30 cell-type-specific mRNAs or nascent transcripts (Mateo et al. 2019). By leveraging cell-typing information from the marker gene expression, they identified cell-type-specific physical borders between active and Polycomb-repressed DNA, and cell-type diverse enhancer-promoter contacts at the kilobase scale. This study showed the ability of imaging-based methods to resolve fine-scale chromatin interactions in complex tissues at the single-cell level, although it was limited to a relatively small genomic region due to constraints on the total number of barcodes they can detect in single experiment. Later In 2021, Yodai Takei from our lab expanded the barcoding capacity and employed DNA seqFISH+ to profile 3,660 chromosomal loci across the mouse genome in mouse embryonic stem cells (mESCs), with 2,460 loci at 1 Mb resolution and 1,200 loci at 25-kb resolution (Takei et al. 2021). DNA seqFISH+ was integrated with RNA seqFISH probing 70 mRNAs and 17 chromatin markers using sequential immunofluorescence. This research identified many genomic loci as “fixed points” in nuclear organization: they are more frequently located the surface of nuclear bodies, and zones with combinatorial chromatin marks. The same method, applied to the complex tissue of mouse brain, revealed cell-type specific nuclear organization, its relation to gene-transcription, and how the radial organization of chromosomes are shaped (Takei et al. 2021).

To overcome the limitation of 1-Mb whole-genome resolution, we developed an innovative two-layer barcoding scheme of DNA seqFISH+, enabling the profiling of the entire mouse genome at 25-kb resolution (100,049 genomic loci), together with 18,000 nascent transcripts, 60 mRNAs, and 26 subnuclear marks (Takei et al. 2025). By combining the chromatin organization data from two-layer DNA seqFISH+ with cell-type information from RNA seqFISH, our study revealed that repressive chromatin compartments exhibit greater diversity across cell types than active chromatin compartments. Notably, a group of specific genomic loci associated with H4K20me3 marked constitutive repressive compartments in the interior of Purkinje nucleus were found to locate at the nuclear periphery, associated with Lamin, in Bergmann glial cells. Despite achieving a 25-kb resolution for DNA seqFISH+, the practical genomic resolution of chromatin profiling is limited by the diffraction-limited immunofluorescence (200–300 nm, corresponding to 100-200-kb of genomic DNA) in conventional confocal microscopy (Takei et al. 2025). Expansion microscopy has been applied to increase the resolution of histone modification immunofluorescence staining to 75nm, around 10-kb resolution (Woodworth et al. 2021; Labade et al. 2024). Additionally, Epigenomic MERFISH, which uses antibodies targeting specific histone modifications followed by *in situ* tagmentation and transcription, improves the detection of epigenetic modifications to the promoter-enhancer level (Lu et al. 2022).

In addition to imaging-base spatial technologies, sequencing technologies have been applied in spatial profiling. Instead of assigning unique barcodes to individual cells, these methods pattern slides with oligos or beads, each linked to a specific spatial location as “pixels.” For Visium and Visium HD, oligos are arrayed on a solid slide surface in a grid, with each spot containing a unique barcode identifying its position at 55 μm (Visium) (Kuppe et al. 2022) and 2 μm (Visium HD) resolution (Oliveira et al. 2024). Stereo-seq uses DNA nanoball (DNB) patterned arrays on lithographically etched chips with ultra-small 500 nm spots (A. Chen et al. 2022; Gong et al. 2025), while Slide-seq uses densely packed barcoded beads to achieve 10 μm resolution (Rodriques et al. 2019). DBiT-seq creates a spatial map by delivering two sets of molecular barcodes through perpendicular microfluidic channels, achieving up to 10 μm resolution (Liu et al. 2020). Single-cell multi-omics sequencing

methods mentioned in section 1.2 can be directly applied to sequencing-based spatial methods to enable spatial multi-omics profiling. For example, spatial ATAC-RNA seq was able to simultaneously apply ATAC-seq and RNA-seq from individual pixels from mouse and human tissue (Zhang et al. 2023). Another method, Patho-DBit, profiles all RNA types, including small RNAs like microRNAs, from FFPE tissue samples, mapping their spatial patterns, which are less explored in spatial research (Bai et al. 2024). Sequencing-based spatial technologies have several advantages: they are relatively straightforward to implement, leveraging established sequencing protocols, and usually provide comprehensive whole-transcriptome or whole-genome coverage. However, these methods cannot achieve the same subcellular spatial resolution as imaging-based approaches and can suffer from low detection efficiency and signal contamination between adjacent locations**.**

1.4 Computational methods to integrate multi-omics datasets

As rapidly developing experimental methods produce single-cell multi-omic datasets, computational tools have become essential for integrating multi-modality information and interpreting the biological meaning behind the data. Each modality provides a large number of features per cell: gene expression data include thousands to tens of thousands of gene expression levels per cell, chromatin accessibility data include hundreds of thousands of accessibility scores across the genome per cell, and protein profiling data include from several to hundreds of protein measurements per cell. These datasets are high-dimensional and noisy; different data modalities have inherently different scales, distributions, and noise levels (Tang et al. 2023). Different technology platforms introduce different technical biases (You et al. 2024; Angerer et al. 2017; Colino-Sanguino et al. 2024), complicating both the integration of modalities within single experiments and the co-analysis of datasets from different platforms

For paired experimental methods where information from different modalities is collected from the same cell or same slide barcode, integrating multi-omics datasets seems computationally straightforward because different modalities are linked through a common cell identifier. It was a common practice to analyze each single-modality dataset

independently using state-of-the-art pipelines first and then project cell typing defined from one modality to another at a pseudo-bulk level. For example, it is a common practice to identify cell types using transcriptome data first and then use these cell type labels to perform differential peak analysis on paired single-cell ATAC datasets. This is because transcriptome datasets are considered more robust for cell type prediction due to its fewer dimensions and less dropouts compared with single-cell ATAC data.

Joint analysis of paired multi-omics datasets is being pursued with advanced computational methods that integrate multiple modalities simultaneously to uncover deeper biological insights. Seurat's WNN Analysis dynamically weights the importance of each data type for individual cells, thus integrated multi-modalities despite the inherent difference of different modalities (Hao et al. 2021). MOFA+ employs a Bayesian model to uncover latent factors that capture both common and unique patterns across multiple omics layers (Argelaguet et al. 2020). In addition to above methods that focus on cell typing through joint analysis of multi-omic data, tools such as SCENIC+ (Bravo González-Blas et al. 2023) are designed specifically to infer gene-regulatory networks from paired single-cell multi-omic datasets.

Despite the rapid growth of multi-omics datasets, the scientific community has a significantly larger collection of unpaired single-modality datasets. To fully utilize these existing unpaired datasets, various computational tools have been developed to integrate them into a shared latent space. For instance, tools like Harmony (Korsunsky et al. 2019) and SpaGE (Abdelaal et al. 2020) are designed to project different modalities into a common space, enabling the alignment of diverse data types. Specifically, Multi-Omics Factor Analysis (MOFA) (Argelaguet et al. 2018) employs factor analysis to uncover shared and modality-specific factors across datasets. Another approach used autoencoder-based models, such as Multi-VI (Ashuach et al. 2023) for integration of scRNA-seq and scATAC-seq data. For spatial data integration, tools like Tangram (Biancalani et al. 2021), SpaGE (Abdelaal et al. 2020), and gimVI (Lopez et al. 2019) map non-spatial scRNA-seq data onto spatial coordinates, thereby reconstructing spatially resolved gene expression patterns.

1.5 Summary

We are at an exciting moment in the single-cell multi-omics era, with both sequencing and spatial technologies advancing rapidly. Imagine that, in a perfect world, one could identify a cell type by its transcriptome, confirm a mutation through genomic data, visualize the 3D chromosome organization together with epigenetic modifications that regulate gene expression, and observe the functional protein outputs, the spatial location of a particular cell, and neighborhood interaction between cells — all in one comprehensive framework. Building on the concept of "virtual cells," which aims to reconstruct cellular activity in silico (Bunne et al. 2024), my vision is to leverage these multidimensional data to develop "virtual tissue," enabling the simulation of tissue- or even organ-level behavior. Such integrative approaches will profoundly update our understanding of biology and disease, at the same time offering new opportunities in drug discovery, precision medicine, and beyond.

1.6 References

Abdelaal, T. et al., 2020. SpaGE: Spatial gene enhancement using scRNA-seq. Nucleic acids research, 48(18), p.e107.

Angerer, P. et al., 2017. Single cells make big data: New challenges and opportunities in transcriptomics. Current opinion in systems biology, 4, pp.85–91.

Angermueller, C. et al., 2016. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nature methods, 13(3), pp.229–232.

Anon, In Situ Hybridization (ISH). Available at: https://www.ncbi.nlm.nih.gov/probe/docs/techish/ [Accessed May 12, 2025].

Arendt, D. et al., 2016. The origin and evolution of cell types. Nature reviews. Genetics, 17(12), pp.744–757.

Argelaguet, R. et al., 2020. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome biology, 21(1), p.111.

Ashuach, T. et al., 2023. MultiVI: deep generative model for the integration of multimodal data. Nature methods, 20(8), pp.1222–1231.

Badia-I-Mompel, P. et al., 2023. Gene regulatory network inference in the era of single-cell multi-omics. Nature reviews. Genetics, 24(11), pp.739–754.

Bai, Z. et al., 2024. Spatially exploring RNA biology in archival formalin-fixed paraffin-embedded tissues. Cell, 187(23), pp.6760–6779.e24.

Baysoy, A. et al., 2023. The technological landscape and applications of single-cell multi-omics. Nature reviews. Molecular cell biology, 24(10), pp.695–713.

Belhocine, K., , PhD, DeMare, L., , PhD & Habern, O., 2021. Single-Cell Multiomics: Simultaneous Epigenetic and Transcriptional Profiling: 10x Genomics shares experimental planning and sample preparation tips for the Chromium Single Cell Multiome ATAC + Gene Expression system. Genetic engineering & biotechnology news, 41(1), pp.66–68.

Biancalani, T. et al., 2021. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. Nature methods, 18(11), pp.1352–1362.

Braun, E. et al., 2023. Comprehensive cell atlas of the first-trimester developing human brain. Science (New York, N.Y.), 382(6667), p.eadf1226.

Bravo González-Blas, C. et al., 2023. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nature methods, 20(9), pp.1355–1367.

Bunne, C. et al., 2024. How to build the virtual cell with artificial intelligence: Priorities and opportunities. Cell, 187(25), pp.7045–7063.

Cao, J. et al., 2020. A human cell atlas of fetal gene expression. Science (New York, N.Y.), 370(6518), p.eaba7721.

Cerrizuela, S. et al., 2022. High-throughput scNMT protocol for multiomics profiling of single cells from mouse brain and pancreatic organoids. STAR protocols, 3(3), p.101555.

Chen, A. et al., 2022. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell, 185(10), pp.1777–1792.e21.

Chen, A.F. et al., 2022. NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. Nature methods, 19(5), pp.547–553.

Chen, K.H. et al., 2015. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science (New York, N.Y.), 348(6233), p.aaa6090.

Clark, S.J. et al., 2018. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nature communications, 9(1), p.781.

Colino-Sanguino, Y. et al., 2024. Performance comparison of high throughput single-cell RNA-Seq platforms in complex tissues. Heliyon, 10(17), p.e37185.

Dey, S.S. et al., 2015. Integrated genome and transcriptome sequencing of the same cell. Nature biotechnology, 33(3), pp.285–289.

Eng, C.-H.L. & Others, 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature, 568(7751), pp.235–239.

Fiskin, E. et al., 2022. Single-cell profiling of proteins and chromatin accessibility using PHAGE-ATAC. Nature biotechnology, 40(3), pp.374–381.

Flusberg, B.A. et al., 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nature methods, 7(6), pp.461–465.

Gong, Y. et al., 2025. Stereo-seq of the prefrontal cortex in aging and Alzheimer's disease. Nature communications, 16(1), p.482.

Hagemann-Jensen, M. et al., 2020. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nature biotechnology, 38(6), pp.708–714.

Hao, Y. et al., 2021. Integrated analysis of multimodal single-cell data. Cell, 184(13), pp.3573–3587.e29.

Hodge, R.D. et al., 2019. Conserved cell types with divergent features in human versus mouse cortex. Nature, 573(7772), pp.61–68.

Im, K. et al., 2019. An introduction to performing immunofluorescence staining. Methods in molecular biology (Clifton, N.J.), 1897, pp.299–311.

Korsunsky, I. et al., 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. Nature methods, 16(12), pp.1289–1296.

Kozareva, V. et al., 2021. A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. Nature, 598(7879), pp.214–219.

Kuppe, C. et al., 2022. Spatial multi-omic map of human myocardial infarction. Nature, 608(7924), pp.766–777.

Labade, A.S. et al., 2024. Expansion in situ genome sequencing links nuclear abnormalities to hotspots of aberrant euchromatin repression. Genomics. Available at: https://www.biorxiv.org/content/10.1101/2024.09.24.614614v1.

Lake, B.B. et al., 2023. An atlas of healthy and injured cell states and niches in the human kidney. Nature, 619(7970), pp.585–594.

Lee, J.H. et al., 2015. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nature protocols, 10(3), pp.442–458.

Levine, M. & Davidson, E.H., 2005. Gene regulatory networks for development. Proceedings of the National Academy of Sciences of the United States of America, 102(14), pp.4936–4942.

Li, H. et al., 2024. Transcriptomic, epigenomic, and spatial metabolomic cell profiling redefines regional human kidney anatomy. Cell metabolism, 36(5), pp.1105–1125.e10.

Lindström, N.O. et al., 2018. Conserved and Divergent Features of Human and Mouse Kidney Organogenesis. Journal of the American Society of Nephrology: JASN, 29(3), pp.785–805.

Lister, R. et al., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature, 462(7271), pp.315–322.

Liu, H. et al., 2023. Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. Nature, 624(7991), pp.366–377.

Liu, Y. et al., 2020. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. Cell, 183(6), pp.1665–1681.e18.

Longo, S.K. et al., 2021. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. Nature reviews. Genetics, 22(10), pp.627–644.

Lopez, R. et al., 2019. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. arXiv [cs.LG]. Available at: https://doi.org/10.48550/arXiv.1905.02269.

Lubeck, E. et al., 2014. Single-cell in situ RNA profiling by sequential hybridization. Nature methods, 11(4), pp.360–361.

Lu, T., Ang, C.E. & Zhuang, X., 2022. Spatially resolved epigenomic profiling of single cells in complex tissues. Cell, 185(23), pp.4448–4464.e17.

Marx, V., 2021. Method of the Year: spatially resolved transcriptomics. Nature Methods, 18(1), pp.9–14.

Ma, S. et al., 2020. Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell, 183(4), pp.1103–1116.e20.

Mateo, L.J. et al., 2019. Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature, 568(7750), pp.49–54.

Mathys, H. et al., 2019. Single-cell transcriptomic analysis of Alzheimer's disease. Nature, 570(7761), pp.332–337.

Mimitou, E.P. et al., 2021. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. Nature biotechnology, 39(10), pp.1246–1258.

Nadalin, F. et al., 2024. Multi-omic lineage tracing predicts the transcriptional, epigenetic and genetic determinants of cancer evolution. Nature communications, 15(1), p.7609.

Nofech-Mozes, I. et al., 2023. Pan-cancer classification of single cells in the tumour microenvironment. Nature communications, 14(1), p.1615.

Oliveira, M.F. et al., 2024. Characterization of immune cell populations in the tumor microenvironment of colorectal cancer using high definition spatial profiling. Cancer Biology. Available at: https://www.biorxiv.org/content/10.1101/2024.06.04.597233v1.

Peterson, V.M. et al., 2017. Multiplexed quantification of proteins and transcripts in single cells. Nature biotechnology, 35(10), pp.936–939.

Ramsköld, D. et al., 2012. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nature biotechnology, 30(8), pp.777–782.

Rang, F.J. et al., 2022. Single-cell profiling of transcriptome and histone modifications with EpiDamID. Molecular cell, 82(10), pp.1956–1970.e14.

Rodriques, S.G. et al., 2019. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science (New York, N.Y.), 363(6434), pp.1463–1467.

Rood, J.E. et al., 2025. The Human Cell Atlas from a cell census to a unified foundation model. Nature, 637(8048), pp.1065–1071.

Rozenblatt-Rosen, O. et al., 2021. Building a high-quality Human Cell Atlas. Nature biotechnology, 39(2), pp.149–153.

Shah, S. et al., 2018. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. Cell, 174(2), pp.363–376.e16.

Shah, S. et al., 2017. SeqFISH accurately detects transcripts in single cells and reveals robust spatial organization in the hippocampus. Neuron, 94(4), pp.752–758.e1.

Shalek, A.K. et al., 2013. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature, 498(7453), pp.236–240.

Stoeckius, M. et al., 2017. Simultaneous epitope and transcriptome measurement in single cells. Nature methods, 14(9), pp.865–868.

Su, J.-H. et al., 2020. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. Cell, 182(6), pp.1641–1659.e26.

Su, Y. et al., 2020. Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. Cell, 183(6), pp.1479–1495.e20.

Tabula Sapiens Consortium* et al., 2022. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science (New York, N.Y.), 376(6594), p.eabl4896.

Takei, Y. et al., 2021. Single-cell nuclear architecture across cell types in the mouse brain. Science (New York, N.Y.), 374(6567), pp.586–594.

Takei, Y. et al., 2025. Spatial multi-omics reveals cell-type-specific nuclear compartments. Nature. Available at: http://dx.doi.org/10.1038/s41586-025-08838-x.

Tang, F. et al., 2009. mRNA-Seq whole-transcriptome analysis of a single cell. Nature methods, 6(5), pp.377–382.

Tang, X. et al., 2023. Explainable multi-task learning for multi-modality biological data analysis. Nature communications, 14(1), p.2546.

Wang, X. et al., 2018. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science (New York, N.Y.), 361(6400). Available at: http://dx.doi.org/10.1126/science.aat5691.

Woodworth, M.A. et al., 2021. Multiplexed single-cell profiling of chromatin states at genomic loci by expansion microscopy. Nucleic acids research, 49(14), p.e82.

Xia, C. et al., 2019. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proceedings of the National Academy of Sciences of the United States of America, 116(39), pp.19490–19499.

Xiong, X. et al., 2024. Deciphering breast cancer dynamics: insights from single-cell and spatial profiling in the multi-omics era. Biomarker research, 12(1), p.107.

You, Y. et al., 2024. Systematic comparison of sequencing-based spatial transcriptomic methods. Nature methods, 21(9), pp.1743–1754.

Zeng, H., 2022. What is a cell type and how to define it? Cell, 185(15), pp.2739–2755.

Zhang, D. et al., 2023. Spatial epigenome-transcriptome co-profiling of mammalian tissues. Nature, 616(7955), pp.113–122.

Zheng, G.X.Y. et al., 2017. Massively parallel digital transcriptional profiling of single cells. Nature communications, 8(1), p.14049.

Zhu, C. et al., 2019. An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. Nature structural & molecular biology, 26(11), pp.1063–1070.

Zhu, C. et al., 2021. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nature methods, 18(3), pp.283–292.

Zhu, C., Preissl, S. & Ren, B., 2020. Single-cell multimodal omics: the power of many. Nature methods, 17(1), pp.11–14.

C h a p t e r 2

SPATIAL MULTI-OMICS REVEALS CELL-TYPE-SPECIFIC NUCLEAR COMPARTMENTS

A modified version of this chapter was published as:

*Takei Y, *Yang Y, White J, Goronzy IN, Yun J, Prasad M, Ombelets LJ, Schindler S, Bhat P, Guttman M, Cai L. Spatial multi-omics reveals cell-type-specific nuclear compartments. Nature. 2025 Apr 9. doi: 10.1038/s41586-025-08838-x. *These authors contributed equally.

2.1 Abstract

The mammalian nucleus is compartmentalized by diverse subnuclear structures. These subnuclear structures, marked by nuclear bodies and histone modifications, are often cell-type specific and affect gene regulation and 3D genome organization (Solovei et al. 2016; Bhat et al. 2021; Belmont 2021). Understanding nuclear organization and its role in gene expression requires identifying the molecular constituents of subnuclear structures and mapping their associations with specific genomic loci and their transcriptional levels in individual cells, within complex tissues. Here, we introduce two-layer DNA seqFISH+, which allows simultaneous mapping of 100,049 genomic loci, together with nascent transcriptome for 17,856 genes and a diverse set of immunofluorescently labeled subnuclear structures all in single cells in both cell lines and complex tissues. These data enable imaging-based chromatin profiling of diverse subnuclear markers and capture changes in chromatin organization at genomic scales from 100–200 kb to approximately 1 Mb, depending on the subnuclear marker and DNA locus. Using multi-omics datasets in the adult mouse cerebellum, we showed that repressive chromatin regions are more variable by cell type than active regions across the genome. We also discovered RNA polymerase II (RNAPII)-enriched foci were locally associated with long, cell-type specific genes (> 200kb), in a manner distinct from nuclear speckles. Further, our analysis revealed that cell-type specific facultative and constitutive heterochromatin regions marked by H3K27me3 and H4K20me3 are enriched at specific genes and gene clusters, respectively, and shape radial chromosomal

positioning and inter-chromosomal interactions in neurons and glial cells. Together, our results provide a single-cell high-resolution multi-omics view of subnuclear structures and chromatin marks, associated genomic loci, and their impacts on gene regulation, directly within complex tissues.

2.2 Introduction

Recent imaging-based genome-wide multimodal technologies have enabled direct profiling of the 3D organization of the nucleus in single cells (Nguyen et al. 2020; Su et al. 2020; Payne et al. 2021; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021; Lu et al. 2022), providing spatial context to our understanding of nuclear architecture derived from genome-wide sequencing-based approaches (Dekker et al. 2017; Kempfer & Pombo 2020; Jerkovic & Cavalli 2021; Conte et al. 2020; Wang et al. 2021; Boninsegna et al. 2022). For example, we had previously shown that specific associations between genomic loci and subnuclear structures are conserved across single cells, despite the apparent variability in the 3D genome structure of individual cells (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). Furthermore, imaging-based transcriptomic approaches have revealed the organization of the nascent transcriptome within the nucleus (Shah et al. 2018) and identified cell-type specific transcriptional programs in tissues (Eng et al. 2019). However, due to optical crowding, current genome-wide imaging approaches are limited to resolving genomic sites at the megabase level (Su et al. 2020; Payne et al. 2021; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021), limiting the achievable level of insight to larger genomic regions, rather than individual genes across the genome.

To enable a more detailed understanding of nuclear organization and its relationship to gene regulation, higher genomic coverage measurements are required together with measurements of diverse subnuclear structures in single cells. Here we introduce a two-layer barcoding DNA seqFISH+ scheme that increases the multiplexing capability of single-cell multi-omics to ~100,000 species, up from the previous ~10,000 (Shah et al. 2018; Eng et al. 2019; Xia et al. 2019), corresponding to 25-kb coverage across the genome (Methods) (Fig. 1b, Extended Data Fig. 1). This two-layer DNA seqFISH+ can be combined with transcriptomic and

immunofluorescence measurements to perform simultaneous investigation of gene expression profiles and chromatin organization across the genome in individual cells (Fig. 1a, Extended Data Figs. 1-7), and enables identification of nuclear compartments and their associated genes in a cell-type specific fashion. We note that our measurements can capture the cell-type specific changes of chromatin organization at 100–200 kb to approximately 1-Mb genomic scales, depending on the immunofluorescence target and chromosome locus, due to the limited spatial resolution of diffraction-limited immunofluorescence images (~200–300 nm) and target-specific issues related to the processing of imaging-based chromatin profiles as described below.

2.3 Results

Two-layer seqFISH+: Single-cell spatial multi-omics technology

Building upon our previous multimodal technology (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021), our goal was to employ a multi-omics approach to characterize the genomic landscape of individual cells in greater depth. The motivation for the two-layer strategy was sparked by previous observations that while some DNA loci are close together, those separated by more than 3 Mb on the genome are on average likely more than 300 nm apart from each other (Wang et al. 2016; Bintu et al. 2018; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021; Chen et al. 2023; Hafner et al. 2023), beyond the diffraction limit. We therefore divided each chromosome into 1.5 Mb blocks and distributed them into three orthogonal fluorescent channels such that adjacent chromosome blocks within the same fluorescent channel are separated by 3 Mb. We then sequentially imaged DNA loci at 25-kb increments in a different block in parallel, circumventing optical crowding, because within the same fluorescent channel, the loci imaged in any given round are genomically at least 3 Mb apart from each other. We used 96 rounds of imaging to impart two-layer barcodes on 100,049 loci across the genome. The initial 60 rounds resolve 25-kb segments within each chromosome block while the subsequent 36 rounds decode chromosome block identities based on the unique combinations of the 9-pseudocolors across 4 barcoding rounds by leveraging the previous seqFISH+ pseudocolor approach (Eng et al. 2019; Takei, Yun, et al.

2021; Takei, Zheng, et al. 2021) to the chromosome blocks (Fig. 1b). We have named this imaging-based two-layer barcoding method as two-layer seqFISH+.

We then combined two-layer DNA seqFISH+ with transcriptomic and subnuclear structure measurements to generate multi-omics datasets that integrate information on individual genomic loci (100,049 loci on 20 chromosomes by two-layer DNA seqFISH+), with information on mature and nascent transcripts (up to 1,247 genes based on mRNAs detection by mRNA seqFISH+ or non-barcoded mRNA seqFISH and 17,856 genes based on intron detection by intron seqFISH+) and as well as information on subnuclear structures and chromatin marks based on sequential immunofluorescence assays (using up to 65 antibodies) in thousands of single cells (Fig. 1a-g, Extended Data Fig. 1).

We initially applied this technology to two different mouse cell lines (mouse embryonic stem cells (mESCs), and mammary gland epithelial NMuMG cells), and then compared the results to those from established approaches for benchmarking. We began with two-layer DNA seqFISH+ in mESCs, detecting 63,466 ± 20,525 (median ± s.d.) DNA dots per cell across 100,049 genomic loci in 1,076 cells from two biological replicates (Fig. 1h, top). The estimated detection efficiency was 21.1% while the false positive rate was estimated at 1.6% (Fig. 1h, bottom). In addition, throughout the imaging rounds, we imaged the fiducial marker targeting endogenous genomic loci (Takei, Yun, et al. 2021), achieving a precise alignment of images from different rounds with a median alignment accuracy of 47.3 nm for DNA seqFISH+ (Extended Data Fig. 3a-c). We then compared our two-layer DNA seqFISH+ data with previously generated DNA seqFISH+ data (Takei, Yun, et al. 2021) as well as the sequencing-based method, Hi-C (Lieberman-Aiden et al. 2009; Bonev et al. 2017), and confirmed that A/B compartments (Lieberman-Aiden et al. 2009; Bonev et al. 2017) (Spearman's correlation coefficient of 0.82 between two-layer DNA seqFISH+ and Hi-C in the genome-wide comparison) and other measures are overall consistent amongst the datasets (Fig. 1i, Extended Data Fig. 3). However, we note that the two-layer barcoding scheme under-detects the DNA loci near the boundary of 1.5 Mb chromosome blocks by approximately 20%, leading to artificial boundaries in the pairwise spatial distance

calculations with 25 kb bins (Extended Data Fig. 3l-t), but minimally affects chromatin profiling and megabase-scale pairwise spatial distance analyses shown below (Extended Data Fig. 3f, g; see Supplementary Notes for detailed discussion).

Next we examined transcriptomic data from mRNA seqFISH+ (detecting 6,496 ± 2,068 (median ± s.d.) spots per cell and 204 genes with averaged spot counts of >10 per cell), and intron seqFISH+ (1,197 ± 421 (median ± s.d.) transcription active sites per cell and 3,092 genes with averaged spot counts of >0.1 per cell) in mESCs, which correspond to 78.9% and 29.1% estimated detection efficiency, respectively (Extended Data Fig. 2a-k). These detection efficiencies are similar or favorable compared to 5-45% detection efficiencies by typical single-cell RNA-seq measurements (Svensson et al. 2017; Marshall et al. 2020). Our imaging-based transcriptomic measurements also showed a high degree of agreement with sequencing-based orthogonal measurements (Extended Data Fig. 2a, h).

To gain further insight into the spatial organization of these loci, we employed imaging-based chromatin profiling (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021), which measures spatial proximity between genomic loci visualized by two-layer DNA seqFISH+ and subnuclear structures visualized by sequential immunofluorescence (Fig. 1g, j, Extended Data Figs. 4-7). To examine subnuclear structures, we included various histone modifications (Kouzarides 2007) (e.g., H3K27ac, H3K9me3) as well as other chromatin-associated factors associated with nuclear bodies such as nuclear speckles (Chen & Belmont 2019) (e.g., SF3A66) and the nucleolus (Pederson 2011) (e.g., Fibrillarin), and the nuclear lamina (Van Steensel & Belmont 2017) (e.g., Lamin B1). The imaging-based chromatin profiles in mESCs were overall consistent with those generated by sequencing-based methods (e.g., CUT&RUN (Skene & Henikoff 2017; Skene et al. 2018), pA-DamID (van Schaik et al. 2020), and RD-SPRITE (Quinodoz et al. 2021)) as well as previous imaging-based approaches for a wide range of subnuclear structures and histone modifications (e.g., Lamin B1, H3K4me3) (Fig. 1j, k, Extended Data Figs. 4-7). Our imaging-based chromatin profiling also captured biological changes of lamina-associated domains (LADs) between

mESCs and NMuMG cells, similarly to the orthogonal sequencing datasets (Extended Data Figs. 4k, 5a).

We note that there are fundamental differences between imaging- and sequencing-based chromatin profiling approaches. Specifically, our imaging-based chromatin profiling approach captures spatial proximity of DNA loci relative to subnuclear structures or histone modifications (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) (either average spatial proximity or fraction of loci within certain thresholds; see Methods), rather than detecting their individual molecular binding at individual genomic sites, which sequencing-based methods such as ChIP-seq (Johnson et al. 2007), CUT&RUN (Skene & Henikoff 2017; Skene et al. 2018), and CUT&Tag (Kaya-Okur et al. 2019) as well as imaging-based epigenomic MERFISH for targeted loci (Lu et al. 2022) typically characterize (Extended Data Fig. 4g, j). Similarly, it is difficult to distinguish a DNA locus being enriched with a given chromatin marker from being spatially associated with neighboring DNA loci enriched with the marker in single cells. In addition, spatial resolution in our imaging-based chromatin profiling is fundamentally limited by the diffraction-limited immunofluorescence images (~200–300 nm spatial resolution) (Woodworth et al. 2021). Thus, we used 100–200 kb genomic bins, which are typically 100-300 nm away from the adjacent bin for the spatial chromatin profiling analysis. Third, because we used the z-score based linear scaling in the chromatin profiling to achieve a systematic way to extract the chromatin marker enrichment information, the dynamic range of fluorescence intensities of a given marker can affect the interpretation of the marker enriched DNA loci (Extended Data Fig. 4g, j). Finally, the analysis methods for imaging-based chromatin profiles may reflect different chromatin features (see Methods).

Nonetheless, because of the improved coverage of genomic loci provided by two-layer DNA seqFISH+, this newly generated chromatin profiling revealed chromatin organization at 100 kb–1 Mb practical resolution (Fig. 1i, j, l, Extended Data Fig. 5-7), a level that was inaccessible in previous imaging-based genome-wide studies (Su et al. 2020; Payne et al. 2021; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021), including the previous DNA

seqFISH+ with genomic bins approximately 1 Mb apart (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). For example, for nuclear speckle associations, we detected DNA seqFISH+ SF3A66 peaks that correspond to Rn7sk peaks from the RD-SPRITE profile with 100-kb binning (Fig. 1l, Extended Data Fig. 6c-e). While we also identified the same 100-kb bin peaks between imaging-based and sequencing-based chromatin profiling in some cases (e.g. RNAPIISer5-P), the overall correlation between datasets improved significantly up to ~350–400 kb on average, depending on the marker and chromosome (Extended Data Figs. 5e, 7a). These results showed context-dependent practical resolution of chromatin profiling at genomic scales from 100–200 kb to approximately 1 Mb.

Together, these results confirm the high quality of the spatial multi-omic datasets, which enable us to perform an integrated analysis of nuclear organization across multiple imaging modalities. We will refer to this two-layer DNA seqFISH+ as DNA seqFISH+ for simplicity below.

Single-cell spatial multi-omics in the mouse cerebellum

Having established the validity of our approach in mouse cell lines, we proceeded to examine the relationships between chromatin organization and gene regulation in the naive tissue context. We chose the adult mouse cerebellum, a well-defined brain structure with diverse cell types (Kozareva et al. 2021), to examine the cell-type specificity of nuclear organization. Thus, we applied DNA seqFISH+ (100,049 loci), intron seqFISH+ (17,856 genes), mRNA seqFISH (60 marker genes), and sequential immunofluorescence (27 markers) to the adult mouse cerebellum.

We identified distinct cell types by mRNA seqFISH transcriptomic profiling and spatially resolved the organization of these diverse cell types in the tissue sections (Fig. 2a-c, Extended Data Figs. 8-10), capturing the layer organization of the mouse cerebellum at the single-cell resolution. We were able to identify major cell types in the adult mouse brain cerebellum (Kozareva et al. 2021), including neurons such as Purkinje cells, Purkinje layer interneurons (PLI), subtypes of molecular layer interneurons (MLI1 and MLI2), and granule cells, as well

as non-neuronal cells such as Bergmann glia, astrocytes, oligodendrocyte precursor cells and oligodendrocytes (OPC/ODC), and endothelial cells. We confirmed the expected layer organization of the adult mouse cerebellum (Kozareva et al. 2021), such as spatially close arrangement of Purkinje cells and Bergmann glia at the Purkinje cell layer, which is adjacent to the granule cell (GC) layer, largely consisting of granule cells (Fig. 2a, Extended Data Fig. 8a-c). Our multi-omics profiling across 4,015 cells further allowed us to compare the genomic landscapes of the different cell types at multiple levels, including nascent transcriptional states by intron seqFISH+ and chromatin states by sequential fluorescence (Fig. 2b, Extended Data Fig. 8). These analyses revealed highly consistent cell-type specific states at each level of analysis (i.e., transcriptional and chromatin).

Cell-type specific active and repressive chromatin regions in mouse cerebellar cells

Based on the imaging of diverse subnuclear markers together with genome-wide DNA loci by DNA seqFISH+, we were able to identify several different types of chromatin regions. Active chromatin regions involved nuclear speckles (SF3A66), which are known to enrich in pre-mRNA splicing factors (Chen & Belmont 2019), and other active chromatin markers such as H3K27ac and RNA polymerase II (RNAPIISer5-P) (Fig. 2d, e, Extended Data Fig. 9d, e). In addition, we also identified at least four major repressive chromatin regions (Solovei et al. 2016; Falk et al. 2019) by examining eight repressive markers, including constitutive heterochromatin (major satellite DNA repeats (MajSat), H4K20me3), facultative heterochromatin (H3K27me3), nucleolus (ITS1 RNA), and nuclear periphery (Lamin B1) (Fig. 2d-f, Extended Data Fig. 9d-g).

Interestingly, we observed that repressive markers were associated with genomic loci in a highly cell-type specific fashion, while active chromatin profiles were largely conserved between cell types using 200 kb bins (Fig. 2g). For example, by comparing Purkinje and MLI1 cells, we observed that the profiles of active chromatin regions associated with nuclear speckles (marked by SF3A66, as described above) were highly correlated with Pearson correlation coefficient of 0.89, whereas those of repressive regions, marked by H3K27me3, were less so with Pearson correlation coefficient of 0.24 (Fig. 2g, right). These observations

were consistent across various active and repressive markers and in all cell types and cell lines examined (Extended Data Fig. 9h).

Furthermore, some of the repressive markers also showed cell-type specific subnuclear localization patterns. For example, H4K20me3 marker stained subnuclear territories that were segregated from pericentromeric heterochromatin characterized by MajSat (Linhoff et al. 2015) in Purkinje cells and MLI1, while those markers stained spatially overlapped regions in Bergmann glia (Fig. 2f, Extended Data Fig. 9f). In addition, regions marked with mH2A1 and H3K27me2 staining, which correlated with H3K27me3 foci in all cell types, also stained pericentromeric regions only in Purkinje cells (Extended Data Fig. 9f). Thus, genomic loci showed cell-type dependent associations with subnuclear territories stained by those repressive markers (Extended Data Fig. 9f, g). Finally, we observed cell-type specific organization of subnuclear foci (Extended Data Fig. 10b-e). In particular, while larger MajSat foci exhibited more interior localization in the nucleus in Purkinje cells and MLI1, consistent with a previous report in Purkinje cells (Solovei et al. 2004), such a trend was not observed in Bergmann glia (Extended Data Fig. 10e), suggesting a different pericentromeric heterochromatin organization between those neurons and glial cells.

We examine these active and repressive chromatin regions and their cell-type specificities using DNA seqFISH+ datasets in more detail below (Extended Data Figs. 11-16). We also note that major biological observations were further validated by additional sequential FISH experiments (~60% detection efficiency) with a smaller set of 22 DNA loci with two biological replicates (Extended Data Fig. 17).

Active chromatin regions show distinct sequence features and cell-type specific gene expression patterns

We first focused on transcriptionally active chromatin regions. Previous work had shown that speckle-associated genomic regions are largely conserved across cell types (Zhang et al. 2020; Takei, Zheng, et al. 2021), but associations with RNAPII have not been fully explored in tissues. In cerebellar cells, genomic regions associated with nuclear speckles tended to be

higher in GC content, enriched in shorter-length genes, and enriched in RNAPIISer5-P marks, regardless of cell type (Fig. 3a-c, Extended Data Figs. 10b, 11a, b), consistent with previous works(Chen et al. 2018; Quinodoz et al. 2018; Zhang et al. 2020; Takei, Zheng, et al. 2021; Alexander et al. 2021). The RNAPIISer5-P broad regions, lacking SF3A66 nuclear speckle marks, formed a distinct territory far from the speckles (Fig. 3b, Extended Data Fig. 11c), similar to those observed in human cells (Chen et al. 2018). Those speckle-enriched and RNAPIISer5-P broad regions not only showed different sequencing features (e.g., GC content) (Fig. 3c, Extended Data Figs. 11b, 12a-c) but also encoded a distinct set of Gene Ontology (GO) terms and a different percentage of housekeeping genes (i.e., a higher percentage of housekeeping genes in speckle-enriched regions) in each cell type (Extended Data Figs. 11d, 12d).

Both nuclear speckles and RNAPIISer5-P broad associated loci showed subtle cell type dependent patterns in their genomic locations (Extended Data Fig. 11e). Nevertheless, we observed that increased association of genomic loci to one of these markers are typically correlated with cell-type specific gene expression programs (Fig. 3d, e, Extended Data Fig. 11f, g). For example, increased locus associations toward either nuclear speckles or RNAPIISer5-P foci are coupled with increased gene expression of the loci when comparing Bergmann glia and MLI1 (Fig. 3d, e, bottom). We note that Purkinje cells were an exception, however, where only increased associations to RNAPIISer5-P foci correlated with increased gene expression (Fig. 3e, Extended Data Fig. 11g).

Furthermore, we classified a third active chromatin region associated with RNAPIISer5-P sharp peaks (< 2 Mb) and exhibited a distinct set of characteristics. Genomic loci in this region tended to belong to genomic regions with lower GC content (Extended Data Fig. 12a-c), and contained longer genes than the speckle region in all cell types as well as the RNAPIISer5-P broad region in some cell types (Fig. 3c, Extended Data Fig. 11b). In addition, these RNAPIISer5-P sharp loci showed more cell-type specific features, compared to the other active regions (Extended Data Fig. 11e). This observation motivated us to further

characterize the chromatin states and gene expression profiles of long genes (> 200 kb) across cell types.

We were particularly interested in long genes because neuronal tissues (including cerebellum) generally express a greater number of long genes, relative to other tissue types (Gabel et al. 2015). In addition, long genes are overrepresented among genes involved in synaptic processes (Koopmans et al. 2019) and brain disorders such as autism spectrum disorder (ASD) (King et al. 2013; Zhao et al. 2018). Furthermore, transcriptionally-active long genes in tissues have been reported to have unique chromatin signatures including activation-dependent rearrangements (Walczak et al. 2013), broad enhancer-like chromatin domains (Zhao et al. 2018), chromatin decondensation (Winick-Ng et al. 2021), and transcriptional loops (Leidescher et al. 2022). Therefore, it is of crucial importance to understand their subnuclear organization in the naive tissue context.

By using our multi-omics datasets, we found a strong link between the association of long genes with the RNAPIISer5-P foci and their cell-type specific gene expression (Fig. 3f, g). Examples of such genes included Cntnap5b and Dpp10 in Purkinje cells, Pam in MLI1 and MLI2/PLI, Adgrl3 in MLI1 and Bergmann, Slc4a4 in Bergmann, and Kcnd2 and Cadps2 in Granule cells. These long genes tend to play cell-type specific roles in the corresponding cells (e.g., Cadps2, a secretory granule-associated protein indispensable for normal cerebellar functions (Sadakata et al. 2007)), and, indeed, their RNAPIISer5-P enrichment was typically observed only in specific cell types (Fig. 3g). We further confirmed that those long genes are generally enriched for synaptic genes and ASD candidate genes (Extended Data Fig. 12d), consistent with previous observations (King et al. 2013; Zhao et al. 2018; Koopmans et al. 2019). The cell-type specificity was also reflected in the nascent transcription patterns observed by intron seqFISH+ profiles (Fig. 3g), open chromatin states (Extended Data Fig. 11i), and the local positional changes of DNA loci from the nuclear periphery to nuclear interior (Extended Data Fig. 12e-g), which coincides with a transcription-dependent detachment of genomic loci from the nuclear lamina upon synthetic transcriptional activation (Brueckner et al. 2020). Chromatin profiling, again combining DNA seqFISH+ with

sequential immunofluorescence, further identified other subnuclear markers associated with these long genes, including H3K4me2, H4K8ac, and CBP (Extended Data Fig. 11h, 12h). However, the nuclear speckle marker SF3A66 typically did not appear at those loci (Fig. 3h, i, Extended Data Fig. 11h), demonstrating a spatial partitioning between nuclear speckles and other active markers (e.g., RNAPIISer5-P, H3K4me2) (Fig. 3b, Extended Data Fig. 11c). This cell-type specific drastic chromatin reorganization at long gene loci contrasts with the relatively small positional changes observed at gene-dense genomic loci around nuclear speckles (Zhang et al. 2020; Takei, Zheng, et al. 2021). The differential subnuclear organization between nuclear speckle-associated loci and transcriptionally active long gene loci was also recently reported in the adult mouse brain cerebellum in an independent study (Zhao et al. 2023). The distinct subnuclear organization of transcriptionally active loci between nuclear speckles and RNAPII may be functionally important in gene regulation, such as gene expression amplification (Alexander et al. 2021) and mRNA splicing (Bhat et al. 2023), and could also be critical in promoting gene misexpression in disease (e.g., speckle-associated loci in schizophrenia (Ahanger et al. 2021)).

Cell-type specific organization of H3K27me3-associated repressive chromatin regions

In contrast to active genomic regions, genomic loci associated with repressive chromatin markers were highly cell-type specific (Fig. 2f, g). We first examined H3K27me3-marked regions, given the role of this repressive histone modification in silencing neurodegeneration-related genes (Von Schimmelmann et al. 2016). We found there are two major subsets of H3K27me3-associated loci in cerebellar cells. The first set of loci were non-cell-type specific and enriched with genes involved in general developmental processes, represented by gene ontology (GO) terms related to pattern specification, such as the Hox gene clusters on Chr2 and Chr6 (Extended Data Fig. 13a). These associations are consistent with previous data in other biological contexts such as embryonic development (O'Geen et al. 2007; Vieux-Rochas et al. 2015). The other set of H3K27me3-associated loci were cell-type specific. Specifically, Purkinje cells showed the H3K27me3 enrichment at genomic loci with longer genes and lower GC content, compared to other cell types such as MLI1 and Bergmann glia

(Fig. 4a, Extended Data Fig. 13b). This Purkinje-specific region included genes such as Grin2b, whose repression is functionally important in Purkinje cells (Galliano et al. 2018), and Ptprd, which regulates developmental neurogenesis (Tomita et al. 2020) (Fig. 4b-d, Extended Data Fig. 13c, d). Some of these long genes (36 out of 46 genes) are developmentally down-regulated from the newborn to the adult Purkinje cells (Stoyanova et al. 2021) (Extended Data Fig. 13e).

In addition, we found that H3K27me3- and Lamin B1-enriched loci are negatively correlated in most cell types (Fig. 4b, Extended Data Fig. 13f). However, a subset of H3K27me3-associated loci is enriched with Lamin B1 in Purkinje cells. For example, Ptprd gene locus, spanning for ~2.2 Mb in Chr4, localized to the nuclear periphery and were marked by Lamin B1 in all major cell types (Purkinje cells, MLI1, and Bergmann glia), but were also marked by H3K27me3 only in Purkinje cells (Fig. 4b-d and Extended Data Fig. 13d). Despite these global organizational differences of the H3K27me3 foci around the nuclear periphery, the increased association toward either H3K27me3 or Lamin B1 showed decreased nascent transcription levels when comparing the pairs of cell types (Extended Data Fig. 13g). Together, Purkinje cells show a unique organization of their H3K27me3 repressive staining, perhaps related to their overall highly transcriptionally active nuclei (Van't Sant et al. 2021) and increased H3K27me3 modification levels (Extended Data Fig. 8k, l).

The H4K20me3 subnuclear organization is associated with specific gene families

We next examined the subnuclear organization of the repressive region marked by H4K20me3. H4K20me3 is a repressive histone modification at constitutive heterochromatin regions and involved in gene silencing (Agredo & Kasinski 2023). Among the repressive markers we imaged, H4K20me3 staining marked unique territories adjacent to pericentromeric heterochromatin (Linhoff et al. 2015) in Purkinje cells (Fig. 5a, Extended Data Fig. 14a). While H4K20me3 was previously characterized in specific biological contexts such as olfactory receptor choice (Magklara et al. 2011; Clowney et al. 2012; Armelin-Correa et al. 2014; Monahan et al. 2019; Tan et al. 2019) and Rett Syndrome (Linhoff et al. 2015), the cell-type specific association of DNA loci with this modification as

well as its broader impact on chromatin organization remain elusive. We therefore examined the cell-type specific DNA loci associated with H4K20me3 regions in the nucleus to characterize their underlying genomic features (Fig. 5b, Extended Data Fig. 14a-c).

We found that H4K20me3-enriched loci were depleted from H3K27me3 and Lamin B1 markers in Purkinje cells (Extended Data Fig. 14b, c). Notably, we also found that H4K20me3-enriched loci in Purkinje cells occur in non-CG methylation (mCH) deserts, i.e. regions that do not accumulate mCH during development (Lister et al. 2013; Tan et al. 2021), as well as in mouse strain-specific diverse regions (SSDRs) (Lilue et al. 2018; Lilue et al. 2019) (Fig. 5c, Extended Data Fig. 14d). Those enrichments contrasted from spatially adjacent pericentromeric MajSat-associated loci, which are mostly found within 2 Mb from chromosome start coordinates (Fig. 5c, Extended Data Fig. 14a). Moreover, the H4K20me3-enriched loci appeared on specific chromosomes, including Chr7 and Chr17 (Fig. 5b) and included both gene-coding regions and gene deserts. In the gene-coding regions, we identified gene clusters, such as vomeronasal receptors (Vmn), secretoglobins (Scgb), and zinc finger proteins (Zfp) (Extended Data Fig. 14e), some of which (e.g. Vmn, Zfp) are genomically distributed across multiple chromosomes. We also found that Vmn gene clusters are consistently marked by H4K20me3, and do not show nuclear lamina association (Fig. 5d-f, Extended Data Fig. 14h, i). In contrast, other gene family clusters such as olfactory receptors (Olfr) had a more mixed profile, associating with H4K20me3 in some cells while with Lamin B1 at the nuclear periphery in others. This feature gave Olfr genes overall lower H4K20me3 enrichments in Purkinje cells (Fig. 5d-f, Extended Data Fig. 14h, i), despite showing similar sequencing features (i.e., mCH deserts and SSDRs) as Vmn family genes (Lister et al. 2013; Lilue et al. 2018; Lilue et al. 2019). We note although our measurements capture spatial proximity rather than molecular interactions, some of those gene family loci (e.g., Vmn, Zfp, Olfr) were previously reported to have molecular interactions with the H4K20me3 modification by ChIP-qPCR in the mouse main olfactory epithelium (Magklara et al. 2011). Taken together, these data reveal that H4K20me3-marked regions constitute a separate subnuclear territory with highly specific locus associations.

H4K20me3 supports cell-type specific radial chromatin organization

We noticed that certain chromosomal loci such as those in Chr4, Chr7, and Chr17 showed neuron-specific interior radial positioning, different from their arrangement in Bergmann glia (Fig. 5g, Extended Data Fig. 14j). Given that a majority of these loci were enriched for H4K20me3 in Purkinje cells, we wondered whether H4K20me3 may be related to the cell-type specific radial positioning of chromosomes. Consistent with this notion, we found that H4K20me3 territories tend to form in the nuclear interior in Purkinje cells (Solovei et al. 2004) (Fig. 5a), and H4K20me3-associated loci were therefore also localized to the interior (Fig. 5g, Extended Data Fig. 14k). For example, H4K20me3-associated Vmn gene clusters are found at 2.7 ± 0.6 μm (median ± s.d.) interior from the nuclear periphery, in contrast to 1.5 ± 0.5μm (median ± s.d.) for the weakly H4K20me3-associated Olfr gene clusters and 1.3 ± 0.5 μm (median ± s.d.) from all genomic loci in Purkinje cells (Extended Data Fig. 14e). At the same time, those genomic loci were excluded from RNAPIISer5-P enrichment as outliers compared to other loci with similar radial positions in the nucleus (Extended Data Fig. 14k).

Interestingly, we found the loci enriched in H4K20me3 in Purkinje cells were highly conserved with other neurons in their association with the H4K20me3 territories (Fig. 5h-j, Extended Data Fig. 14l, m; 90.5% with MLI1 and 92.5% with MLI2/PLI). However, those loci were not enriched with H4K20me3 in Bergmann glia (8.7%), where they localized at the nuclear periphery and showed enrichment for Lamin B1 (Fig. 5h-j, Extended Data Fig. 14m). For example, the Skint gene cluster and the gene desert regions, which are ~6.8 Mb apart in Chr4, were enriched with H4K20me3 at the nuclear interior and depleted for Lamin B1 in three types of neurons, while the same genomic regions were not marked by H4K20me3 and instead showed enrichment with Lamin B1 and localization to the nuclear periphery in Bergmann glia (Fig. 5g-k). By contrast, chromosomal loci in Chr11 and Chr19 showed more interior radial positioning in glial nuclei compared to neurons (Fig. 5g). We had observed similar differences between neuron versus glia radial chromosomal positioning in the adult mouse cerebral cortex (Takei, Zheng, et al. 2021), where Chr7 and Chr17 are

positioned in the nuclear interior in neurons whereas Chr11 and Chr19 are interior in astrocytes. Similarly, 60.5% of H4K20me3-associated loci identified in Purkinje cells corresponded to radial repositioning loci from the nuclear periphery to interior in forebrain neurons, but not in glial cells, during postnatal brain development (Tan et al. 2021; Tan et al. 2023) (Extended Data Fig. 14n). These results suggest that different brain regions can exhibit conserved nuclear organization patterns that are distinct between neurons and glial cells.

Furthermore, despite the association differences of the identified genomic loci between H4K20m3 and Lamin B1 in neurons and glial cells, the majority of genes in those genomic loci remain silenced regardless of the cell types (Extended Data Fig. 9j). These results suggest that the cell-type specific association changes between those two repressive markers could potentially have structural roles of chromatin organization, rather than contributing to cell-type specific gene expression changes. We speculate that those genes in different repressive subnuclear structures could have differential sensitivities to the disruption of each subnuclear structure in specific diseases or aging processes.

Subnuclear structures underpin the 3D organization of the genome

Having examined the features of each separate chromatin region, we next investigated the spatial relationship of the genome organization with the subnuclear structures in the different cell types of the adult mouse cerebellum (Fig. 6, Extended Data Figs. 15, 16). To do so, we systematically calculated the average inter-chromosomal distances between pairs of genomic loci enriched with subnuclear markers using top 5% genomic loci associated with each marker. We observed that, in all cell types, pairs of genomic loci enriched with the same specific markers - such as markers for nuclear speckles and pericentromeric heterochromatin - show closer average inter-chromosomal distances compared to those with random selection (Fig. 6a, c, Extended Data Fig. 15a-c, f), consistent with previous literature (Hewitt et al. 2008; Clowney et al. 2012; Quinodoz et al. 2018; Quinodoz et al. 2021; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021; Arrastia et al. 2021). Lamin B1-enriched loci were an exception, showing longer spatial distances between pairs of inter-chromosomal loci relative to controls, suggesting a lack of interaction (Fig. 6a, c, Extended Data Fig. 15a, b). At the

other end of the spectrum, H3K27me3-enriched loci were on average closer to each other than random pairs in MLI1 and Bergmann glia but not in Purkinje cells (Fig. 6c). These observations can be explained by the fact that loci localizing to the nuclear periphery, which generally encompasses a larger area, are more likely to be distant from each other, than pairs of loci in the nuclear interior. Thus, pairs of chromosomes with exterior radial positioning at the nuclear periphery tend to be spatially farther away from each other. Consistent with this notion, chromatin profiles showed that H3K27me3-enriched loci localized to the interior in MLI1 and in Bergmann glia, but appear at the nuclear periphery in Purkinje cells, while Lamin B1-enriched loci are at the nuclear periphery in all cell types (Extended Data Fig. 13f).

We further examined the relationships between subnuclear structure and 3D genome organization using the H4K20me3-enriched loci identified in Purkinje cells (Figs. 5g, 6b, Extended Data Figs. 15c, d, 16, Supplementary Movie 1). Consistent with the above model, DNA loci enriched with H4K20me3, which are generally located at the nuclear interior in neurons such as Purkinje and MLI1, exhibited increased inter-chromosome association (i.e., shorter distances) (Fig. 6b, d, Extended Data Fig. 15c, d). In contrast, in Bergmann glia, where those H4K20me3 loci are present at the nuclear periphery enriched with Lamin B1 (Fig. 5j), we observed decreased inter-chromosomal association of the loci (i.e., increased distances) (Fig. 6a, Extended Data Fig. 15c). Furthermore, we found long-range intra-chromosomal associations between the Purkinje H4K20me3-enriched loci (e.g., Chr4 122.0 and 144.0 Mb) were conserved in all those three cell types but not in mESCs (Fig. 6a, Extended Data Fig. 15e). These results suggest that long-range intra-chromosomal interactions are promoted by either H4K20me3 (in Purkinje cells and MLI1) or the nuclear lamina (in Bergmann glia) in the mouse cerebellum. The recent report similarly identified long-range intra-chromosomal interactions in neurons and glial cells from other regions of the mouse brain but not in mESCs (Winick-Ng et al. 2021). We observed that many of these loci are H4K20me3-associated (e.g., Chr7, Vmn2r, and Mrgpr gene clusters in Fig. 5h) in the Purkinje cells, suggesting that those intra-chromosomal interactions could also be mediated by distinct subnuclear structures (H4K20me3 or Lamin B1) in a cell-type specific

fashion. Finally, we showed that H4K20me3-associated loci spatially cluster in single cells (Extended Data Fig. 16). In Purkinje cells and MLI1, 50.7% and 46.7% of clusters respectively are formed by inter-chromosomal hubs, including higher-order organization by three chromosomes (e.g., Chr4, Chr7, and Chr12), while only 18.0% of clusters show such inter-chromosomal organization in Bergmann glia (Extended Data Fig. 16). Thus, there are neuron-specific higher-order inter-chromosomal organization around the H4K20me3 in the mouse cerebellum. Together, these data support a model in which subnuclear structures help to shape the 3D genome organization in the nucleus in a cell-type specific fashion (Fig. 6d).

2.4 Discussion

Here we demonstrated how high-resolution seqFISH-based single-cell multi-omics profiling can reveal cell-type specific chromatin organization in native tissues and identify specific genes associated with each chromatin marker. We showed that our new barcoding strategy (two-layer DNA seqFISH+) can effectively cover a large number (>100,000) of genomic loci across the genome, larger than what is feasible with existing imaging-based omics methods (Shah et al. 2018; Eng et al. 2019; Xia et al. 2019). By combining with sequential immunofluorescence and transcriptome measurements, our imaging-based chromatin profiling can map spatial proximity of genomic loci to various subnuclear structures and chromatin marks using 100-200 kb bin sizes in a cell-type specific fashion. While our imaging-based approach does not capture individual molecular binding at cis-regulatory elements at <1-5 kb resolution (i.e., enhancer, promoter resolution of individual genes) as obtained in sequencing-based single-cell chromatin profiling (Zhu et al. 2021; Bartosovic & Castelo-Branco 2022) and imaging-based epigenomic MERFISH for targeted loci(Lu et al. 2022), our datasets uniquely contain a large number of chromatin marks for each of the DNA loci as well as inter-chromosomal interactions and radial positioning of the DNA loci within the same nucleus. These datasets with improved genomic coverage and diverse subnuclear markers uniquely allow us to perform integrated analysis and discover cell-type specific subnuclear organizations, their associated genomic loci, and their impact on 3D genome organization and gene regulation.

The high-resolution spatial multi-omics data revealed that repressive chromatin regions are globally organized in a more cell-type specific fashion compared to active chromatin regions in the adult mouse cerebellum, while subtle changes of genomic loci toward active subnuclear markers between cell types are tightly linked to cell-type specific gene expression programs. Specifically, we systematically show that transcriptionally active genomic loci of long genes (>200 kb), including those for neuronal identities such as synaptic processes (Koopmans et al. 2019) and those involved in brain disorders such as ASD (King et al. 2013; Zhao et al. 2018), tend to localize at the RNAPIISer5-P foci away from nuclear speckles. These observations about subnuclear organization at the level of individual long genes became possible with improved genomic coverage of two-layer DNA seqFISH+ developed in this study and were previously inaccessible with DNA seqFISH+ (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) or DNA-MERFISH (Su et al. 2020) or in situ genome sequencing(Payne et al. 2021) due to their genome-wide coverage (up to ~ 1 Mb).

The cell-type specific repressive and active subnuclear organizations are correlated with variations in the chromosomal structures in different cell types. In particular, we demonstrated that H4K20me3-stained territories, localized adjacent to pericentromeric heterochromatin (Linhoff et al. 2015), confine genomic loci enriched with a subset of specific gene families (e.g., Vmn, Zfp) with a high specificity. Within the adult mouse cerebellum, this confinement was seen at the nuclear interior in neuronal cells, but not Bergmann glia. In contrast, the same set of loci switched to the nuclear lamina association at the nuclear periphery in Bergmann glia cells. Thus, those genomic loci contribute to cell-type specific 3D genome organization, including radial positioning of chromosomes and inter-chromosomal interactions. Previously, H4K20me3 had been shown to be involved in chromosomal organization of olfactory receptor genes in the olfactory sensory neurons (Magklara et al. 2011; Clowney et al. 2012; Armelin-Correa et al. 2014; Monahan et al. 2019; Tan et al. 2019). Our results now reveal a broader role for H4K20me3 in organizing the 3D genome with a different set of genomic loci (e.g. Vmn, Skint gene clusters), in a cell-type-dependent manner. The H4K20me3 association in neurons may account for a wide-range of 3D genome organization, including the neuron-specific radial reorganization of

chromosomes during development (Tan et al. 2021; Tan et al. 2023) and long-range intra- and inter-chromosomal interactions in the other mouse brain regions (Winick-Ng et al. 2021; Tan et al. 2021). The H4K20me3 organization could also be disrupted in neurological and neurodegenerative disorders (Baltanás et al. 2011; Linhoff et al. 2015; Chomiak et al. 2022) and can be investigated further with the seqFISH-based high-resolution spatial multi-omics approach. Lastly, cell-type specific repressive chromatin organization may reflect regulatory history of cells and the path used to arrive at specific cell states (Letsou & Cai 2016).

We note several technical limitations of our spatial multi-omics measurements. First, the practical genomic resolution of the imaging-based chromatin profiling analysis, measuring spatial proximity of DNA loci and subnuclear markers rather than their molecular interactions, is fundamentally limited by the diffraction-limited immunofluorescence images (~200–300 nm spatial resolution), which can be further improved by super-resolution imaging of individual chromatin marks or protein molecules (Woodworth et al. 2021; Pownall et al. 2023). Second, to generate the chromatin profiles systematically across different immunofluorescence marks, we linearly normalized fluorescence intensities of each marker in each cell and determined whether a given DNA locus is enriched in a high intensity region of the immunofluorescence mark. This approach was chosen to be systematic, but can result in differences for marks with different dynamic ranges in the nuclei. Third, the two-layer barcoding scheme under-detects DNA loci near the chromosome block boundaries, which leads to block patterns in pairwise spatial distance calculations of DNA loci (Supplementary Notes) and potentially limits the ability to detect DNA loci with local DNA looping. However, we observed little effects on the cell-type specific chromatin profiles, but note that analyses examining how subnuclear markers and 3D chromosome structures co-vary in single cells could be affected. Lastly, while we did not observe a large-scale distortion of the nuclei, the fixation and in situ hybridization steps could perturb the smaller scale chromatin structures in single cells.

2.5 References

Abrahams, B.S. et al., 2013. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Molecular autism, 4(1), p.36.

Agasti, S.S. et al., 2017. DNA-barcoded labeling probes for highly multiplexed Exchange-PAINT imaging. Chemical science , 8(4), pp.3080–3091.

Agredo, A. & Kasinski, A.L., 2023. Histone 4 lysine 20 tri-methylation: a key epigenetic regulator in chromatin structure and disease. Frontiers in genetics, 14, p.1243395.

Ahanger, S.H. et al., 2021. Distinct nuclear compartment-associated genome architecture in the developing mammalian brain. Nature neuroscience, 24(9), pp.1235–1242.

Alexander, K.A. et al., 2021. p53 mediates target gene association with nuclear speckles for amplified RNA expression. Molecular cell, 81(8), pp.1666–1681.e6.

Amemiya, H.M., Kundaje, A. & Boyle, A.P., 2019. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Scientific reports, 9(1), p.9354.

Antebi, Y.E. et al., 2017. Combinatorial Signal Perception in the BMP Pathway. Cell, 170(6), pp.1184–1196.e24.

Armelin-Correa, L.M. et al., 2014. Nuclear compartmentalization of odorant receptor genes. Proceedings of the National Academy of Sciences of the United States of America, 111(7), pp.2782–2787.

Arrastia, M.V. et al., 2021. Single-cell measurement of higher-order 3D genome organization with scSPRITE. Nature biotechnology. Available at: http://dx.doi.org/10.1038/s41587-021-00998-1.

Baltanás, F.C. et al., 2011. Purkinje Cell Degeneration in pcd Mice Reveals Large Scale Chromatin Reorganization and Gene Silencing Linked to Defective DNA Repair*. The Journal of biological chemistry, 286(32), pp.28287–28302.

Bartosovic, M. & Castelo-Branco, G., 2022. Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. Nature biotechnology. Available at: http://dx.doi.org/10.1038/s41587-022-01535-4.

Becht, E. et al., 2018. Dimensionality reduction for visualizing single-cell data using UMAP. Nature biotechnology. Available at: http://dx.doi.org/10.1038/nbt.4314.

Beliveau, B.J. et al., 2012. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. Proceedings of the National Academy of Sciences of the United States of America, 109(52), pp.21301–21306.

Belmont, A.S., 2021. Nuclear Compartments: An Incomplete Primer to Nuclear Compartments, Bodies, and Genome Organization Relative to Nuclear Architecture. Cold Spring Harbor perspectives in biology. Available at: http://dx.doi.org/10.1101/cshperspect.a041268.

Bhat, P. et al., 2023. 3D genome organization around nuclear speckles drives mRNA splicing efficiency. bioRxiv, p.2023.01.04.522632. Available at: https://www.biorxiv.org/content/10.1101/2023.01.04.522632v1 [Accessed January 6, 2023].

Bhat, P., Honson, D. & Guttman, M., 2021. Nuclear compartmentalization as a mechanism of quantitative control of gene expression. Nature reviews. Molecular cell biology, 22(10), pp.653–670.

Bintu, B. et al., 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science, 362(6413). Available at: http://dx.doi.org/10.1126/science.aau1783.

Bizhanova, A. et al., 2020. Distinct features of nucleolus-associated domains in mouse embryonic stem cells. Chromosoma, 129(2), pp.121–139.

Bonev, B. et al., 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell, 171(3), pp.557–572.e24.

Boninsegna, L. et al., 2022. Integrative genome modeling platform reveals essentiality of rare contact events in 3D genome organizations. Nature methods, 19(8), pp.938–949.

Boyd, N., Schiebinger, G. & Recht, B., 2017. The Alternating Descent Conditional Gradient Method for Sparse Inverse Problems. SIAM Journal on Optimization. Available at: https://epubs.siam.org/doi/10.1137/15M1035793 [Accessed October 4, 2024].

Brueckner, L. et al., 2020. Local rewiring of genome-nuclear lamina interactions by transcription. The EMBO journal, 39(6), p.e103159.

Cerrato, V. et al., 2018. Sox2 conditional mutation in mouse causes ataxic symptoms, cerebellar vermis hypoplasia, and postnatal defects of Bergmann glia. Glia, 66(9), pp.1929–1946.

Chen, L.-F. et al., 2023. Structural elements promote architectural stripe formation and facilitate ultra-long-range gene regulation at a human disease locus. Molecular cell, 83(9), pp.1446–1461.e6.

Chen, X. et al., 2022. Transcriptomic mapping uncovers Purkinje neuron plasticity driving learning. Nature, 605(7911), pp.722–727.

Chen, Y. et al., 2018. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. The Journal of cell biology, 217(11), pp.4025–4048.

Chen, Y. & Belmont, A.S., 2019. Genome organization around nuclear speckles. Current opinion in genetics & development, 55, pp.91–99.

Chomiak, A.A. et al., 2022. Nde1 is required for heterochromatin compaction and stability in neocortical neurons. iScience, 25(6), p.104354.

Clowney, E.J. et al., 2012. Nuclear aggregation of olfactory receptor genes governs their monogenic expression. Cell, 151(4), pp.724–737.

Conte, M. et al., 2020. Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. Nature communications, 11(1), p.3289.

Dekker, J. et al., 2017. The 4D nucleome project. Nature, 549(7671), pp.219–226.

De Zeeuw, C.I., 2021. Bidirectional learning in upbound and downbound microzones of the cerebellum. Nature reviews. Neuroscience, 22(2), pp.92–110.

Di Tommaso, P. et al., 2017. Nextflow enables reproducible computational workflows. Nature biotechnology, 35(4), pp.316–319.

Eberhart, A. et al., 2012. Reliable detection of epigenetic histone marks and nuclear proteins in tissue cryosections. Chromosome research: an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology, 20(7), pp.849–858.

Edelstein, A. et al., 2010. Computer control of microscopes using μManager. Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.], Chapter 14, p.Unit14.20.

Eng, C.-H.L. et al., 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature, 568(7751), pp.235–239.

Ewels, P.A. et al., 2020. The nf-core framework for community-curated bioinformatics pipelines. Nature biotechnology, 38(3), pp.276–278.

Falk, M. et al., 2019. Heterochromatin drives compartmentalization of inverted and conventional nuclei. Nature, 570(7761), pp.395–399.

Finn, E.H. et al., 2019. Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. Cell, 176(6), pp.1502–1515.e10.

Gabel, H.W. et al., 2015. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature, 522(7554), pp.89–93.

Galliano, E. et al., 2018. Impact of NMDA Receptor Overexpression on Cerebellar Purkinje Cell Activity and Motor Learning. eNeuro, 5(1). Available at: http://dx.doi.org/10.1523/ENEURO.0270-17.2018.

Gong, H. et al., 2016. Simple Method To Prepare Oligonucleotide-Conjugated Antibodies and Its Application in Multiplex Protein Detection in Single Cells. Bioconjugate chemistry, 27(1), pp.217–225.

Gu, B. et al., 2018. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science, 359(6379), pp.1050–1055.

Hafner, A. et al., 2023. Loop stacking organizes genome folding from TADs to chromosomes. Molecular cell, 83(9), pp.1377–1392.e6.

Hershberg, E.A. et al., 2021. PaintSHOP enables the interactive design of transcriptome- and genome-scale oligonucleotide FISH experiments. Nature methods, 18(8), pp.937–944.

Hewitt, S.L. et al., 2008. Association between the Igk and Igh immunoglobulin loci mediated by the 3′ Igk enhancer induces' decontraction'of the Igh locus in pre--B cells. Nature immunology, 9(4), pp.396–404.

Hie, B., Bryson, B. & Berger, B., 2019. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nature biotechnology, 37(6), pp.685–691.

Hormoz, S. et al., 2016. Inferring Cell-State Transition Dynamics from Lineage Trees and Endpoint Single-Cell Measurements. Cell systems, 3(5), pp.419–433.e8.

Hounkpe, B.W. et al., 2021. HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. Nucleic acids research, 49(D1), pp.D947–D955.

Jerkovic, I. & Cavalli, G., 2021. Understanding 3D genome organization by multidisciplinary methods. Nature reviews. Molecular cell biology, 22(8), pp.511–528.

Johnson, D.S. et al., 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science, 316(5830), pp.1497–1502.

Jonkers, I., Kwak, H. & Lis, J.T., 2014. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife, 3, p.e02407.

Kaya-Okur, H.S. et al., 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nature communications, 10(1), p.1930.

Kempfer, R. & Pombo, A., 2020. Methods for mapping 3D chromosome architecture. Nature reviews. Genetics, 21(4), pp.207–226.

King, I.F. et al., 2013. Topoisomerases facilitate transcription of long genes linked to autism. Nature, 501(7465), pp.58–62.

Kinsella, R.J. et al., 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database: the journal of biological databases and curation, 2011, p.bar030.

Koopmans, F. et al., 2019. SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. Neuron, 103(2), pp.217–234.e4.

Kouzarides, T., 2007. Chromatin modifications and their function. Cell, 128(4), pp.693–705.

Kozareva, V. et al., 2021. A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. Nature, 598(7879), pp.214–219.

Kwak, Y.D. et al., 2021. Chromatin architecture at susceptible gene loci in cerebellar Purkinje cells characterizes DNA damage–induced neurodegeneration. Science Advances, 7(51), p.eabg6363.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature methods, 9(4), pp.357–359.

Leidescher, S. et al., 2022. Spatial organization of transcribed eukaryotic genes. Nature cell biology, 24(3), pp.327–339.

Letsou, W. & Cai, L., 2016. Noncommutative Biology: Sequential Regulation of Complex Networks. PLoS computational biology, 12(8), p.e1005089.

Lieberman-Aiden, E. et al., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 326(5950), pp.289–293.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics , 25(16), pp.2078–2079.

Lilue, J. et al., 2019. Mouse protein coding diversity: What's left to discover? PLoS genetics, 15(11), p.e1008446.

Lilue, J. et al., 2018. Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. Nature genetics, 50(11), pp.1574–1583.

Linhoff, M.W., Garg, S.K. & Mandel, G., 2015. A high-resolution imaging approach to investigate chromatin architecture in complex tissues. Cell, 163(1), pp.246–255.

Lister, R. et al., 2013. Global epigenomic reconfiguration during mammalian brain development. Science, 341(6146), p.1237905.

Liu, M. et al., 2020. Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. Nature communications, 11(1), p.2907.

Liu, S.-L. et al., 2013. Fast and high-accuracy localization for three-dimensional single-particle tracking. Scientific reports, 3, p.2462.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology, 15(12), p.550.

Lu, T., Ang, C.E. & Zhuang, X., 2022. Spatially resolved epigenomic profiling of single cells in complex tissues. Cell, 185(23), pp.4448–4464.e17.

MacDonald, J.L. & Roskams, A.J., 2008. Histone deacetylases 1 and 2 are expressed at distinct stages of neuro-glial development. Developmental dynamics: an official publication of the American Association of Anatomists, 237(8), pp.2256–2267.

Magklara, A. et al., 2011. An epigenetic signature for monoallelic olfactory receptor expression. Cell, 145(4), pp.555–570.

Marshall, J.L. et al., 2020. HyPR-seq: Single-cell quantification of chosen RNAs via hybridization and sequencing of DNA probes. Proceedings of the National Academy of Sciences of the United States of America, 117(52), pp.33404–33413.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal. Available at: http://journal.embnet.org/index.php/embnetjournal/article/view/200.

Meuleman, W. et al., 2013. Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. Genome research, 23(2), pp.270–280.

Monahan, K., Horta, A. & Lomvardas, S., 2019. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. Nature, 565(7740), pp.448–453.

Nagano, T. et al., 2017. Cell-cycle dynamics of chromosomal organization at single-cell resolution. Nature, 547(7661), pp.61–67.

Nguyen, H.Q. et al., 2020. 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. Nature methods, 17(8), pp.822–832.

Nichols, M.H. & Corces, V.G., 2021. Principles of 3D compartmentalization of the human genome. Cell reports, 35(13), p.109330.

O'Geen, H. et al., 2007. Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. PLoS genetics, 3(6), p.e89.

Pachitariu, M. & Stringer, C., 2022. Cellpose 2.0: how to train your own model. Nature methods, 19(12), pp.1634–1641.

Parthasarathy, R., 2012. Rapid, accurate particle tracking by calculation of radial symmetry centers. Nature methods, 9(7), pp.724–726.

Pascual-Reguant, L. et al., 2018. Lamin B1 mapping reveals the existence of dynamic and functional euchromatin lamin B1 domains. Nature communications, 9(1), p.3420.

Payne, A.C. et al., 2021. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. Science, 371(6532). Available at: http://dx.doi.org/10.1126/science.aay3446.

Pederson, T., 2011. The nucleolus. Cold Spring Harbor perspectives in biology, 3(3). Available at: http://dx.doi.org/10.1101/cshperspect.a000638.

Peric-Hupkes, D. et al., 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Molecular cell, 38(4), pp.603–613.

Pownall, M.E. et al., 2023. Chromatin expansion microscopy reveals nanoscale organization of transcription and chromatin. Science, 381(6653), pp.92–100.

Quinlan, A.R. & Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics , 26(6), pp.841–842.

Quinodoz, S.A. et al., 2018. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell, 174(3), pp.744–757.e24.

Quinodoz, S.A. et al., 2021. RNA promotes the formation of spatial compartments in the nucleus. Cell, 184(23), pp.5775–5790.e30.

Ramachandran, P. & Varoquaux, G., 2011. Mayavi: 3D Visualization of Scientific Data. Computing in science & engineering, 13(2), pp.40–51.

Ramírez, F. et al., 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic acids research, 44(W1), pp.W160–5.

Sadakata, T. et al., 2007. Impaired cerebellar development and function in mice lacking CAPS2, a protein involved in neurotrophin release. The Journal of neuroscience: the official journal of the Society for Neuroscience, 27(10), pp.2472–2482.

van Schaik, T. et al., 2020. Cell cycle dynamics of lamina-associated DNA. EMBO reports, 21(11), p.e50636.

van Schaik, T. et al., 2022. CTCF and cohesin promote focal detachment of DNA from the nuclear lamina. Genome biology, 23(1), p.185.

von Schimmelmann, M. et al., 2016. Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. Nature neuroscience, 19(10), pp.1321–1330.

Shah, S. et al., 2018. Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. Cell, 174(2), pp.363–376.e16.

Shah, S. et al., 2016. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron, 92(2), pp.342–357.

Skene, P.J., Henikoff, J.G. & Henikoff, S., 2018. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nature protocols, 13(5), pp.1006–1019.

Skene, P.J. & Henikoff, S., 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife, 6. Available at: http://dx.doi.org/10.7554/eLife.21856.

Solovei, I. et al., 2004. Positional changes of pericentromeric heterochromatin and nucleoli in postmitotic Purkinje cells during murine cerebellum development. Cytogenetic and genome research, 105(2-4), pp.302–310.

Solovei, I., Thanisch, K. & Feodorova, Y., 2016. How to rule the nucleus: divide et impera. Current opinion in cell biology, 40, pp.47–59.

van Steensel, B. & Belmont, A.S., 2017. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell, 169(5), pp.780–791.

Stevens, T.J. et al., 2017. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature, 544(7648), pp.59–64.

Stoyanova, E. et al., 2021. 5-Hydroxymethylcytosine-mediated active demethylation is required for mammalian neuronal differentiation and function. eLife, 10. Available at: http://dx.doi.org/10.7554/eLife.66973.

Stringer, C. et al., 2021. Cellpose: a generalist algorithm for cellular segmentation. Nature methods, 18(1), pp.100–106.

Stuart, T. et al., 2019. Comprehensive Integration of Single-Cell Data. Cell, 177(7), pp.1888–1902.e21.

Su, J.-H. et al., 2020. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. Cell, 182(6), pp.1641–1659.e26.

Svensson, V. et al., 2017. Power analysis of single-cell RNA-sequencing experiments. Nature methods, 14(4), pp.381–387.

Takei, Y., Yun, J., et al., 2021. Integrated spatial genomics reveals global architecture of single nuclei. Nature, 590(7845), pp.344–350.

Takei, Y. et al., 2017. Multiplexed Dynamic Imaging of Genomic Loci by Combined CRISPR Imaging and DNA Sequential FISH. Biophysical journal, 112(9), pp.1773–1776.

Takei, Y., Zheng, S., et al., 2021. Single-cell nuclear architecture across cell types in the mouse brain. Science, 374(6567), pp.586–594.

Tan, L. et al., 2021. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. Cell, 184(3), pp.741–758.e17.

Tan, L. et al., 2023. Lifelong restructuring of 3D genome architecture in cerebellar granule cells. Science, 381(6662), pp.1112–1119.

Tan, L. et al., 2019. Three-dimensional genome structures of single sensory neurons in mouse visual and olfactory systems. Nature structural & molecular biology, 26(4), pp.297–307.

Tasic, B. et al., 2016. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nature neuroscience, 19(2), pp.335–346.

Tomita, H. et al., 2020. The Protein Tyrosine Phosphatase Receptor Delta Regulates Developmental Neurogenesis. Cell reports, 30(1), pp.215–228.e5.

Traag, V.A., Waltman, L. & van Eck, N.J., 2019. From Louvain to Leiden: guaranteeing well-connected communities. Scientific reports, 9(1), p.5233.

Van't Sant, L.J. et al., 2021. In vivo 5-ethynyluridine (EU) labelling detects reduced transcription in Purkinje cell degeneration mouse mutants, but can itself induce neurodegeneration. Acta neuropathologica communications, 9(1), p.94.

Vieux-Rochas, M. et al., 2015. Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. Proceedings of the National Academy of Sciences of the United States of America, 112(15), pp.4672–4677.

Waisman, A. et al., 2021. Automatic and unbiased segmentation and quantification of myofibers in skeletal muscle. Scientific reports, 11(1), p.11793.

Walczak, A. et al., 2013. Novel higher-order epigenetic regulation of the Bdnf gene upon seizures. The Journal of neuroscience: the official journal of the Society for Neuroscience, 33(6), pp.2507–2511.

Wang, S. et al., 2016. Spatial organization of chromatin domains and compartments in single chromosomes. Science, 353(6299), pp.598–602.

Wang, Y. et al., 2021. SPIN reveals genome-wide landscape of nuclear compartmentalization. Genome biology, 22(1), p.36.

Winick-Ng, W. et al., 2021. Cell-type specialization is encoded by specific chromatin topologies. Nature, 599(7886), pp.684–691.

Wolf, F.A., Angerer, P. & Theis, F.J., 2018. SCANPY: large-scale single-cell gene expression data analysis. Genome biology, 19(1), p.15.

Woodworth, M.A. et al., 2021. Multiplexed single-cell profiling of chromatin states at genomic loci by expansion microscopy. Nucleic acids research. Available at: http://dx.doi.org/10.1093/nar/gkab423.

Xia, C. et al., 2019. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proceedings of the National Academy of Sciences of the United States of America, 116(39), pp.19490–19499.

Yang, M. et al., 2022. Chemical-induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. Cell stem cell. Available at: http://dx.doi.org/10.1016/j.stem.2022.01.010.

Yu, G. et al., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology, 16(5), pp.284–287.

Zhang, L. et al., 2020. TSA-seq reveals a largely conserved genome organization relative to nuclear speckles with small position changes tightly correlated with gene expression changes. Genome research. Available at: http://dx.doi.org/10.1101/gr.266239.120.

Zhang, W. et al., 2019. Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. Nucleic acids research, 47(16), pp.8485–8501.

Zhao, Y.-T. et al., 2018. Long genes linked to autism spectrum disorders harbor broad enhancer-like chromatin domains. Genome research, 28(7), pp.933–942.

Zhao, Z. et al., 2023. Mega-Enhancer Bodies Organize Neuronal Long Genes in the Cerebellum. bioRxiv : the preprint server for biology. Available at: http://dx.doi.org/10.1101/2023.07.19.549737.

Zhu, C. et al., 2021. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nature methods, 18(3), pp.283–292.

2.6 Main figures

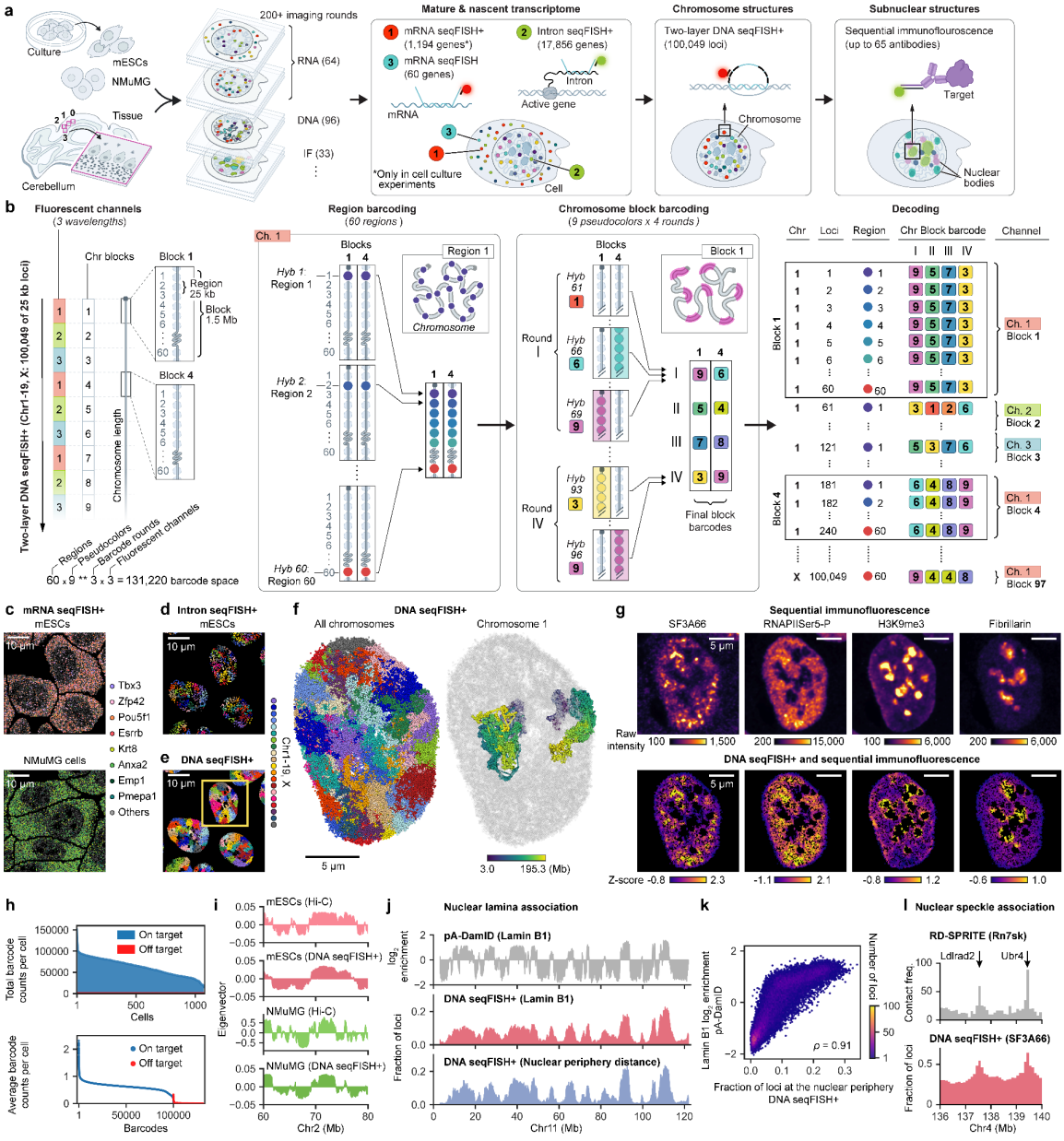


Fig. 1| Development and validation of single-cell spatial multi-omics technology.

a, Schematic of spatial multi-omics, imaging mature and nascent transcriptome as well as chromosome and subnuclear structures within the same cell. **b**, Detailed schematic of two-layer DNA seqFISH+ barcoding strategy. Chromosomes are splitted into 1.5-Mb blocks, in which each 25-kb locus is imaged as a diffraction limited spot during the first 60 rounds of

imaging with three orthogonal fluorescent channels. Then each 1.5-Mb block is uniquely barcoded across 4 rounds with 9-pseudocolors during the subsequent 36 rounds of imaging. Finally, by decoding the region and block identities, 100,049 of 25-kb loci can be uniquely resolved across 20 chromosomes. One barcoding round was redundant and used for stringent barcode decoding. **c**, Representative visualization of decoded mRNA spots for marker genes in single cells for mESCs (top) and NMuMG cells (bottom). **d**, **e**, Representative visualization of decoded intron or DNA locus spots in single nuclei by intron seqFISH+ (**d**) or two-layer DNA seqFISH+ (**e**) colored by chromosome identities (Chr1-19, X). Same mESCs are shown in **c**. **f**, Representative 3D visualization of decoded DNA locus spots in the single nucleus of mESCs. The DNA loci are colored by chromosome identities (left) similarly to **e** or chromosome coordinates (right). The selected nucleus is highlighted with a yellow box in **e**. **g**, Representative immunofluorescence raw images (top) and decoded DNA locus spots colored by z-score normalized immunofluorescence intensity of the corresponding marker (bottom) in the same nucleus of mESCs in **f**. **h**, Total on- and off-target barcode counts per cell (top). Average counts of individual barcode per cell (bottom). n = 100,049 and 31,171 DNA loci for on- and off-target barcodes by two-layer DNA seqFISH+. **i**, Representative A/B compartment eigenvector profiles between Hi-C (Bonev et al. 2017; Pascual-Reguant et al. 2018) and two-layer DNA seqFISH+ for mESCs and NMuMG cells. **j**, Representative visualization of Lamin B1 chromatin profiles by pA-DamID (van Schaik et al. 2022) (top) and two-layer DNA seqFISH+ (middle), and fraction of loci within 500 nm from the nuclear periphery (bottom) in mESCs. **k**, Genome-wide comparison of fraction of DNA loci at the nuclear periphery by DNA seqFISH+ and Lamin B1 enrichment over control by pA-DamID (van Schaik et al. 2022) with 100 kb binning (n = 25,110 loci). Spearman correlation coefficient of 0.91. **l**, Representative visualization of chromatin profiles of nuclear speckle markers by RD-SPRITE (Quinodoz et al. 2021) (top) and two-layer DNA seqFISH+ (bottom). The same 100-kb bin peaks detected by both methods are highlighted by black arrows and gene names. n = 1,076 cells from two biological replicates of mESCs in **c-l** and n = 384 cells from one biological replicate of NMuMG cells in **c**, **i**.

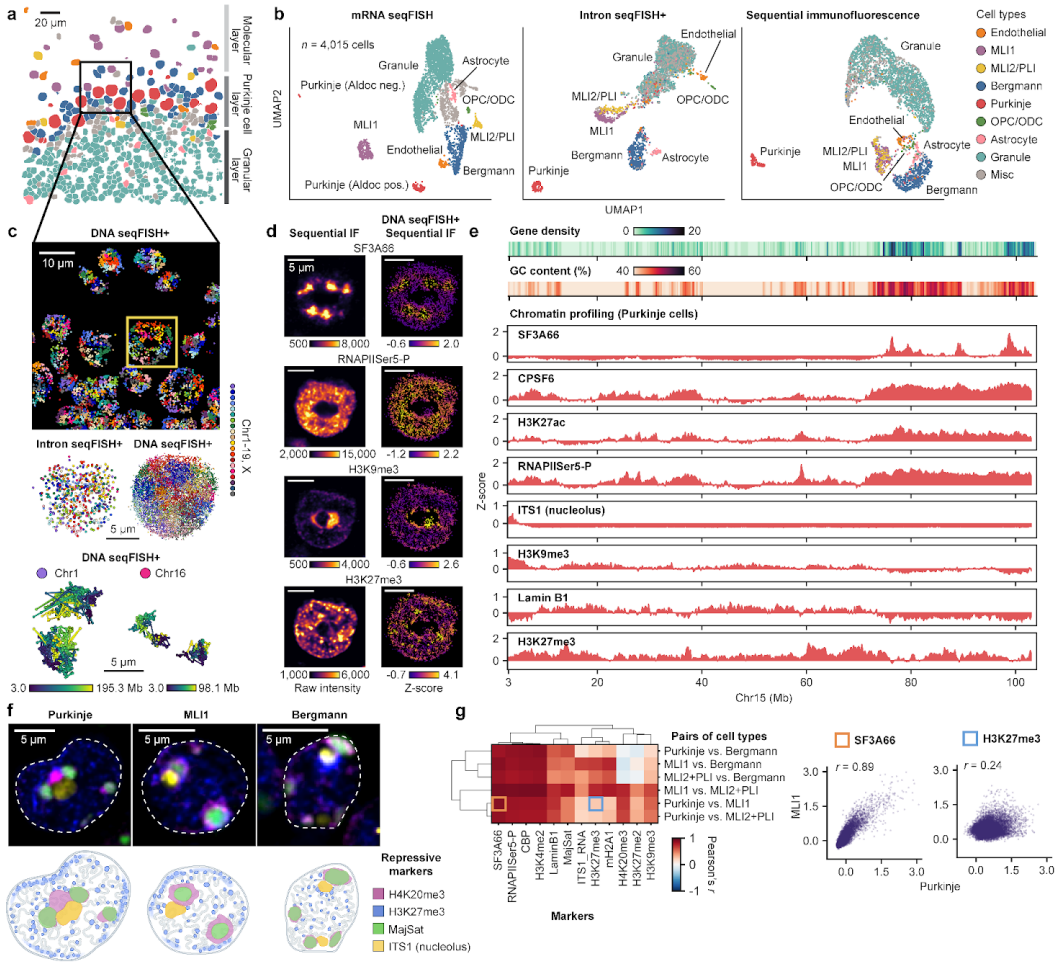


Fig. 2| Single-cell spatial multi-omics in the adult mouse brain cerebellum.

a, Spatial distribution of cell type clusters from a single z-section of the adult mouse cerebellum. **b**, Cell type clusters determined by mRNA seqFISH projected onto UMAP-embedding of mRNA seqFISH (left), intron seqFISH+ (middle), sequential immunofluorescence (right). **c**, Decoded DNA locus spots (top) from a single z-section of cells in the black box in **a**. Zoomed-in 3D views of decoded intron spots (middle left) and DNA locus spots (middle right, bottom) in the nucleus from a yellow-boxed Purkinje cell (top). The decoded spots are colored by chromosome identities (top, middle) or chromosome coordinates (bottom). **d**, Immunofluorescence raw images (left) and DNA spots colored by z-score normalized immunofluorescence intensity (right) in the nucleus from a cell highlighted in **c**. **e**, Genomic features and representative imaging-based

chromatin profiling in Purkinje cells in chromosome 15. **f**, Representative raw immunofluorescence images from a single z-section for each cell type (top). Illustration showing cell-type specific organization of repressive markers (bottom). The intense H3K27me3 cluster visible in Bergmann glia, representing the inactive X chromosome territory (Linhoff et al. 2015), is not depicted. **g**, Degree of similarity of chromatin profiles between pairs of cell types (left) and corresponding examples (right). *n* = 12,562 loci. 200 kb binning was used for the visualization and analysis (**e**, **g**). *n* = 4,015 cells from two biological replicates of the adult mouse cerebellum in **a**-**g**.

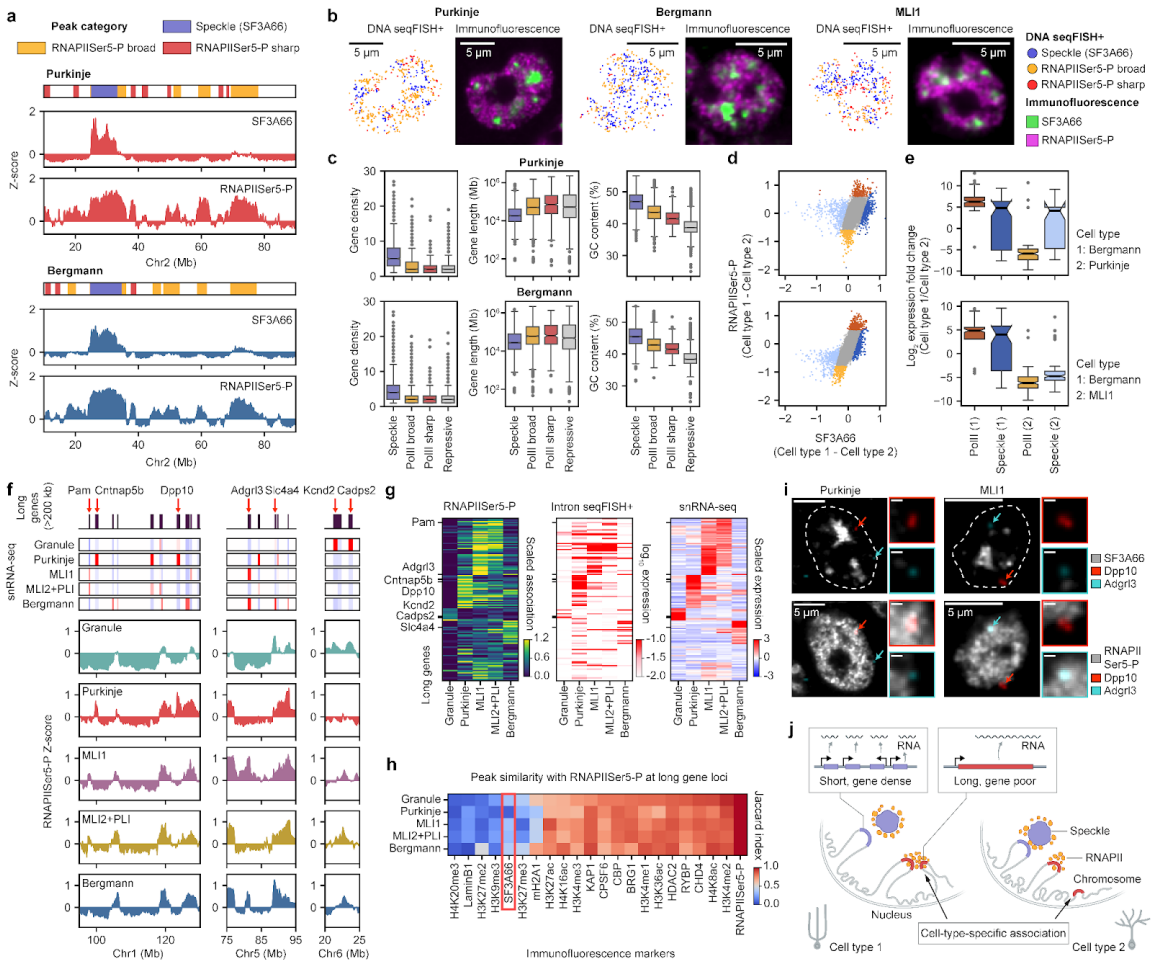


Fig. 3| Distinct active chromatin markers organize transcriptionally-active genomic loci.

a, Representative classification of three types of active domains (nuclear speckle, RNAPIISer5-P broad and sharp) in Purkinje cells and Bergmann glia along with the chromatin profiling of SF3A66 (nuclear speckle marker) and RNAPIISer5-P. **b**, Visualization of genomic loci colored by the ensemble-average active domain classification in **a** (left) and corresponding raw immunofluorescence image (right) from a single z-section for each cell type. **c**, Comparison of genomic features across different active domains in each cell type. *n* = 1,405, 4,102, 679, 4,892 loci (Purkinje) and 3,161, 3,115, 411, 4,080 loci (Bergmann) from left to right category. **d**, Comparison of differential association of genomic loci with SF3A66 and RNAPIISer5-P between pairs of cell types. **e**, Comparison

of differential mRNA expression between pairs of cell types at differentially associated loci for either PolII (RNAPIISer5-P) or Speckle (SF3A66) in each cell type, colored in **d**. *n* = 110, 55, 121, 94 loci (top) and 134, 94, 109, 67 loci (bottom) from left to right category. In box plots, the center lines for the median, boxes for the interquartile range, whiskers for values within 1.5 times the interquartile range, and points for outliers (**c**, **e**). **f**, Representative genomic regions with long genes (>200 kb) (top), corresponding mRNA expression(Kozareva et al. 2021) (middle), and RNAPIISer5-P chromatin profiles (bottom) in each cell type. **g**, Heatmaps for RNAPIISer5-P chromatin profiles (left), nascent RNA expression (middle), and mRNA expression (Kozareva et al. 2021) (right) for highly correlated long genes (*n* = 132 genes, Methods). **h**, Similarity of RNAPIISer5-P peaks with other markers on the long genes in **g** in each cell type. **i**, Representative single cell visualization of long gene loci with cell-type specific gene expression (Dpp10 in Purkinje cells and Adgrl3 in MLI1), relative to nuclear speckles (SF3A66) and RNAPIISer5-P with a maximum z-projection of two z-sections. Scale bars, 500 nm. **j**, Illustration showing nuclear speckle and RNAPIISer5-P subnuclear markers associated with distinct genomic loci in a cell-type specific fashion. 200 kb binning (*n* = 12,562 loci in total) was used for the analysis and visualization (**a**, **c**-**h**). *n* = 2,336, 128, 263, 88, and 518 cells for Granule, Purkinje, MLI1, MLI2+PLI, and Bergmann glia cells from two biological replicates of the adult mouse cerebellum in **a**-**i**.

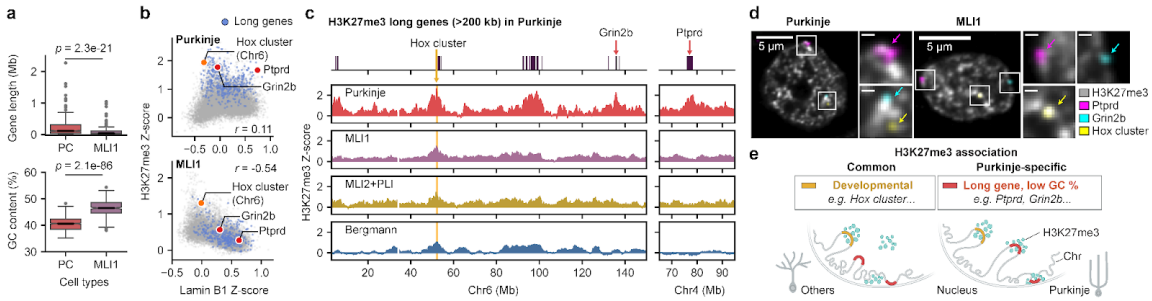


Fig. 4| Cell-type specific organization of H3K27me3-associated chromatin regions.

a, Comparison of genomic features between loci that were differentially associated with H3K27me3 in two cell types. n = 600, 260 loci for Purkinje cells (PC) and MLI1. p values by two-sided Wilcoxon's signed rank-sum test. The center lines for the median, boxes for the interquartile range, whiskers for values within 1.5 times the interquartile range, and points for outliers. **b**, Comparison of Lamin B1 and H3K27me3 chromatin profiles in each cell type. r represents a Pearson correlation coefficient. **c**, H3K27me3 profiles across cell types, highlighted with Hox cluster. Identified Purkinje-specific long gene loci associated with H3K27me3 are shown as blue dots (**b**) and binary heatmap (**c**). **d**, Representative single cell visualization of genomic loci highlighted (**b**, **c**), relative to H3K27me3. Scale bars, 500 nm. **e**, Illustration showing H3K27me3 subnuclear foci associated with common and Purkinje-specific genomic loci at the nuclear interior or periphery. 200 kb binning (n = 12,562 loci in total) was used for the analysis and visualization (**a**-**c**). n = 128, 263, 88, and 518 cells for Purkinje, MLI1, MLI2+PLI, and Bergmann glia cells from two biological replicates of the adult mouse cerebellum in **a**-**d**.

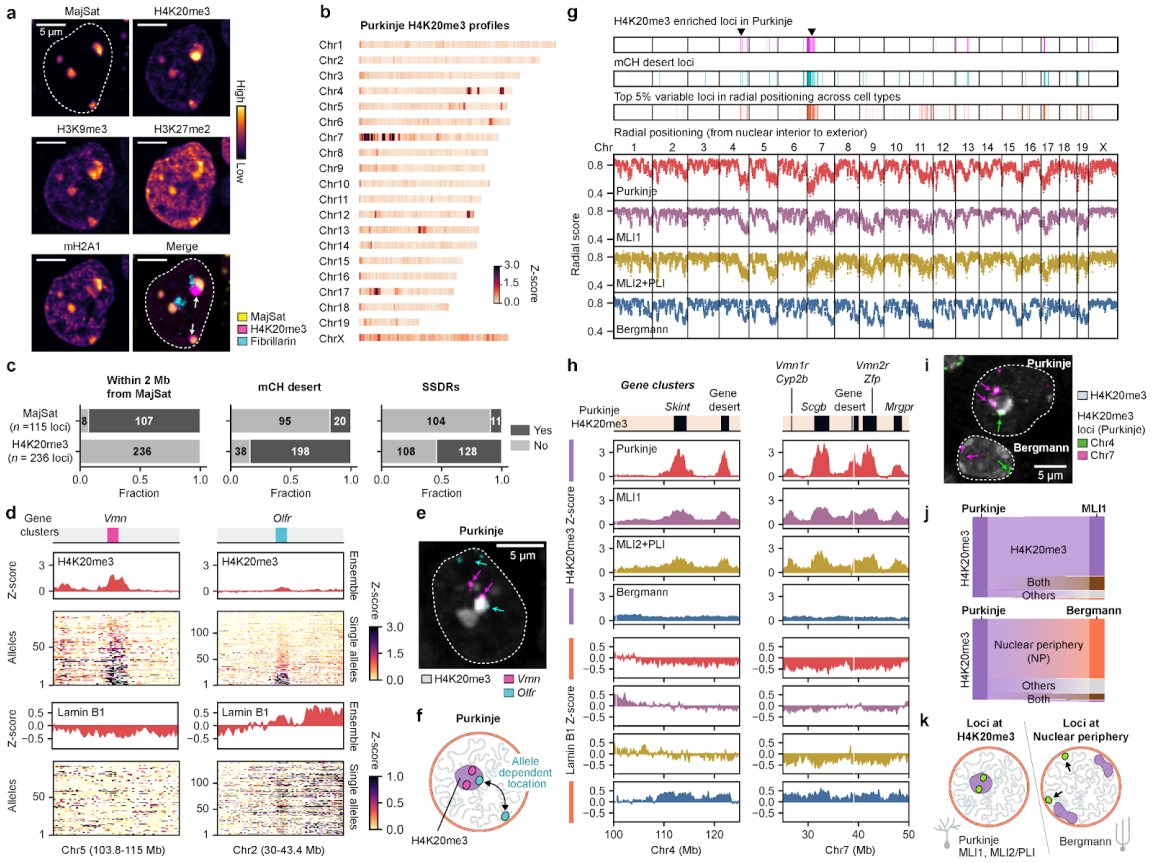


Fig. 5| The H4K20me3-marked territories confine specific genomic loci in neurons.

a, DNA FISH (MajSat) and sequential immunofluorescence images for markers enriched near pericentromeric repressive heterochromatin in Purkinje cells. **b**, H4K20me3 enrichment across chromosomes in Purkinje cells. **c**, Barplots comparing the locus characteristics such as mCH desert (Lister et al. 2013) and SSDRs(Lilue et al. 2018; Lilue et al. 2019) between MajSat- and H4K20me3-enriched loci in Purkinje cells. **d**, Ensemble-averaged and single allele chromatin profiles sorted by H4K20me3 enrichment from bottom to top in Purkinje cells. **e**, Visualization of H4K20me3-enriched *Vmn* and *Olfr* gene family loci with H4K20me3 staining. **f**, Illustration showing the differences of subnuclear localization between *Vmn* (magenta) and *Olfr* (cyan) gene family loci. **g**, Comparison of other genomic features (top) along with radial positioning of chromosomal loci from nuclear interior to exterior across cell types (bottom). **h**, H4K20me3 and Lamin B1 chromatin profiles at the H4K20me3-enriched regions, highlighted by triangles (**g**). **i**,

Visualization of H4K20me3 enriched loci in Chr4 and Chr7 (**g**) overlaid on the H4K20me3 immunofluorescence image in Purkinje and Bergmann cells. **j**, Transition of H4K20me3-enriched loci from Purkinje cells to MLI1 or Bergmann glia. *n* = 252 loci. **k**, Illustration showing the localization switching of genomic loci between neurons and Bergmann glia. 200 kb binning (*n* = 12,562 loci in total) was used for the analysis and visualization. *n* = 128, 263, 88, and 518 cells for Purkinje, MLI1, MLI2+PLI, and Bergmann glia cells from two biological replicates of the adult mouse cerebellum in **a**-**e**, **g**-**j**.

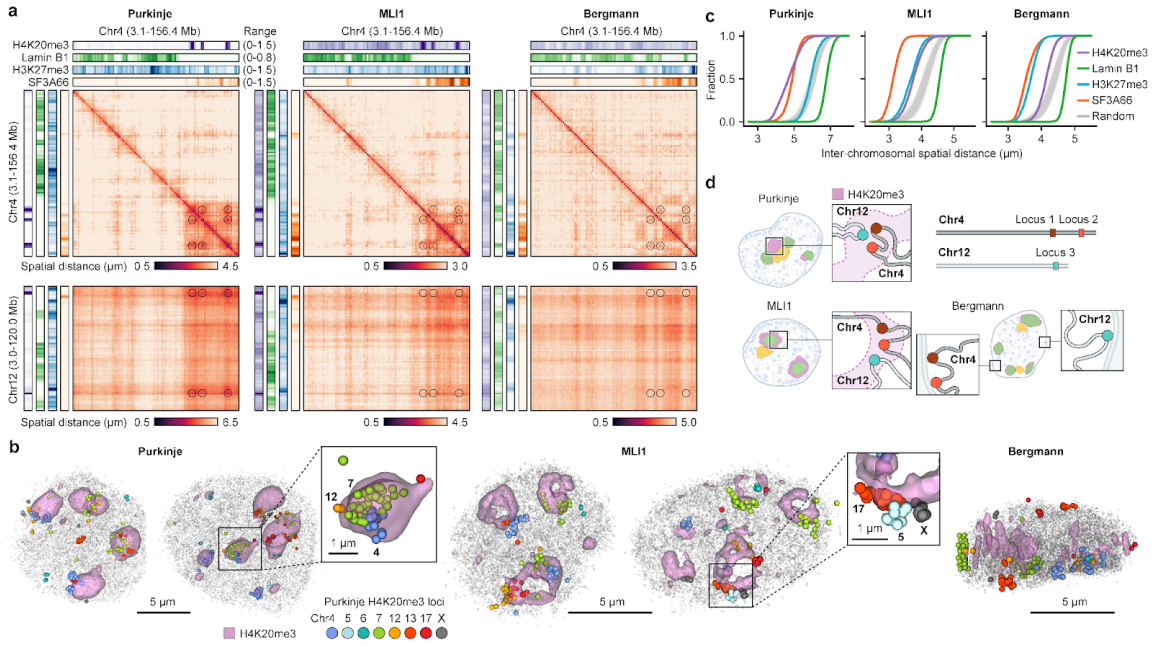
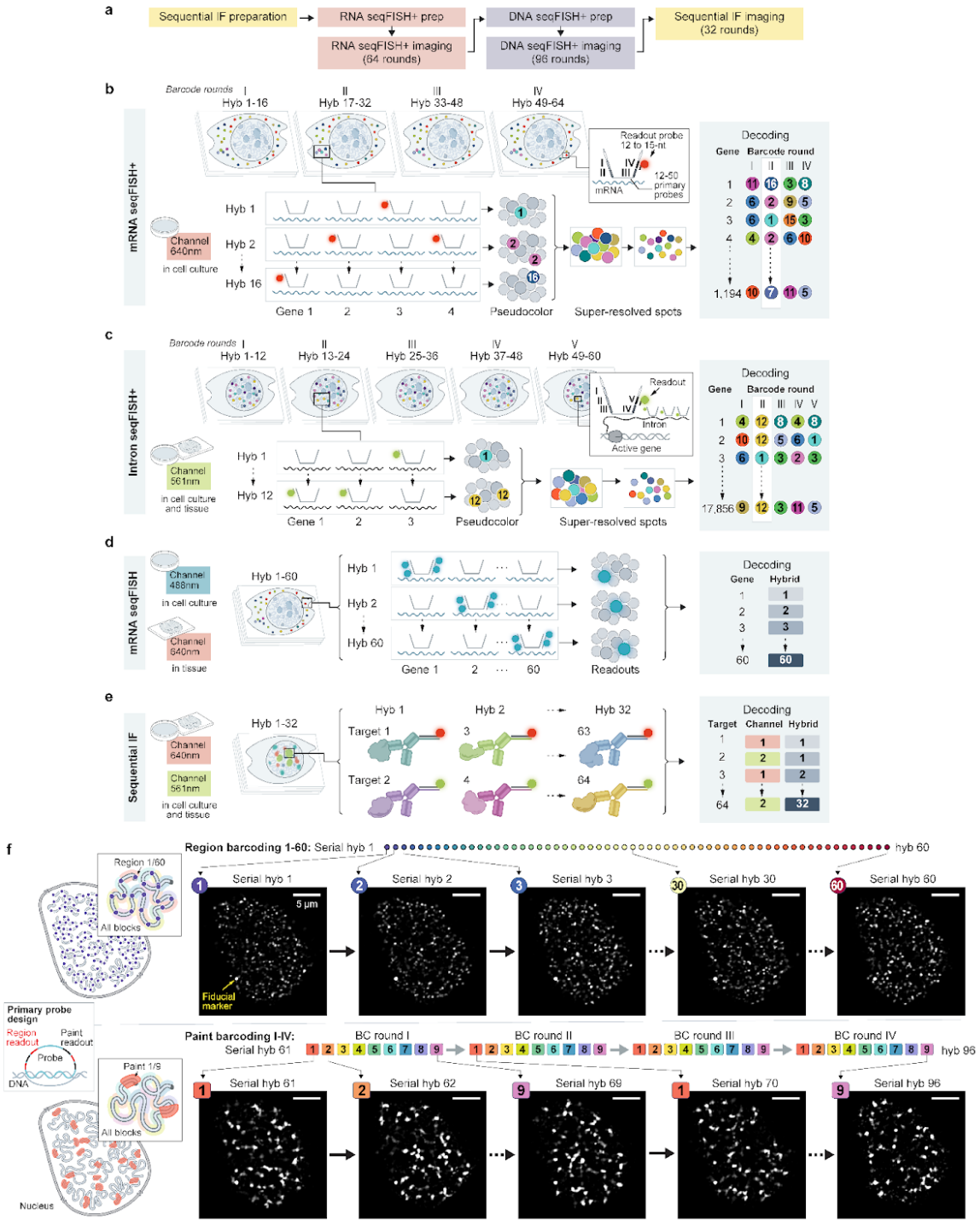


Fig. 6| Cell-type specific subnuclear structures and 3D genome organization.

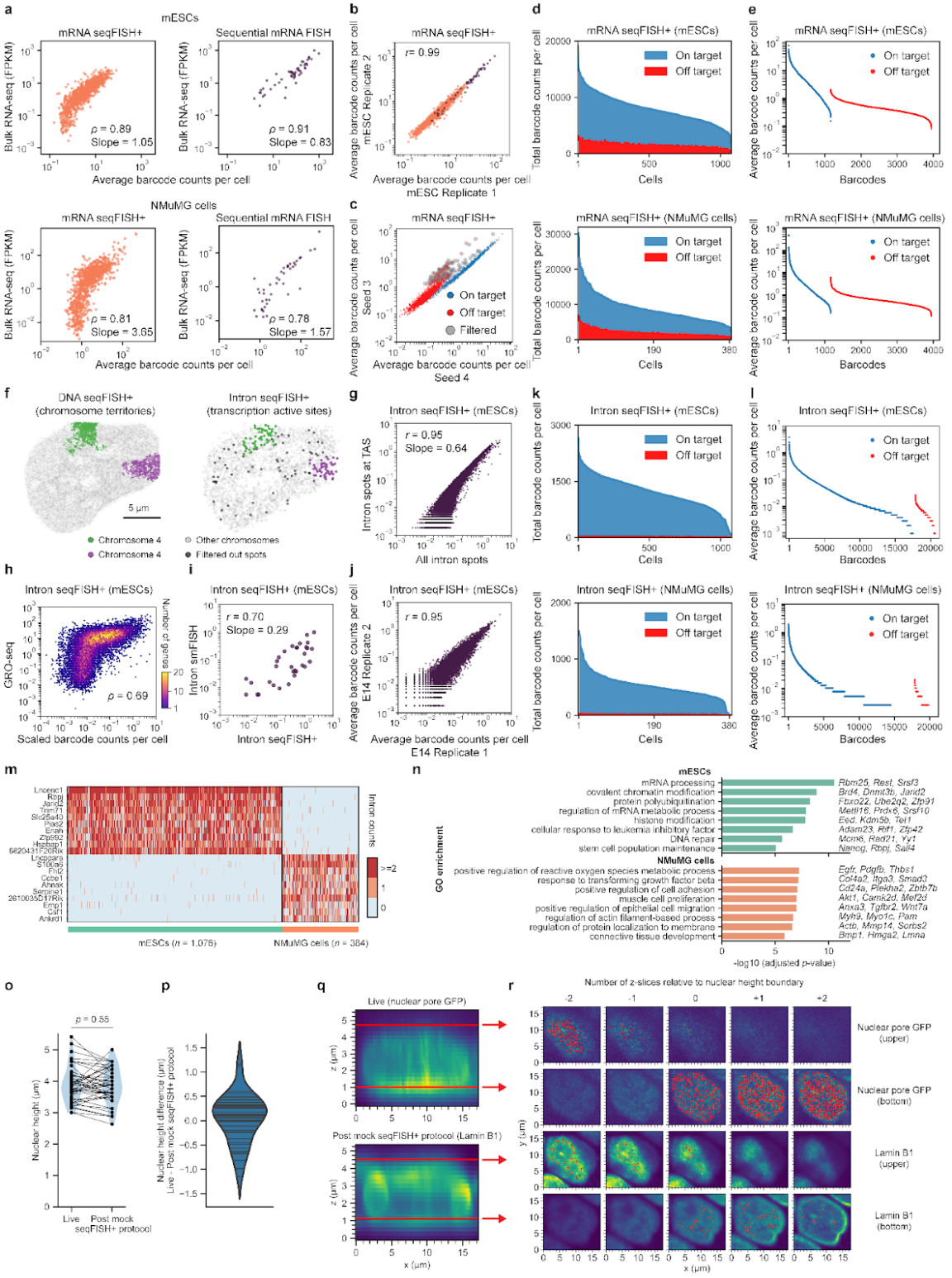
a, Ensemble-averaged spatial distances between pairs of genomic loci along with chromatin profiling with 1.5 Mb bins from each cell type. Locations of pairs of H4K20me3-associated loci in Purkinje cells in Fig. 5c are highlighted as black circles. **b**, Representative 3D images of the nucleus with H4K20me3 staining and chromosomal loci in each cell type. The H4K20me3-associated loci in Purkinje cells identified in Fig. 5c are highlighted as colored spots. **c**, Cumulative distribution of inter-chromosomal distances between pairs of loci with top 5% association to a given marker compared to random pairs of loci (n = 1,000 trials). 1.5 Mb binning (n = 1,678 loci in total), grouped by the chromosome paint block barcodes, was used. **d**, Illustration showing the differences of cell-type specific intra- and inter-chromosomal spatial arrangements around H4K20me3-enriched subnuclear territories in neurons or at the nuclear periphery in Bergmann glia. n = 128, 263, and 518 cells for Purkinje, MLI1, and Bergmann glia cells from two biological replicates of the adult mouse cerebellum in **a**-**c**.

2.7 Extended data figures



Extended Data Fig. 1| Detailed schematic of single-cell spatial multi-omics.

a, Flow chart of the sample preparation and imaging in single-cell spatial multi-omics. **b**, Schematic of mRNA seqFISH+ with 4 barcoding rounds with 16-pseudocolors including one round of error correction using one fluorescent channel (640 nm). This coding scheme can resolve up to 4,096 barcodes, while a subset of 1,194 barcodes were used to resolve mRNA species in cell culture experiments. **c**, Schematic of intron seqFISH+ with 5 barcoding rounds with 12-pseudocolors including one round of error correction using one fluorescent channel (561 nm). This coding scheme can resolve up to 20,736 barcodes, while a subset of 17,856 barcodes were used to resolve intronic RNA species in cell culture and tissue experiments. **d**, Schematic of non-barcoded mRNA seqFISH using one fluorescent channel (488 nm in cell culture experiments; 640 nm in tissue experiments). Unlike the exponential barcoding scheme in **b** and **c** whose coding capacity increases exponentially to the number of barcoding rounds, the number of RNA species that can be distinguished increases linearly to the number of hybridization rounds. **e**, Schematic of sequential immunofluorescence using two fluorescent channels (640 nm and 561 nm) in cell culture and tissue experiments. Similar to **d**, the number of antibody species that can be multiplexed scales linearly to the number of hybridization rounds. **f**, Implementation of DNA seqFISH+ (left), by leveraging the combinations of region barcoding (top) and chromosome block barcoding (bottom). To implement this, each primary probe (left, middle) contained three identical 15-nt readout binding sites (black) for one of the 60 regions in hybridizations 1-60 as well as three 15-nt readout binding sites (red) for three out of four chromosome paint barcoding rounds (hybridizations 61-96) in each fluorescent channel. Representative images for the nucleus from mESCs across 96 rounds of serial hybridization and imaging of two-layer DNA seqFISH+ (right) for region barcoding (top) and chromosome block barcoding (bottom). A fiducial marker targeting a locally repetitive 3632454L22Rik locus by a single DNA FISH probe (Takei, Yun, et al. 2021) appears in all rounds of imaging. Background subtracted images used for an analysis (Methods) are shown from a single z-section for visual clarity.

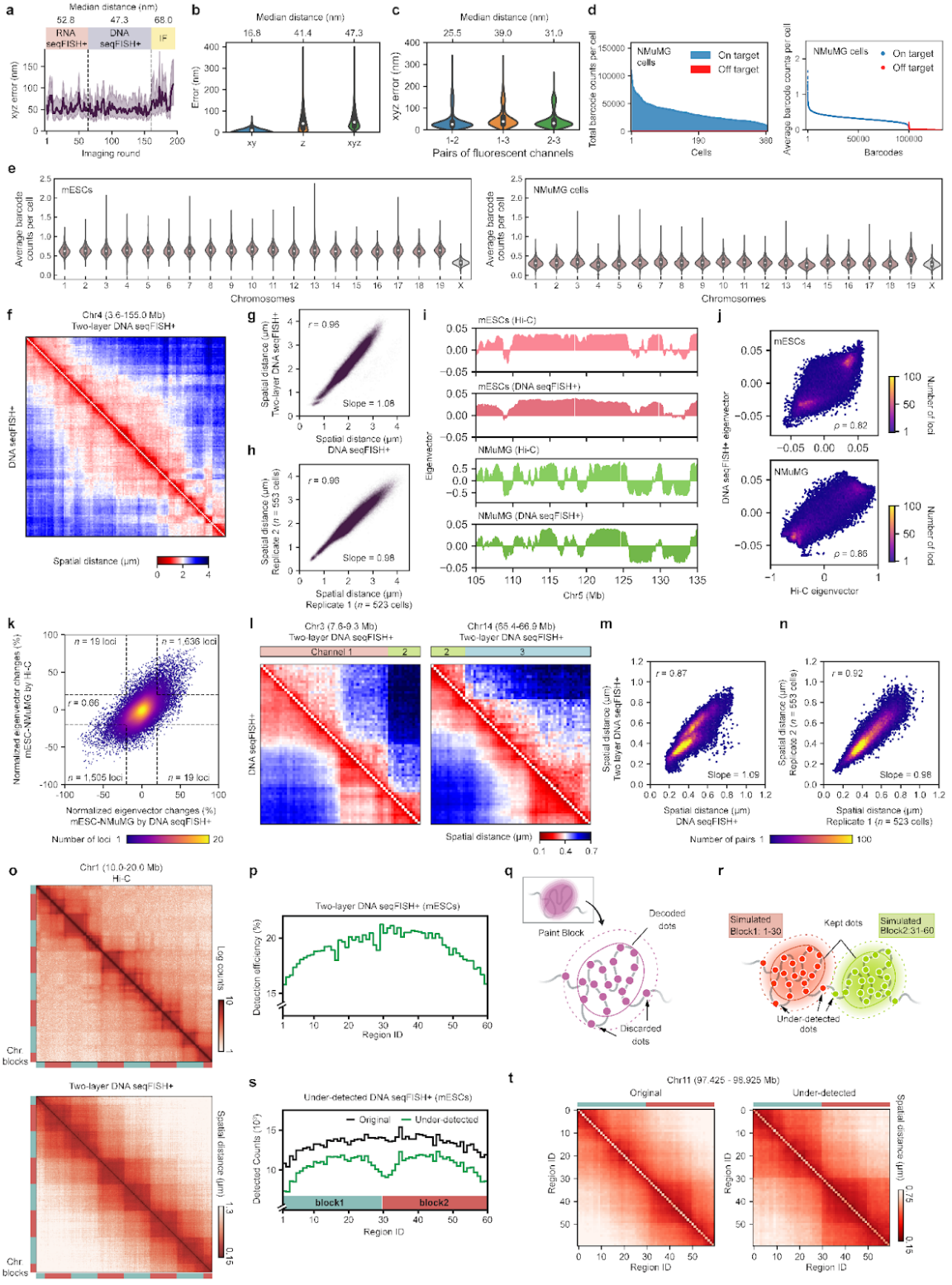


Extended Data Fig. 2| Validation of mRNA and intron seqFISH+ in cell culture

experiments.

a, Spearman correlation between average mRNA counts by mRNA seqFISH+ (left) or non-barcoded sequential mRNA FISH (right) and bulk RNA-seq in mESCs(Antebi et al. 2017) (top) and NMuMG cells(Pascual-Reguant et al. 2018) (bottom). By comparing the slope values for mRNA seqFISH+ and sequential mRNA FISH relative to bulk RNA-seq in each cell line, we estimated the detection efficiency of mRNA seqFISH+ as 78.9% and 43.0% in mESCs and NMuMG cells, respectively. **b**, Pearson correlation of average mRNA counts profiled by mRNA seqFISH+ (orange) and non-barcoded mRNA seqFISH (purple) between two biological replicates from mESCs. n = 1,194 and 53 genes for mRNA seqFISH+ and non-barcoded mRNA seqFISH profiling, respectively, in **a**, **b**. **c**, Visual representation of on- and off-target barcode counts and filtered barcodes by comparing mRNA seqFISH+ results from mESCs between seeds 3 and 4 decoding stringency (Methods). Those filtered barcodes (n = 150 barcodes, including both on- and off-target barcodes) were excluded from the downstream analysis. **d**, Total on- and off-target barcodes detected per cell by mRNA seqFISH+ in mouse ES cells (top) and NMuMG cells (bottom). **e**, Average counts per each on- and off-target barcode per cell by mRNA seqFISH+ in mESCs (top) and NMuMG cells (bottom). n = 1,163, 2,783, and 150 for on-target, off-target, and filtered barcodes, respectively in **c**-**e**. **f**, Representative visualization of homologous chromosomes (chromosome 4) by DNA seqFISH+ and transcription active sites by intron seqFISH+ near chromosome territories in the nucleus of mESCs. Intron spots detected within 500 nm from chromosome territories of their own chromosome captured by DNA seqFISH+ were considered to be transcription active sites, and other spots were filtered out from the downstream analysis. **g**, Pearson correlation between average intron counts per cell computed from all intron spots and transcription active sites defined by the criteria in **f**. **h**, Spearman correlation between scaled intron counts at transcription active sites by intron seqFISH+ and bulk GRO-seq (Jonkers et al. 2014), measuring the amount of transcriptionally active RNA polymerase II (RNAPII), in mESCs. **i**, Pearson correlation of average intron counts at the transcription active sites by intron seqFISH+ and non-barcoded intron seqFISH for n = 33 genes (Shah et al. 2018) in mESCs. The slope of 0.291 indicates a relative detection efficiency of 29.1%. **j**, Pearson correlation of average barcode counts at

transcription active sites per cell by intron seqFISH+ between two biological replicates from mouse ES cells. *n* = 17,856 genes for intron seqFISH+ in **g**, **h**, **j**. **k**, Total barcode counts at transcription active sites per cell by intron seqFISH+ in mESCs (top) and NMuMG cells (bottom). **l**, Average counts at transcription active sites per each on- and off-target barcode per cell by intron seqFISH+ in mESCs (top) and NMuMG cells (bottom). *n* = 17,856 and 2,880 for on- and off-target barcodes, respectively in **k** and **l**. **m**, Heatmap of gene expression profiles across single cells for top differentially expressed genes between mESCs and NMuMG cells by intron seqFISH+ (*n* = 10 genes for each cell line). **n**, The GO terms (top five plus manually selected three) identified from intron seqFISH+ profiles between mESCs and NMuMG cells represent their corresponding cell type identities. Three genes from each term were randomly selected for a display. **o**, Violin plot overlaid with reaction norm plot showing the paired distribution of nuclear heights of mESCs in live (left) and fixed cells after a mock seqFISH protocol (right). *P* value by a two-sided paired t-test. **p**, Violin plot showing the distribution of differences of nuclear heights between live cells and after a mock seqFISH+ protocol. Horizontal lines represent the height differences of individual nuclei. **q**, Side views of a representative nuclear height from live cells (top) and fixed cells after a mock seqFISH+ protocol (bottom). Horizontal red lines mark the determined top and bottom of the cells. While live, the nucleus was imaged using a nuclear pore (Nup37) GFP transgene. After the mock protocol, the nucleus was imaged using Lamin B1 immunofluorescence. We note that subnuclear staining enrichment patterns may differ between nuclear pore and Lamin B1 targets. **r**, Each row shows overlays of z-slices and coordinates of fit point spread functions (rounded to the nearest z-slice) for the range of z-slices 2 below to 2 above a boundary of the cell illustrated in **r**. *n* = 1,076 cells from two biological replicates of mESCs and *n* = 384 cells from one biological replicate of NMuMG cells in **a**-**p**. *n* = 43 cells from two biological replicates of Nup37-GFP mESCs for live cell experiments in **o**-**r**.

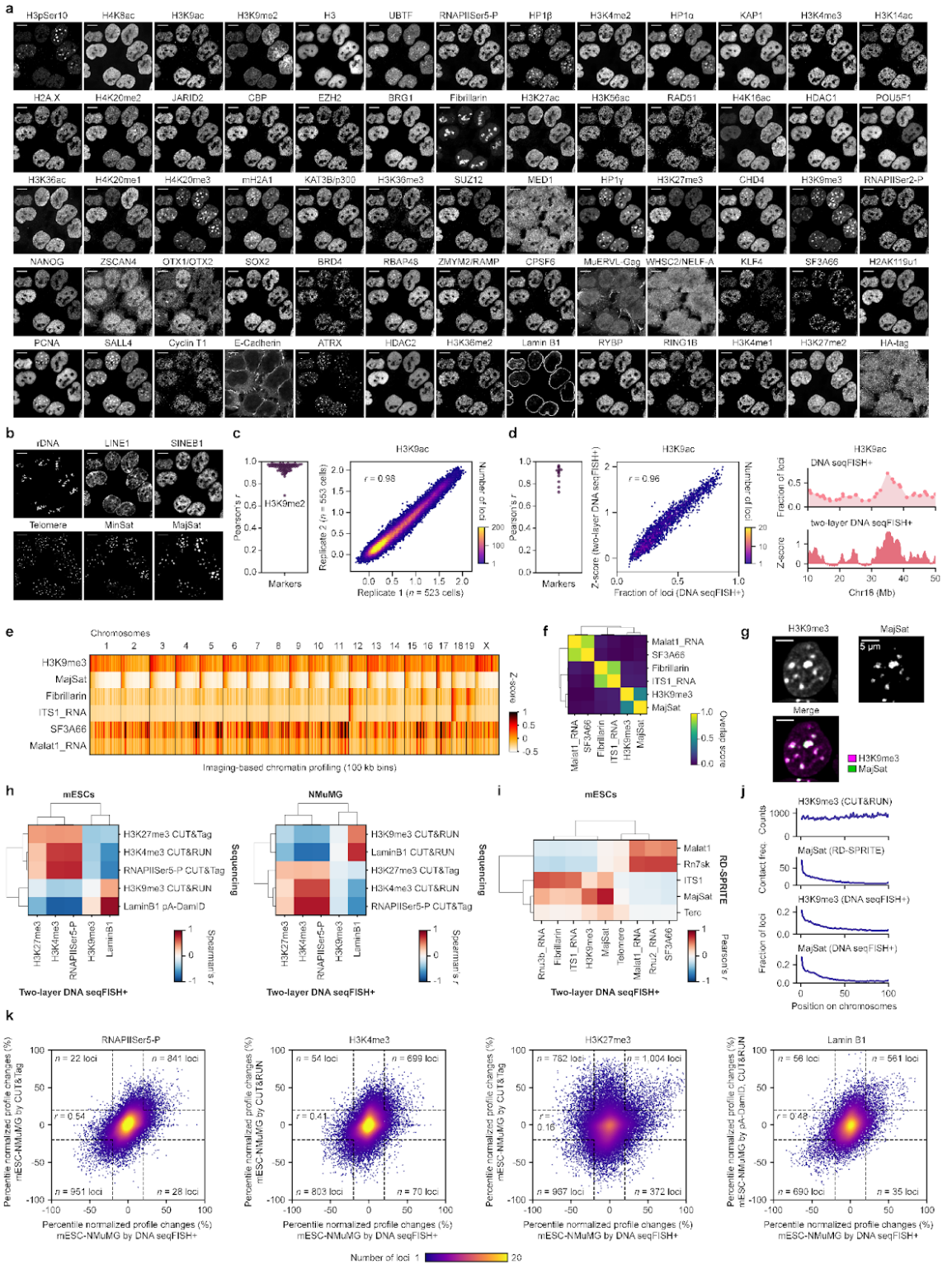


Extended Data Fig. 3| Validation of DNA seqFISH+ in cell culture experiments.

a, Quantification of the median fiducial marker localization relative to the reference image across 96 rounds of two-layer DNA seqFISH+ imaging together with additional rounds of RNA seqFISH+ and sequential immunofluorescence imaging in mESCs, representing a high image alignment accuracy with a median error of 47.3 nm in 3D across imaging rounds. Shaded areas represent the interquartile range. n = 1,332-2,402 matched fiducial markers in each hybridization round from three fluorescent channels. On average, 54.4 ± 18.2% (median ± s.d.) of fiducial markers detected in the reference image were used for the alignment, which corresponds to 0.66 ± 0.20 DNA spots per cell. **b**, The distribution of fiducial marker localization error in xy or z or xyz directions from 96 rounds of two-layer DNA seqFISH+ imaging in mESCs. The localization errors were calculated as translations of matched fiducial marker spots at given dimensions. **c**, Quantification of the median reference fiducial marker localization between pairs of fluorescence channels after the chromatic shift correction in 3D, suggesting a minimum chromatic effects across fluorescent channels after the correction. n = 2,301 matched spots in mESCs. **d**, Total on- and off-target barcodes detected per cell by two-layer DNA seqFISH+ in NMuMG cells (left). Average counts per each on- and off-target barcode per cell by two-layer DNA seqFISH+ in NMuMG cells (right). n = 100,049 and 31,171 for on- and off-target barcodes, respectively. **e**, The distribution of mean on-target barcode counts per cell grouped by chromosome identities by two-layer DNA seqFISH+ in mESCs (top) and NMuMG cells (bottom). The differences in detection efficiency between autosomal and X chromosomal loci reflect the identities of male mESCs and female NMuMG cells. We note chromosome 19 in NMuMG cells could be trisomy because of the 43.0% greater average barcode counts per cell than those in the other chromosomes. n = 100,049 loci in total. **f**, Heatmap comparing average spatial distances of pairs of loci between two-layer DNA seqFISH+ (upper right) and previous DNA seqFISH+ (Takei, Yun, et al. 2021) (lower left) at the DNA seqFISH+ 1-Mb resolution loci in Chr4 in mESCs. n = 149 loci. **g**, Pearson correlation of mean spatial distances of pairs of intra-chromosomal loci between two-layer DNA seqFISH+ and previous DNA seqFISH+ (Takei, Yun, et al. 2021) at the DNA seqFISH+ 1-Mb resolution 25-kb loci across the genome in mESCs. **h**, Pearson correlation of mean spatial distances of pairs of intra-chromosomal loci between two biological replicates of two-layer DNA seqFISH+ using the same loci in **g** in

mESCs. *n* = 159,397 pairs that were detected in both measurements in **g**, **h**. **i**, Representative visualization of eigenvectors between Hi-C (Bonev et al. 2017; Pascual-Reguant et al. 2018) (top) and two-layer DNA seqFISH+ (bottom) for mESCs and NMuMG cells, confirming the highly concordant compartment organization between the measurements. **j**, Spearman correlation of eigenvectors between Hi-C and DNA seqFISH+ across the mouse genome for mESCs (top) and NMuMG cells (bottom). *n* = 23,547, 23,582 loci that were commonly profiled between the measurements for mESCs and NMuMG cells. **k**, Comparison of changes of eigenvectors between mESCs and NMuMG cells computed by DNA seqFISH+ (x-axis) and Hi-C(Bonev et al. 2017; Pascual-Reguant et al. 2018) datasets (y-axis), representing concordant biological changes between the measurements. The A/B compartment scores from the principal eigenvector for both DNA seqFISH+ and Hi-C datasets were percentile normalized for comparison. The boxed regions with dashed lines represent genomic loci with more than 20% changes between cell lines with indicated numbers of loci. *r* represents Pearson correlation coefficient. 100 kb binning was used in **i-k**. **l**, Heatmap comparing median spatial distances of pairs of loci between two-layer DNA seqFISH+ (upper right) and previous DNA seqFISH+ (Takei, Yun, et al. 2021) (lower left) at the previously profiled 25-kb resolution loci in Chr3 and Chr14 in mESCs. *n* = 60 loci for each chromosome. Fluorescent channels used for two-layer DNA seqFISH+ loci are shown (top). We note that visible block patterns between loci in different fluorescent channels are likely due to the under-detection of DNA loci near the boundary of chromosome blocks introduced by the two-layer barcode scheme (see Supplementary Notes for details). **m**, Pearson correlation of median spatial distances of pairs of intra-chromosomal loci between two-layer DNA seqFISH+ and previous DNA seqFISH+ (Takei, Yun, et al. 2021) at previously profiled 25-kb resolution loci at the selected regions in Chr1-19, X in mESCs. **n**, Pearson correlation of median spatial distances of pairs of intra-chromosomal loci between two biological replicates of two-layer DNA seqFISH+ using the same 25-kb loci in **l** in mESCs. *n* = 35,400 pairs that were commonly profiled between the measurements in **m**, **n**. **o**, Representative Hi-C (Bonev et al. 2017; Pascual-Reguant et al. 2018) map (top) and median spatial distance map between pairs of intra-chromosomal loci by the two-layer DNA seqFISH+ in mESCs (bottom). The left and bottom lines with cyan and red colors mark

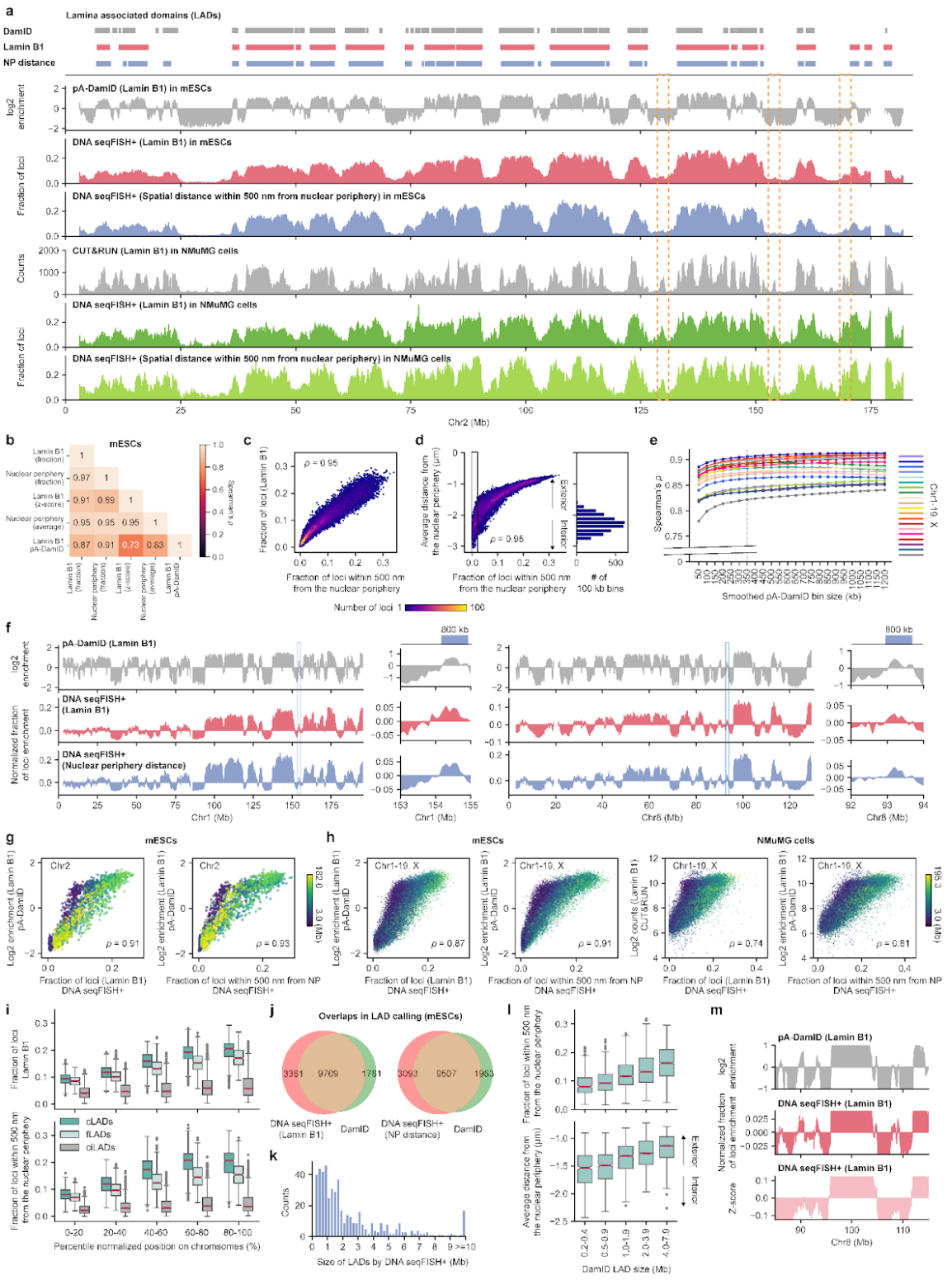
chromosomal block boundaries used for the two-layer DNA seqFISH+ barcoding. **p**, Average detection efficiency of 25-kb genomic loci with locus identities from 1-60 within 1.5-Mb blocks (n = 1,678) used in the two-layer DNA seqFISH+ barcoding. **q**, Illustration showing the decrease of fluorescence intensity of a chromosomal block paint. DNA spots located at the periphery of the paint block are preferentially dropped during the two-layer DNA seqFISH+ decoding. **r**, Illustration of the under-detection simulation on 60 genomic loci. The 60 continuous 25-kb genomic loci are splitted into the first and second 30-loci blocks, and identified genomic loci furthest from block centroid are dropped in simulated under-detection. **s**, Total counts of detected 25-kb genomic loci with identities from region ID 1-60 across 20 chromosomal regions before and after the under-detection in DNA seqFISH+ data (Takei, Yun, et al. 2021). The cyan and red bars mark simulated block boundaries. **t**, Representative median spatial distance maps between pairs of intra-chromosomal loci before and after the under-detection in DNA seqFISH+ data (Takei, Yun, et al. 2021). The cyan and red bars above the distance matrices mark simulated block boundaries. See Supplementary Notes for additional visualization and detailed explanation. n = 1,076 cells from two biological replicates of mESCs, n = 384 cells from one biological replicate of NMuMG cells in this study, and n = 446 cells from two biological replicates of mESCs by previous DNA seqFISH+ (Takei, Yun, et al. 2021) in **b**-**t**.



Extended Data Fig. 4| Validation of imaging-based chromatin profiling in cell culture experiments.

a, b, Representative images by sequential immunofluorescence (*n* = 65 markers) in **a** and DNA FISH of repetitive elements (*n* = 6 markers) in **b** in mESCs. **c**, Pearson correlation of imaging-based chromatin profiling for each marker (*n* = 69 markers in **a**, **b**, except E-Cadherin and HA-tag that do not stain nucleus) across the genome between two biological replicates in mESCs (left), showing H3K9me2 as an outlier, and corresponding density plot for individual loci (*n* = 100,049 loci) for H3K9ac as an example (right). **d**, Pearson correlation of imaging-based chromatin profiling for each marker (*n* = 13 markers) between two-layer DNA seqFISH+ and previous DNA seqFISH+ (Takei, Yun, et al. 2021) at the DNA seqFISH+ 1-Mb resolution 25-kb loci across the genome in mESCs (left), corresponding density plot for individual loci (*n* = 2,460 loci) for H3K9ac (middle), and corresponding chromatin profiles for H3K9ac as an example (right). The exact genomic loci targeted by 1-Mb resolution DNA seqFISH+ are shown by red dots (top right). **e**, Heat map showing the enrichment of subnuclear markers in pericentromeric heterochromatin (H3K9me3, MajSat), the nucleolus (Fibrillarin, ITS1_RNA), and nuclear speckles (SF3A66, Malat1_RNA) in mESCs computed by the imaging-based chromatin profiling. The strong nucleolar enrichments were observed in chromosomes 12, 16, 18, and 19, consistent with the previous report in mESCs derived from the 129/Ola mouse strain (Takei, Yun, et al. 2021). **f**, Quantification of an overlap of top 5% genomic loci associated with each marker between pairs of markers in mESCs. We note that genomic loci associated with nuclear speckles or nucleolus markers show high overlap scores (87.4%, 83.0% respectively) before and after DNA seqFISH+ procedure as represented by Malat1_RNA (imaged before DNA seqFISH+) and SF3A66 (imaged after DNA seqFISH+) as well as ITS1_RNA (imaged before DNA seqFISH+) and Fibrillarin (imaged after DNA seqFISH+) comparison, indicating a high degree of preservation of subnuclear structures in our protocol. **g**, Representative raw images of pericentromeric heterochromatin markers (H3K9me3, MajSat) and their colocalization patterns in mESCs. **h**, Spearman correlation of imaging-based chromatin profiling and sequencing-based chromatin profiling (pA-DamID for Lamin B1(van Schaik et al. 2022), CUT&Tag for other markers (Yang et al. 2022)) across the genome in mESCs. We note that imaging-based chromatin profiling measures spatial proximity between DNA loci and subnuclear structures, while sequencing-based profiling captures molecular interactions

between genomic DNA and antibodies of interest. **i**, Pearson correlation of imaging-based chromatin profiling and published RD-SPRITE datasets (Quinodoz et al. 2021) across the genome in mESCs, representing a similar enrichment around the subnuclear structures. RD-SPRITE characterized the frequency of interactions between each RNA marker and genomic regions. **j**, The enrichment of MajSat and H3K9me3 in mESCs along the chromosomal coordinates. Each chromosome was percentile normalized along the x-axis, similarly to the previous analysis (Quinodoz et al. 2021). While H3K9me3 by CUT&RUN did not show enrichment based on the chromosomal coordinates, MajSat by RD-SPIRTE as well as MajSat and H3K9me3 by DNA seqFISH+ showed overall enrichment based on the chromosomal coordinates. This is because the pericentromeric MajSat DNA repeats are located at the beginning of each mouse chromosome and RD-SPRITE and DNA seqFISH+ can capture distance-based decay. The H3K9me3 chromatin profiles by DNA seqFISH+ capture similar features due to the pericentromeric enrichment of H3K9me3, which spatially overlaps with MajSat as shown in **a**, **b**, **g**. On the other hand, CUT&RUN, which can capture local molecular interactions between H3K9me3 and DNA loci, does not capture the H3K9me3 enrichment in pericentromeric heterochromatin as such enrichment is outside the interaction range of the method. These fundamental differences of the methods could lead to the lack of strong correlation of H3K9me3 by CUT&RUN and DNA seqFISH+ in **h**. For the comparison with orthogonal sequencing-based datasets, imaging-based chromatin profiles with the fraction of loci calculation were used in **h**-**j**. **k**, Comparison of changes of chromatin profiles between mESCs and NMuMG cells computed by DNA seqFISH+ (x-axis) and sequencing-based datasets (CUT&Tag or CUT&RUN generated in this study or published pA-DamID (van Schaik et al. 2022)) (y-axis), representing a degree of similarity of biological changes between the measurements. The boxed regions with dashed lines represent genomic loci with more than 20% changes between cell lines with indicated numbers of loci. *r* represents Pearson correlation coefficient. 100 kb binning was used in **e**, **f**, **h**-**k**. n = 1,076 cells from two biological replicates of mESCs and n = 384 cells from one biological replicate of NMuMG cells.



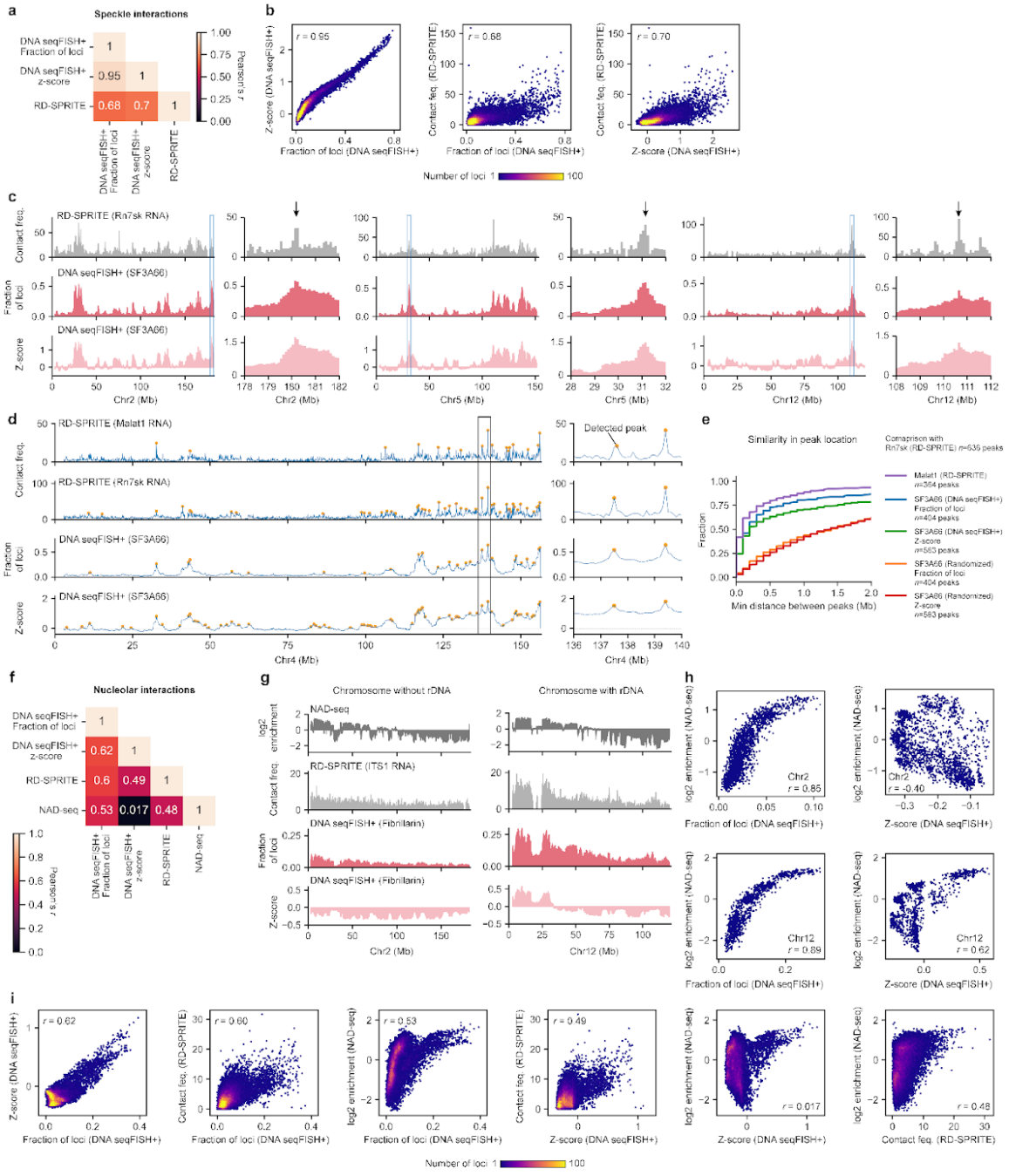
Extended Data Fig. 5| Additional validation of imaging-based chromatin profiling at

the nuclear lamina in cell culture experiments.

a, Representative chromatin profiles for Lamin B1 and at the nuclear periphery along with lamina-associated domains (LADs). The LADs in mESCs were obtained from published DamID datasets(Peric-Hupkes et al. 2010) or DNA seqFISH+ using Lamin B1 immunofluorescence or physical distance from nuclear periphery (see Methods). The Lamin B1 profiles in mESCs were obtained from the published pA-DamID datasets (van Schaik et al. 2022) while those in NMuMG cells were generated by CUT&RUN (Skene & Henikoff 2017; Skene et al. 2018) in this study by using optimized protocol for nuclear lamina (Ahanger et al. 2021) in differentiated cells. Representative cell-line specific peaks are highlighted by orange dashed boxes. **b**, Spearman correlation matrix of DNA seqFISH+ (either fraction of loci or z-score calculations) and pA-DamID (van Schaik et al. 2022) chromatin profiles for nuclear lamina associations in mESCs, using 100 kb bins across the genome. **c**, Genome-wide comparison of fraction of loci within 500 nm from the nuclear periphery or with Lamin B1 association with 100 kb binning by DNA seqFISH+. Lamin B1 enriched DNA loci are physically closer to the nuclear periphery as expected. Spearman correlation coefficient of 0.95. **d**, Genome-wide comparison of fraction of loci within 500 nm from the nuclear periphery or average spatial distance from the nuclear periphery with 100 kb binning by DNA seqFISH+ (left). Spearman correlation coefficient of 0.95. The distribution of the average spatial distance from the nuclear periphery (right) for a subset of DNA loci with less than 2% of each locus within 500 nm of the nuclear periphery (n = 3,485 loci), heighted as a black box (left). The average spatial distance from the nuclear periphery can resolve the radial organization of DNA loci in the nucleus for the loci with low interaction with the nuclear periphery. **e**, The Spearman correlation of the fraction of loci within 500 nm from the nuclear periphery in mESCs with smoothed pA-DamID data (van Schaik et al. 2022) (window size 2-48) was calculated for each chromosome, together with the change point in the correlation curve. Black dashed line represents the average change point of the curve from all chromosomes. 25 kb bins were used. **f**, Representative visualization of Lamin B1 chromatin profiles by pA-DamID(van Schaik et al. 2022) (top) and two-layer DNA seqFISH+ (middle), and fraction of loci within 500 nm from the nuclear periphery (bottom). The baselines of DNA seqFISH+ profiles were normalized by the expected fraction of LADs

by pA-DamID (log2 enrichment > 0) in each chromosome in mESCs for visual clarity. Genomic regions in the cyan boxes (left panels) are further shown with LAD called from the nuclear periphery distance data by DNA seqFISH+ (right panels). **g**, The scatter plots represent the Spearman correlation between DNA seqFISH+ and pA-DamID (van Schaik et al. 2022) datasets in chromosome 2 displayed in **a**. **h**, The scatter plots represent the Spearman correlation between DNA seqFISH+ and pA-DamID (van Schaik et al. 2022) or CUT&RUN datasets across the genome. Some of the outlier loci that were outside the display range in y-axis are not displayed for visual clarity. The genomic loci were colored by chromosomal coordinates, representing enrichment differences between methods based on the chromosomal coordinates in **g**, **h**. **i**, Comparison of Lamin B1 enrichment among loci with different categories as constitutive lamina- associated domains (cLADs), facultative LADs (fLADs), and constitutive inter-LADs (ciLADs) (Peric-Hupkes et al. 2010; Meuleman et al. 2013) Similarly to **g**, **h**, we found Lamin B1 enrichment differences between DNA seqFISH+ and DamID based on the chromosomal coordinates. We also confirmed the higher Lamin B1 enrichments at LADs (cLADs and fLADs) than those at ciLADs in each group. **j**, Genome-wide comparison of LADs calling in mESCs between published DamID (Peric-Hupkes et al. 2010) and DNA seqFISH+. LADs by DNA seqFISH+ were orthogonally computed by using the fraction of loci data of either Lamin B1 (left) or physical distance from the nuclear periphery (right), both of which yielded a similar degree of overlap with LADs identified by DamID. **k**, The size distribution of LADs called from the nuclear periphery distance data by DNA seqFISH+. **l**, The fraction of loci within 500 nm from the nuclear periphery (top) and average distance from the nuclear periphery (bottom) across different LAD sizes captured by DamID (Peric-Hupkes et al. 2010) in mESCs. Smaller LADs are less likely to be near the nuclear lamina and are more often positioned toward the nuclear interior. In box plots, the center lines for the median, boxes for the interquartile range, whiskers for values within 1.5 times the interquartile range, and points for outliers in **i** and **l**. **m**, Additional example of Lamin B1 profiles by pA-DamID (van Schaik et al. 2022) or by two-layer DNA seqFISH+ (middle, fraction of loci; bottom, z-score). The fraction of loci processing of DNA seqFISH+ data (middle) showed improved similarity with pA-DamID profiles compared to the z-score processing (bottom). The entire chromosome 8 is shown in

f. The display ranges in y-axis are optimized for the smaller LADs for visual clarity. We note that median Lamin B1 z-score profiles or average spatial distance profiles from the nuclear periphery capture information even when the DNA loci are far from the nuclear periphery. In contrast, the fraction of loci data as well as DamID and pA-DamID assays capture information only when the loci are in contact with or in close proximity to the nuclear periphery or Lamin B1. These distinct types of information in the chromatin profiles can lead to apparent differences in the chromatin profiles, especially at the small-scale LADs and iLADs. 100 kb binning was used in **a**-**d**, **f**-**m**. n = 1,076 cells from two biological replicates of mESCs and n = 384 cells from one biological replicate of NMuMG cells.

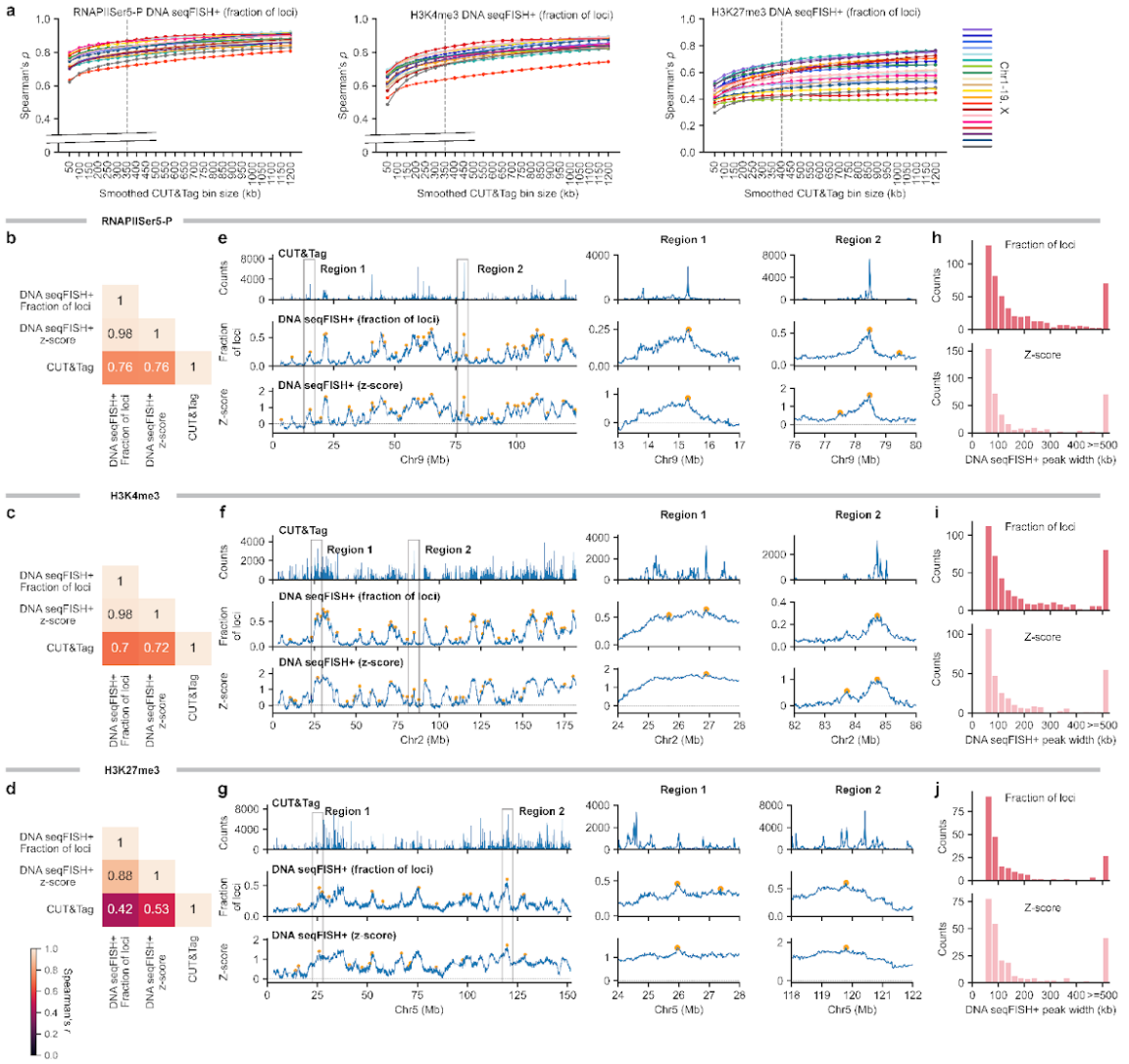


Extended Data Fig. 6| Additional validation of imaging-based chromatin profiling at nuclear speckles and the nucleolus in cell culture experiments.

a, Pearson correlation matrix of SF3A66 DNA seqFISH+ (either fraction of loci or z-score calculations) and Rn7sk RD-SPRITE (Quinodoz et al. 2021)chromatin profiles for nuclear speckle associations in mESCs, using 100 kb bins across the genome. **b**, Pearson correlation

plots of pairs of speckle association profiles in **a**. **c**, Representative chromatin profiles for nuclear speckles by Rn7sk RD-SPRITE (Quinodoz et al. 2021) (top) and SF3A66 DNA seqFISH+ (fraction of loci, middle; z-score, bottom) in mESCs. Genomic regions within the cyan boxes (left panels) are further zoomed-in, with the same 100-kb bin peaks detected by both methods highlighted by black arrows (right panels). **d**, Additional chromatin profiles for nuclear speckles by RD-SPRITE (Quinodoz et al. 2021) (either Malat1 or Rn7sk RNA) and SF3A66 DNA seqFISH+ (either fraction of loci or z-score) in mESCs, overlaid with detected confident peaks (orange dots). The genomic region within the black box (left) is further zoomed-in, with the same 100-kb bin peaks detected by both methods highlighted by black arrows (right). The same genomic region is shown in Fig. 1l. **e**, The comparison of genomic distance from each detected peak in one profile to the closest peak in the RD-SPRITE Rn7sk profile to evaluate the similarity of peak location between the imaging-based and sequencing-based profiles. The peak similarity within RD-SPRITE data was calculated using Rn7sk and Malat1 profiles as a positive control. The randomized peak locations in both fraction or median imaging-based chromatin profiles were used as negative controls. **f**, Pearson correlation matrix of Fibrillarin DNA seqFISH+ (either fraction of loci or z-score calculations) with ITS1 RD-SPRITE (Quinodoz et al. 2021) and NAD-seq (Bizhanova et al. 2020) chromatin profiles for nucleolar associations in mESCs, using 100 kb bins across the genome. **g**, Representative chromatin profiles for the nucleolus by NAD-seq (Bizhanova et al. 2020) and ITS1 RD-SPRITE (Quinodoz et al. 2021) and Fibrillarin DNA seqFISH+ (fraction of loci or z-score) in chromosomes without or with rDNA in mESCs. **h**, Pearson correlation plots of nucleolar association profiles by DNA seqFISH+ (fraction of loci, left panels; z-score, right panels) and NAD-seq(Bizhanova et al. 2020) for chromosomes in **g**. We found that the basal values varied between the fraction of loci and z-score DNA seqFISH+ profiling as represented by the elevated differences in correlation with NAD-seq in non-rDNA containing chromosomes (e.g. chromosome 2). This may reflect the fact that fraction of loci calculation more effectively captures the associations with high Fibrillarin-stained regions (nucleolus), while the median z-score can include the local enrichment of the staining, such as open chromatin region staining, at the background level. Nonetheless, both DNA seqFISH+ profiles (fraction of loci and z-score) showed consistency in top enriched

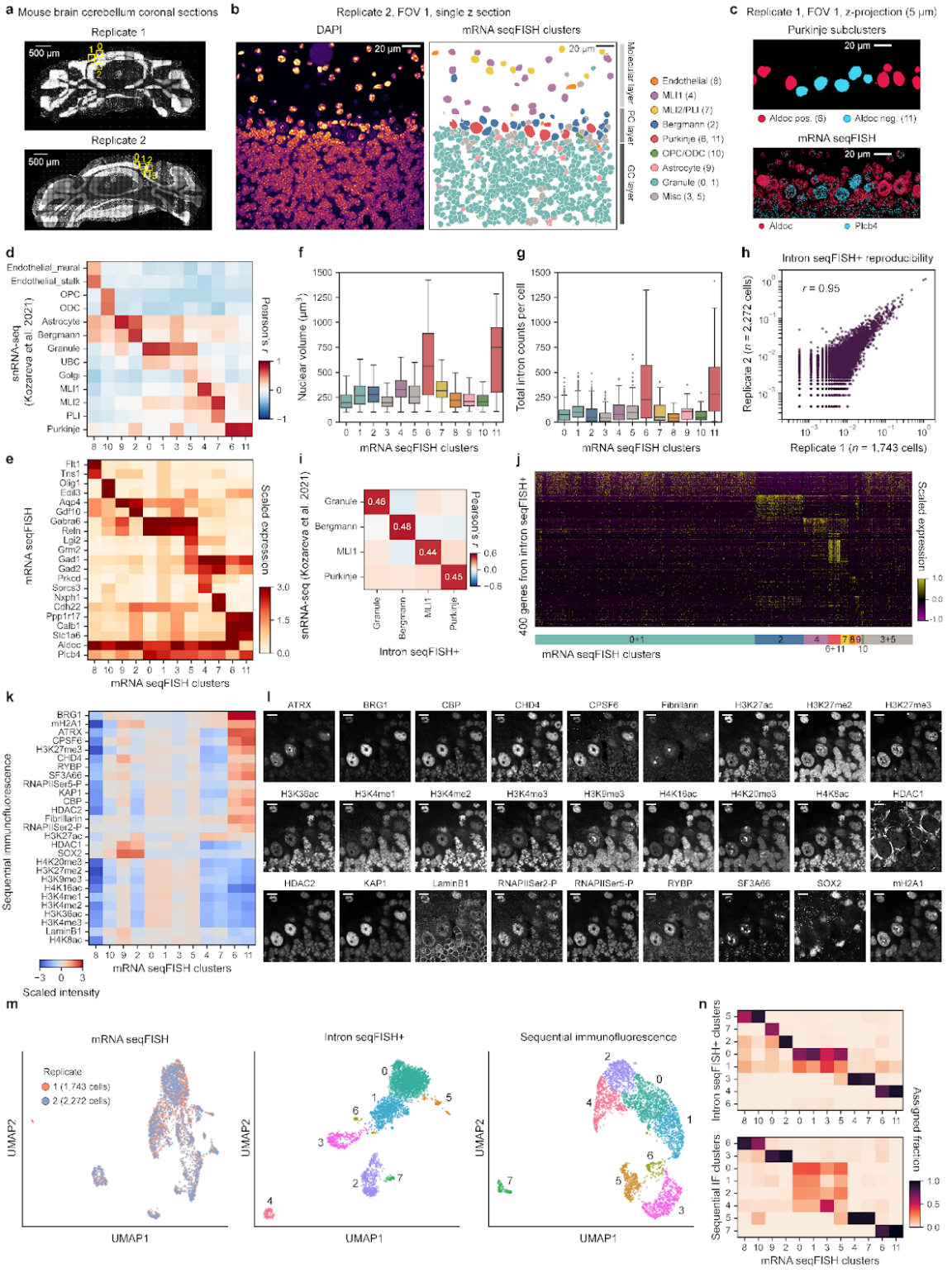
loci (Jaccard score of 0.66 for the top 2.5% of enriched loci) and Pearson correlation coefficient of 0.62 across the genome in **f**, **i**. **i**, Pearson correlation plots of pairs of nucleolar association profiles by DNA seqFISH+ (either fraction of loci or z-score), ITS1 RD-SPRITE (Quinodoz et al. 2021)NAD-seq(Bizhanova et al. 2020) in **f** (Bizhanova et al. 2020) 100 kb binning was used in **a**-**i**. *n* = 1,076 cells for two-layer DNA seqFISH+ from two biological replicates of mESCs.



Extended Data Fig. 7| Additional validation of imaging-based chromatin profiling with additional markers in cell culture experiments.

a, The Spearman correlation between the fraction of loci for RNAPIISer5-P, H3K4me3, and H3K27me3 in mESCs and their corresponding smoothed CUT&Tag enrichment scores (window size 2-48) for each chromosome. The average change points from all chromosomes in the correlation curves were also identified and visualized as black dashed lines. **b-d**, Spearman correlation matrix of DNA seqFISH+ (either fraction of loci or z-score calculations) and CUT&Tag chromatin profiles for RNAPIISer5-P in **b**, H3K4me3 in **c**, and H3K27me3 in **d** in mESCs, using 100 kb bins across the genome. **e-f**, Representative

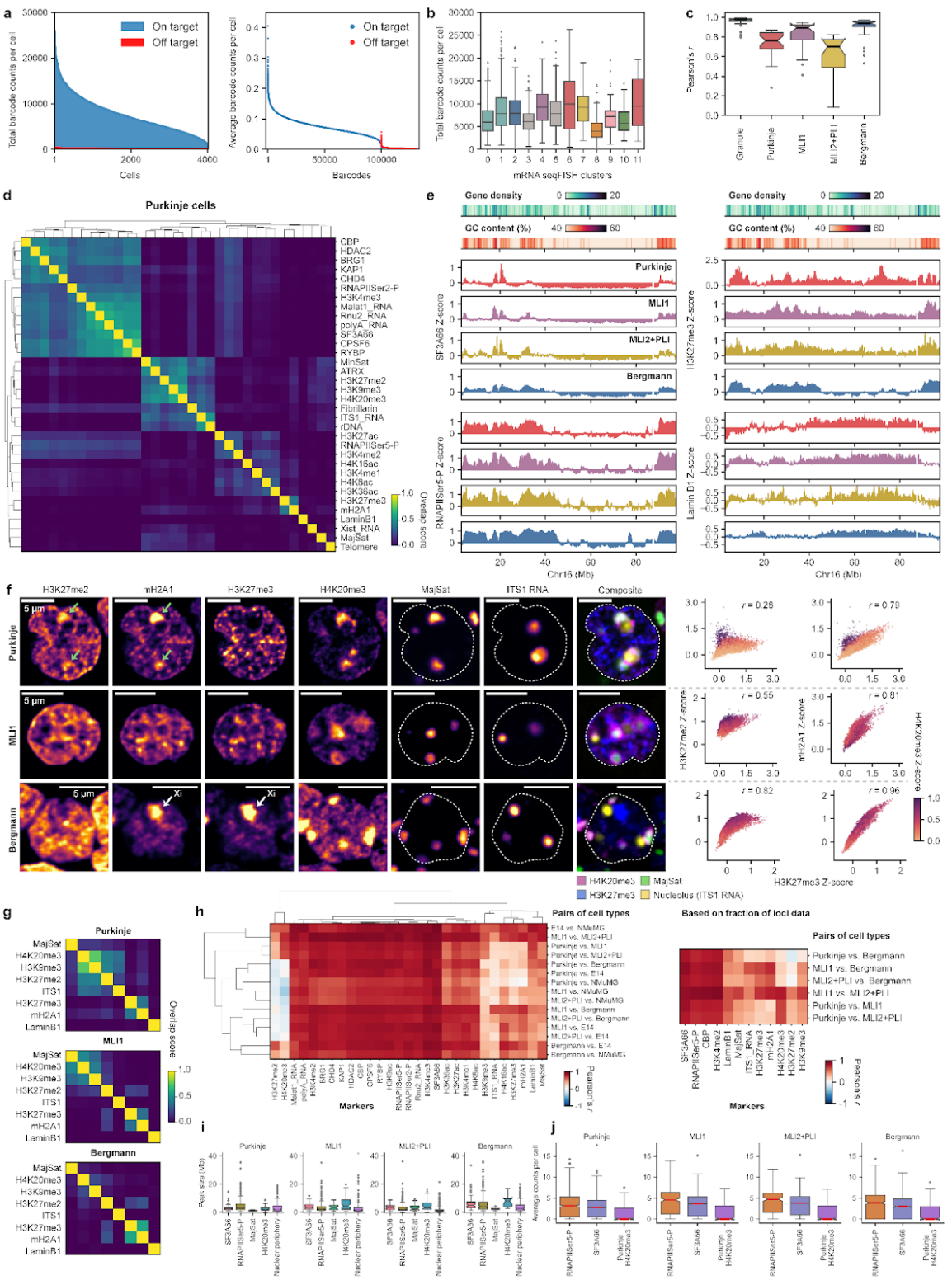
chromatin profiles for RNAPIISer5-P in **e**, H3K4me3 in **f**, and H3K27me3 in **g** by CUT&Tag and DNA seqFISH+ (either fraction of loci or z-score) in mESCs, overlaid with detected confident peaks (orange dots). Genomic regions within the black box (left) are further zoomed-in (right). **h-j**, The distribution of peak width detected by DNA seqFISH+ (fraction of loci, top; z-score, bottom) for RNAPIISer5-P in **h**, H3K4me3 in **i**, and H3K27me3 in **j** in mESCs. We observed that these peak widths typically range from 50-kb to 150-kb sizes across markers. The original 25 kb bins were used in **a, e-j**. *n* = 1,076 cells for two-layer DNA seqFISH+ from two biological replicates of mESCs.



Extended Data Fig. 8| Validation and characterization of spatial multi-omics measurements in the adult mouse brain cerebellum.

a, Large-field DAPI images of the adult mouse brain cerebellum coronal sections from two biological replicates. Yellow boxes represent unique fields of view (FOVs) imaged in each biological replicate. **b**, Representative images from a single z-section in one field of view for DAPI staining (left) and spatial distribution of cell clusters computed from mRNA seqFISH profiles (right). The cell type annotation for each cell cluster (0-11) was determined by the comparison with the single-nucleus RNA sequencing dataset (Kozareva et al. 2021) shown in **d**. **c**, Representative images of spatial distribution of cell clusters for Purkinje cell subtypes (top) and corresponding marker gene expression in the nucleus (bottom), reflecting the known patterns of parasagittal stripes between Aldoc positive and negative Purkinje cells in the mouse cerebellum (De Zeeuw 2021; Kozareva et al. 2021; Chen et al. 2022). **d**, Comparison of cell clusters (0-11) defined by mRNA seqFISH and cell types identified by the single-nucleus RNA sequencing (Kozareva et al. 2021). Based on the degree of Pearson correlation, we annotated our mRNA seqFISH clusters as shown in **b**. **e**, Marker gene expression profiles in each cell cluster. Those include previously characterized marker genes (Kozareva et al. 2021; Chen et al. 2022) such as Flt1 in Endothelial cells (cluster 8), Olig1 in oligodendrocyte precursor cells and oligodendrocytes (cluster 10), Aqp4 in Astrocytes (cluster 9), Gdf10 in Bergmann glia (cluster 2), Gabra6 in Granular cells (clusters 0 and 1), Sorcs3 in MLI1 (cluster 4), Nxph1 in MLI2/PLI (cluster 7), and Ppp1r17 in Purkinje cells (clusters 6 and 11, which can be further divided by Aldoc and Plcb4 as shown in **c**). **f**, Distribution of nuclear volume for cells in each cell cluster. **g**, Distribution of total intron counts per cell in each cell cluster by intron seqFISH+. In box plots, the center lines for the median, boxes for the interquartile range, whiskers for values within 1.5 times the interquartile range, and points for outliers in **f**, **g**. **h**, Reproducibility of intron seqFISH+ from two biological replicates with the adult mouse brain cerebellum. n = 17,849 genes that were detected at least once in each replicate. **i**, Comparison of gene expression profiles by intron seqFISH+ and single-nucleus RNA sequencing (Kozareva et al. 2021) in four major cell types showed overall high consistency of cell-type specific gene expression programs. Genes with averaged intron counts of more than 0.01 per cell by intron seqFISH+ were used for comparison (n = 1,422, 1,115, 2,322, and 5,473 genes for Granule, Bergmann, MLI1, and Purkinje cells). **j**, Heatmap of nascent gene expression profiles of differentially expressed

genes by intron seqFISH+ across single cells grouped by cell clusters defined by mRNA seqFISH. **k**, Heatmap of sequential immunofluorescence intensity profiles across cell clusters defined by mRNA seqFISH. Note that HDAC1 and SOX2 were expressed in Bergmann glia (cluster 2) and astrocyte (cluster 9), consistent with previous studies (MacDonald & Roskams 2008; Cerrato et al. 2018). **l**, Representative sequential immunofluorescence images from a single z-section in the adult mouse brain cerebellum. **m**, UMAP-embedding of cells colored by two biological replicates (left) and each cell cluster defined by intron seqFISH+ profiles (middle) or sequential immunofluorescence intensity profiles (right). **n**, Comparison of cell clusters defined by mRNA seqFISH and intron seqFISH+ (top) or sequential immunofluorescence (bottom), representing robust identification of similar cell clusters regardless of the measurement modalities in the adult mouse brain cerebellum. n = 4,015 cells (n = 1,504, 832, 518, 357, 263, 164, 113, 88, 76, 56, 29, 15 cells from mRNA seqFISH cluster 0 to 11) from two biological replicates of the adult mouse brain cerebellum.

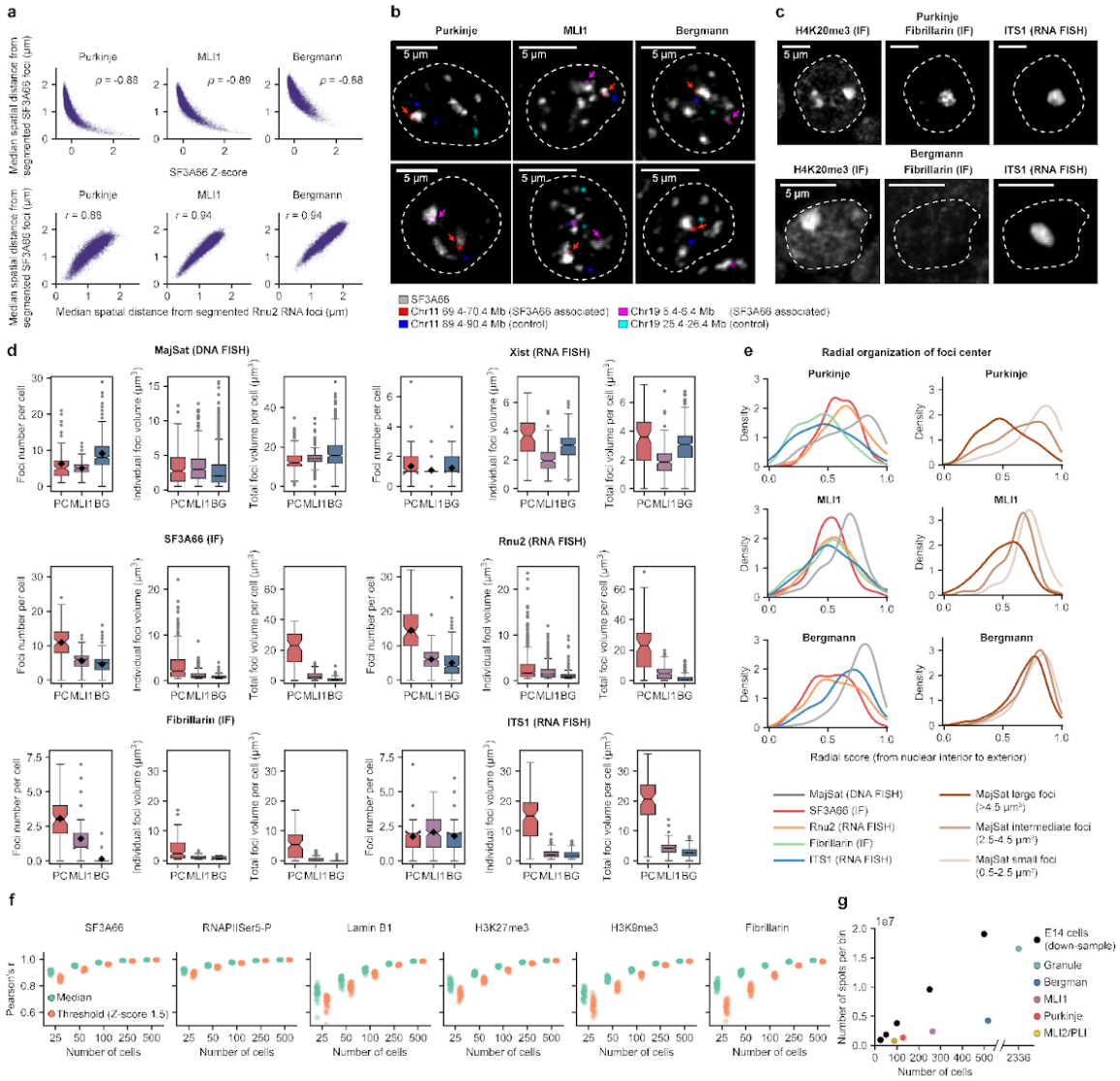


Extended Data Fig. 9| Additional validation and characterization of spatial multi-omics

measurements in the adult mouse brain cerebellum.

a, Total on- and off-target barcodes detected per cell by DNA seqFISH+ in the mouse cerebellum (left). Average counts per each on- and off-target barcode per cell by DNA seqFISH+ in the mouse cerebellum (right). n = 100,049 and 31,171 for on- and off-target barcodes, respectively. **b**, Total on-target barcodes detected per cell by DNA seqFISH+ in each transcriptionally-defined cell cluster in the mouse cerebellum. n = 1,504, 832, 518, 357, 263, 164, 113, 88, 76, 56, 29, 15 cells from mRNA seqFISH cluster 0 to 11. **c**, Pearson correlation of imaging-based chromatin profiling across the genome for each marker (n = 26 markers) in each cell type between two biological replicates of the adult mouse brain cerebellum. **d**, Quantification of an overlap of top 5% genomic loci associated with each marker between pairs of markers in Purkinje cells, largely separating active and repressive chromatin markers. **e**, Genomic features and representative imaging-based chromatin profiling with four markers in each cell type in chromosome 16. **f**, Representative raw immunofluorescence images for various repressive markers from a single z-section for each cell type (left). Purkinje-specific pericentromeric staining by H3K27me2 and mH2A1 is highlighted by green arrows. The intense mH2A1 and H3K27me3 clusters visible in Bergmann glia highlighted by white arrows represent the inactive X chromosome (Xi) territory (Linhoff et al. 2015) in the female mouse cerebellum section. Scatter plots for individual loci show the relationship of each repressive marker enrichment in each cell type (right). **g**, Degree of overlap of top 5% enriched loci between pairs of repressive markers in each cell type. **h**, Degree of similarity of chromatin profiles between pairs of cell types by Pearson correlation, including cell lines (mESCs and NMuMG cells) (left) or using fraction of loci data (right), representing the similar trends regardless of the processing method. **i**, Comparison of the distribution of detected peak sizes in each cell type for selected markers used for downstream analysis. **j**, Comparison of gene expression profiles in RNAPIISer5-P, SF3A66 enriched loci or H4K20me3-associated loci identified in Purkinje cells in each cell type, representing a silencing of gene expression at the majority of the H4K20me3-associated loci regardless of cell types. The gene expression profiles were calculated from the published snRNA-seq datasets in the adult mouse brain cerebellum (Kozareva et al. 2021). In box plots, the center lines for the median, boxes for the interquartile range, whiskers for values within

1.5 times the interquartile range, and points for outliers in **b**, **c**, **i**, **j**. 200 kb binning (n = 12,562 loci in total) was used for the analysis in **c**-**h**. n = 4,015 cells from two biological replicates of the adult mouse brain cerebellum.

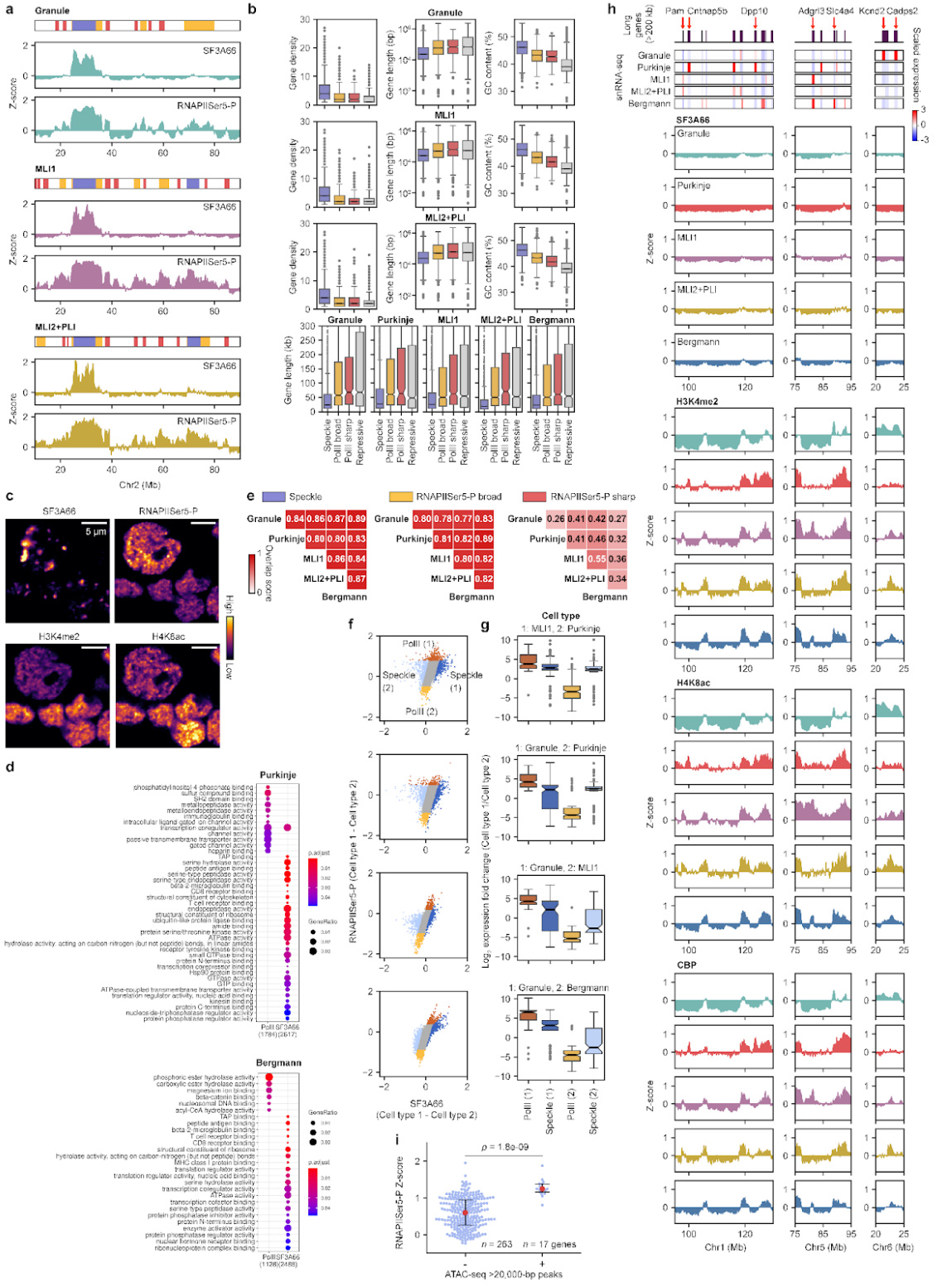


Extended Data Fig. 10| Additional validation of imaging-based chromatin profiling and characterization of subnuclear foci in the adult mouse brain cerebellum.

a, Comparison of z-score normalized SF3A66 chromatin profiles (x-axis, top) and spatial distances from segmented SF3A66 foci computed from the closest spatial distance between DNA loci and binary SF3A66 masks (y-axis, top), representing a high degree of similarity between z-score-based SF3A66 chromatin profiles and physical distances from SF3A66 foci across cell types. Comparison of spatial distances around nuclear speckles (Rnu2 RNA and SF3A66) from segmented Rnu2 RNA foci (x-axis, bottom) and SF3A66 foci (y-axis, bottom), representing a robustness of our approach in measuring genomic loci organization

around subnuclear structures across cell types. We also note that Rnu2 RNA and SF3A66 immunofluorescence imaging was performed before and after DNA seqFISH+ procedure, indicating that our protocol preserved the subnuclear structures in the tissue section experiments. 200 kb binning (n = 12,562 loci in total) was used for the analysis. **b**, Representative raw SF3A66 images overlaid with decoded DNA loci in SF3A66 associated domains or negative control regions with a maximum z-projection of four sections. The spatial overlaps between DNA loci and nuclear speckles marked by SF3A66 are highlighted by colored arrows. The SF3A66-associated loci shown here were identified within SF3A66 domains in all three cell types. **c**, Representative raw images of H4K20me3 and nucleolar markers (Fibrillarin, ITS1 RNA) with a maximum z-projection of four sections in Purkinje cells (top) and Bergmann glia cells (bottom). We did not detect Fibrillarin signals in Bergmann glia cells as quantified in **d**. We note that we performed sequential immunofluorescence preparation using tens of antibodies under one condition. Some antibodies (e.g., Fibrillarin) may require optimized conditions for each antibody in tissues(Eberhart et al. 2012). **d**, Comparison of subnuclear marker foci features for Purkinje cells (PC), MLI1, and Bergmann glia (BG) by quantifying foci number per cell, individual foci volume, and total foci volume per cell (from left to right panel for each marker) characterized cell-type specific foci organization. We note that by comparing foci features from two nuclear speckle markers, Rnu2 RNA (imaged before DNA seqFISH+) and SF3A66 antibody (imaged after DNA seqFISH+), we confirmed that nuclear speckle organization was consistent before and after DNA seqFISH+. In contrast, in the case of nucleoli, the staining of the nucleolar marker, Fibrillarin, was weaker or not detected (e.g., Bergmann glia in **c**). This can be seen by the decreased volumes of nucleolar Fibrillarin foci relative to those from nucleolar ITS1 RNA by RNA FISH. While Fibrillarin staining quality was poor in the tissue section experiments, the nucleolar subnuclear structures themselves were preserved as shown by the comparison between Fibrillarin and ITS1 in MLI1 (**e**) and Purkinje cells (**c**, **e**). We also note that downstream analysis regarding nucleolar association across cell types was performed with ITS1, instead of Fibrillarin. In box plots, the black diamonds for the mean, the center lines for the median, boxes for the interquartile range, whiskers for values within 1.5 times the interquartile range, and points for outliers. Some of the outlier foci (n = 5 in

total) that were outside the display range are not displayed for visual clarity. **e**, Kernel density estimation of the centroid of subnuclear foci from nuclear interior to exterior in each cell type. By comparing the different markers in each cell type (left), we characterized cell-type specific radial organization of subnuclear foci in the nucleus. We confirmed the trend of more nuclear interior organization of nucleoli relative to MajSat in Purkinje cells, consistent with literature(Solovei et al. 2004). In addition, we compared the radial organization of different sizes of MajSat foci in each cell type (right). We confirmed the foci size dependent radial organization in Purkinje cells (i.e., larger foci tend to be more nuclear interior), consistent with literature(Solovei et al. 2004). We further extended such observation in MLI1 and revealed the lack of MajSat foci size dependent radial organization in Bergmann glia. **f**, mESCs were down-sampled to 500, 250, 100, and 50 cells, with 30 replicates per group. The correlation between the downsampled fraction of loci and median z-scores with the respective ensemble scores calculated from all 1,076 mESCs was calculated and plotted for immunofluorescence markers SF3A66, RNAPIISer5-P, LaminB1, H3K27me3, H3K9me3, and Fibrillarin. The median Pearson correlation for the downsampled data showed that median z-score was more robust to down-sampling. **g**, The relationship between the number of cells and the number of spots per bin in mESCs (E14 cells) and mouse brain cerebellum experiments. 200 kb binning was used for the mouse brain cerebellum data. 100 kb binning was used for mESC data in **f**, **g**. mESCs were down-sampled from n = 1,076 cells from two biological replicates in **f**, **g**. n = 2,336, 128, 263, 88, and 518 cells for Granule, Purkinje, MLI1, MLI2+PLI, and Bergmann glia cells from two biological replicates of the adult mouse cerebellum in **a-e, g**.

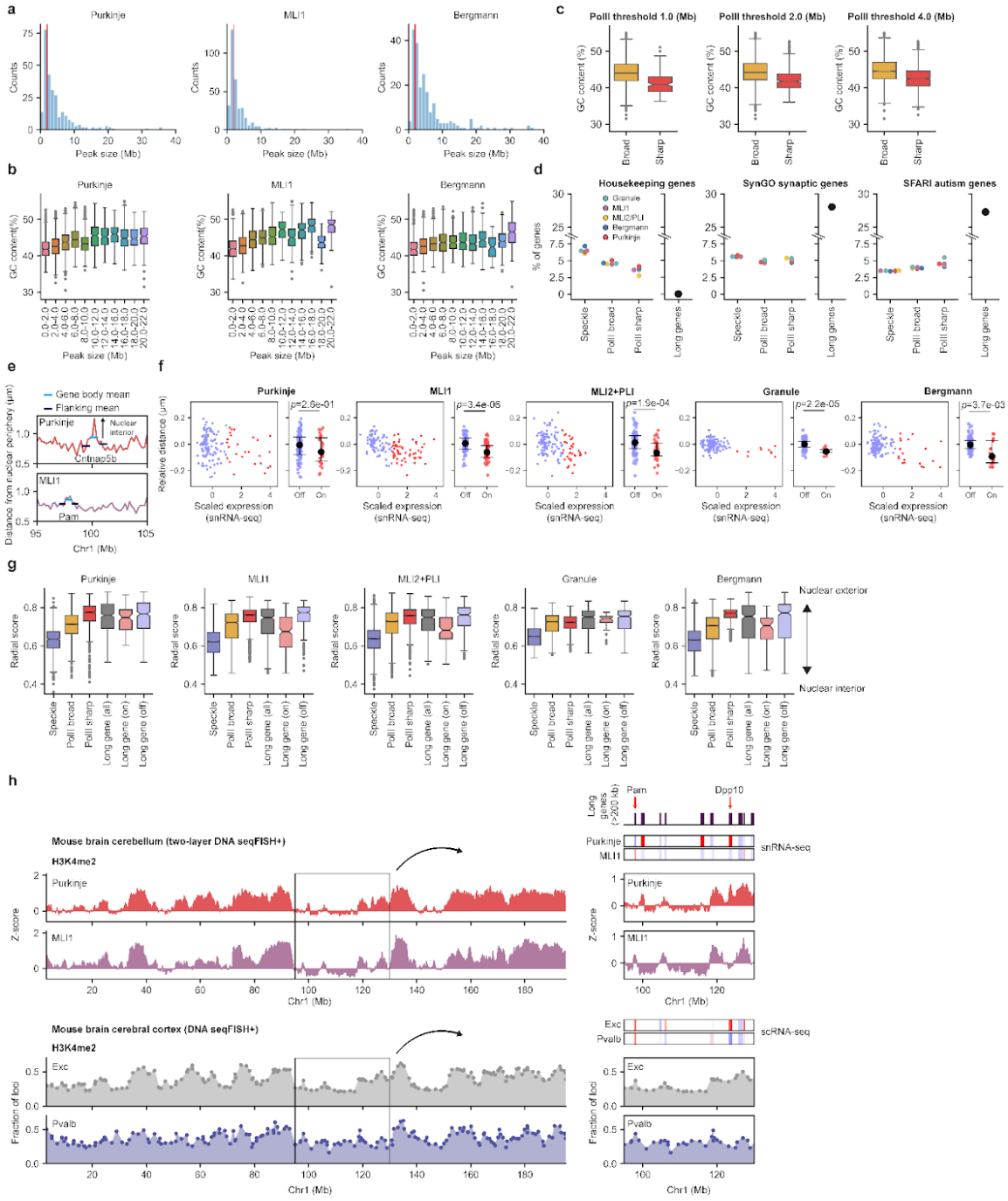


Extended Data Fig. 11| Additional visualization and characterization of distinct active

chromatin regions and long gene loci.

a, Representative classification of three types of active domains (nuclear speckle, RNAPIISer5-P broad and sharp) along with the chromatin profiling of SF3A66 (nuclear speckle marker) and RNAPIISer5-P across cell types. **b**, Comparison of genomic features across different active domains in each cell type. *n* = 2,531, 1,789, 458, 4,220 loci (Granule), *n* = 2,533, 2,060, 1,228, 4,654 loci (MLI1), and *n* = 2,376 2,015 1,367 4,210 loci (MLI2+PLI) from left to right category. The boxplots for median gene length per bin for each cell type are also displayed as a linear scale up to 300 kb (bottom panels) for visual clarity. **c**, Representative raw images of active markers from a single z-section of the adult mouse brain cerebellum, showing distinct organization between nuclear speckles (SF3A66) and other active markers (e.g. RNAPIISer5-P, H3K4me2, H4K8ac) regardless of the cell types. **d**, GO term comparison between SF3A66 and RNAPIISer5-P associated genomic loci in each cell type, representing largely distinct enrichment. Similar observation of distinct pathway enrichments between speckle-associating and speckle-non-associating p53 target genes was previously observed in human cell lines(Alexander et al. 2021). **e**, Overlap of each active domain across cell types, showing more cell-type specific organization of RNAPIISer5-P sharp domains. **f**, Comparison of differential association of genomic loci with SF3A66 and RNAPIISer5-P between pairs of cell types. **g**, Comparison of differential mRNA expression(Kozareva et al. 2021) between pairs of cell types at differentially associated loci with either PolII (RNAPIISer5-P) or Speckle (SF3A66) in each cell type. *n* = 150, 114, 198, 126 loci (MLI1 vs. Purkinje), *n* = 203, 136, 139, 146 loci (Granule vs. Purkinje), *n* = 164, 85, 143, 79 loci (Granule vs. MLI1), and *n* = 147, 96, 117, 87 loci (Granule vs. Bergmann) from left to right category. In box plots, the center lines for the median, boxes for the interquartile range, whiskers for values within 1.5 times the interquartile range, and points for outliers in **b**, **g**. **h**, Representative genomic regions with long genes (>200 kb) (top), corresponding mRNA expression(Kozareva et al. 2021) (middle), and chromatin profiles of different active markers (bottom) in each cell type. Nuclear speckles marked by SF3A66 were not enriched at the differentially expressed long genes highlighted with red arrows (top). **i**, Comparison of Purkinje RNAPIISer5-P enrichment between long genes with the absence or presence of highly open chromatin regions defined by ATAC-seq >20 kb peaks

in Purkinje cells(Kwak et al. 2021). The red dots for the median and error bars for the interquartile range. *p* values by two-sided Wilcoxon's signed rank-sum test. 200 kb binning (*n* = 12,562 loci in total) was used for the analysis in **a**, **b**, **e**-**h**. *n* = 2,336, 128, 263, 88, and 518 cells for Granule, Purkinje, MLI1, MLI2+PLI, and Bergmann glia cells from two biological replicates of the adult mouse cerebellum.

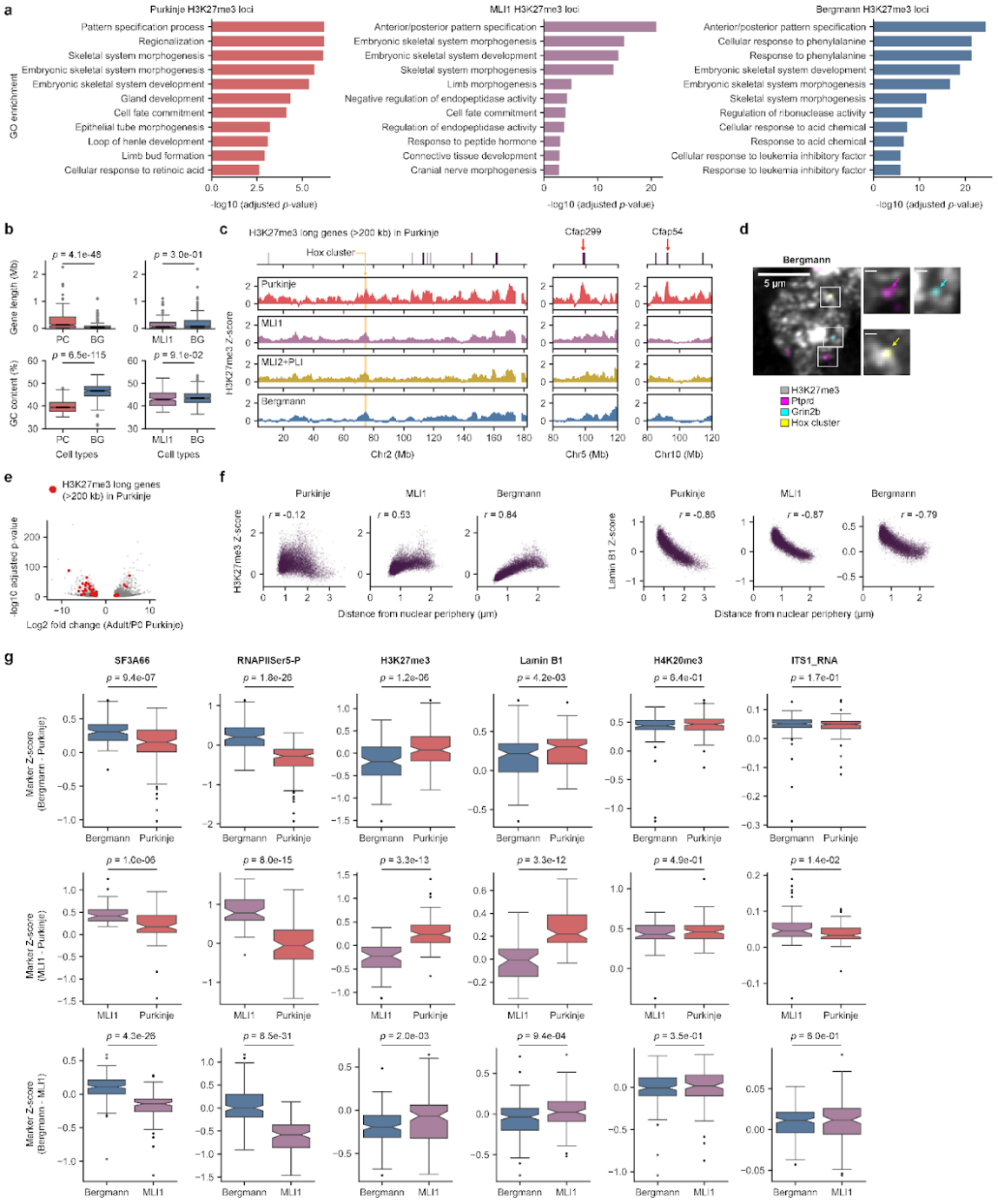


Extended Data Fig. 12| Additional validation and characterization of distinct active chromatin regions and long gene loci.

a, Histogram of RNAPIISer5-P peak size in each cell type. To separate RNAPIISer5-P broad and sharp domains, a threshold value of 2 Mb was used as a default. **b**, Comparison of GC

content across different RNAPIISer5-P domain sizes in each cell type. **c**, The comparison of GC content between RNAPIISer5-P broad and sharp domains by different threshold values in Purkinje cells, representing robust changes of GC content between two domains regardless of the thresholding values. **d**, Comparison of percentage of genes categorized as housekeeping genes (left), synaptic genes (middle), or autism genes (right) in active domains (n = 7,423-11,994, 3,677-9,049, 800-2,651 genes for nuclear speckles, RNAPIISer5-P broad and sharp, respectively) for each cell type as well as in cell-type specific long genes (n = 132 genes) identified in this study. To classify the genes, annotations from previous studies or databases were used (see Methods). **e**, Example visualization of changes of spatial distance from the nuclear periphery for long gene loci relative to their flanking loci. To compute the relative spatial distance change, a mean distance for 1 Mb of each of the flanking loci was subtracted by that for 1 Mb of each long gene locus for each cell type. **f**, Comparison of gene expression and spatial distance changes of long gene loci from the nuclear periphery (left panel for each cell type). The spatial distance changes were quantified as illustrated in **e**. The gene expression states were binarized (on or off) using a threshold value of 1 for the expression level measured by snRNA-seq (Kozareva et al. 2021) The relative distances were then compared between on and off states (right panel for each cell type). On average, the on-state loci positioned 53.6, 66.7, 77.2, 56.0, 89.1 nm more toward the nuclear interior than the off-state loci when spatial distances were computed relative to the flanking loci in Purkinje, MLI1, MLI2+PLI, Granule, and Bergmann glia. The black dots represent the median and whiskers represent the interquartile range. p values by two-sided Wilcoxon's signed rank-sum test. **g**, Comparison of the radial positioning of on- and off-state long gene loci in different cell types (with the same definition in **f**) as well as genomic loci located in different active regions (nuclear speckles, RNAPIISer5-P broad and sharp domains). The radial positioning of transcriptionally active on-state long gene loci appears to be more toward the nuclear interior compared to the off-state long gene loci in most of the cell types. We note that we observed transcriptionally active long gene loci tend to localize more toward the nuclear periphery compared to nuclear speckle-associated genomic loci, consistent with a recent finding (Zhao et al. 2023). **h**, Comparison of H3K4me2 chromatin profiles by two-layer DNA seqFISH+ in the adult mouse brain cerebellum in this study and 1-Mb resolution

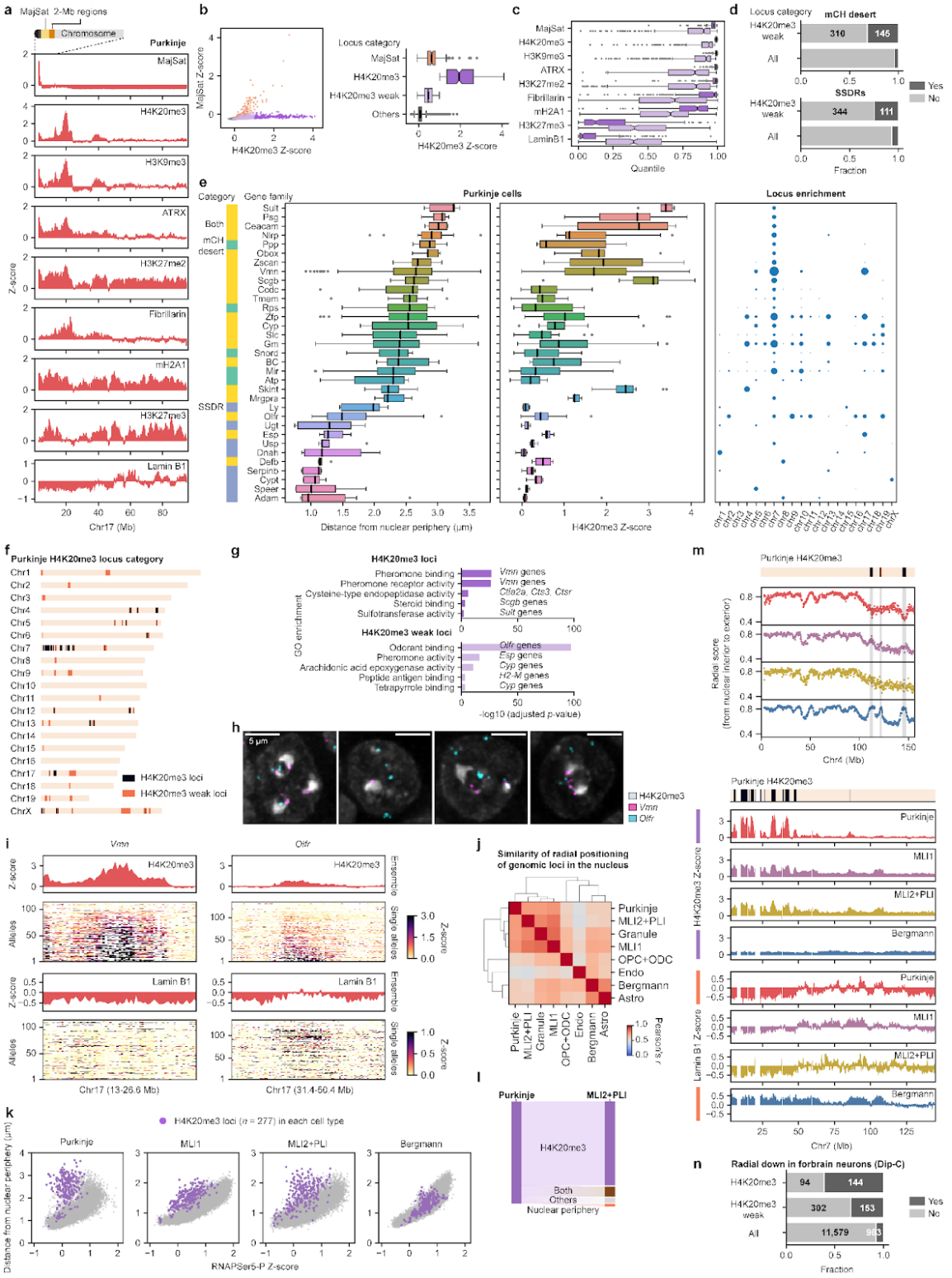
DNA seqFISH+ in the adult mouse brain cerebral cortex (Takei, Zheng, et al. 2021). Genomic regions in the black boxes (left) are further shown with long gene (>200 kb) annotations and scaled mRNA expression profiles from the corresponding brain regions (Tasic et al. 2016; Kozareva et al. 2021)(right). The exact genomic loci targeted by 1-Mb resolution DNA seqFISH+ are represented by colored dots. We note that the H3K4me2 enrichment at the Dpp10 locus was not captured in the excitatory neurons despite the gene being transcriptionally active possibly because the previous DNA seqFISH+ probes (Takei, Zheng, et al. 2021), targeting 25-kb loci with ~1 Mb gap in between on average, did not cover the genomic region of the Dpp10 locus. On the other hand, the two-layer DNA seqFISH+ probes cover all possible 25-kb loci across the genome except for highly repetitive genomic regions and therefore detect the H3K4me2 enrichment at the Dpp10 locus in Purkinje cells. n = 1,895 and 155 cells for excitatory neurons and Pvalb neurons from three biological replicates for the 1-Mb resolution DNA seqFISH+. 200 kb binning (n = 12,562 loci in total) was used for the analysis in **a**-**h**. n = 2,336, 128, 263, 88, and 518 cells for Granule, Purkinje, MLI1, MLI2+PLI, and Bergmann glia from two biological replicates of the adult mouse cerebellum.



Extended Data Fig. 13| Additional visualization and characterization of cell-type specific H3K27me3-associated regions.

a, The top eleven GO terms identified from H3K27me3 chromatin profiles in each cell type represent the enrichment of similar GO terms across cell types. **b**, Comparison of genomic

features between loci that were differentially associated with H3K27me3 between two cell types. *n* = 595, 359 loci for Purkinje cells (PC) vs. Bergmann glia (BG) and *n* = 169, 491 loci for MLI1 vs. BG. **c**, Additional examples of H3K27me3 profiles across cell types along with Hox cluster in Chr2 and Purkinje-specific H3K27me3-associated long gene loci (top). **d**, Representative single cell visualization of genomic loci relative to H3K27me3 in Bergmann glia. Scale bars, 500 nm. **e**, Violin plot of expression changes between P0 and adult Purkinje cells (Stoyanova et al. 2021), representing the down-regulaiton of 36 out of 46 differentially expressed genes categorized as Purkinje-specific H3K27me3-associated long genes from the P0 to adult. **f**, Comparison of spatial distance from nuclear periphery and H3K27me3 or Lamin B1 chromatin profiles in each cell type. **g**, Comparison of differential association with each subnuclear marker at differentially expressed loci defined by intron seqFISH+ between pairs of major cell types. The increased association with SF3A66 at transcriptionally up-regulated loci was similarly observed in the adult mouse cerebral cortex (Takei, Zheng, et al. 2021). *n* = 114, 129 loci for Bergmann glia vs. Purkinje cells, *n* = 48, 81 loci for MLI1 vs. Purkinje cells, and *n* = 151, 114 loci for Bergmann glia vs. MLI1. *p* values by two-sided Wilcoxon's signed rank-sum test in **b**, **g**. In box plots, the center lines for the median, boxes for the interquartile range, whiskers for values within 1.5 times the interquartile range, and points for outliers in **b**, **g**. 200 kb binning (*n* = 12,562 loci in total) was used for the analysis in **a**-**c**, **f**, **g**. *n* = 128, 263, 88, and 518 cells for Purkinje, MLI1, MLI2+PLI, and Bergmann glia cells from two biological replicates of the adult mouse cerebellum.

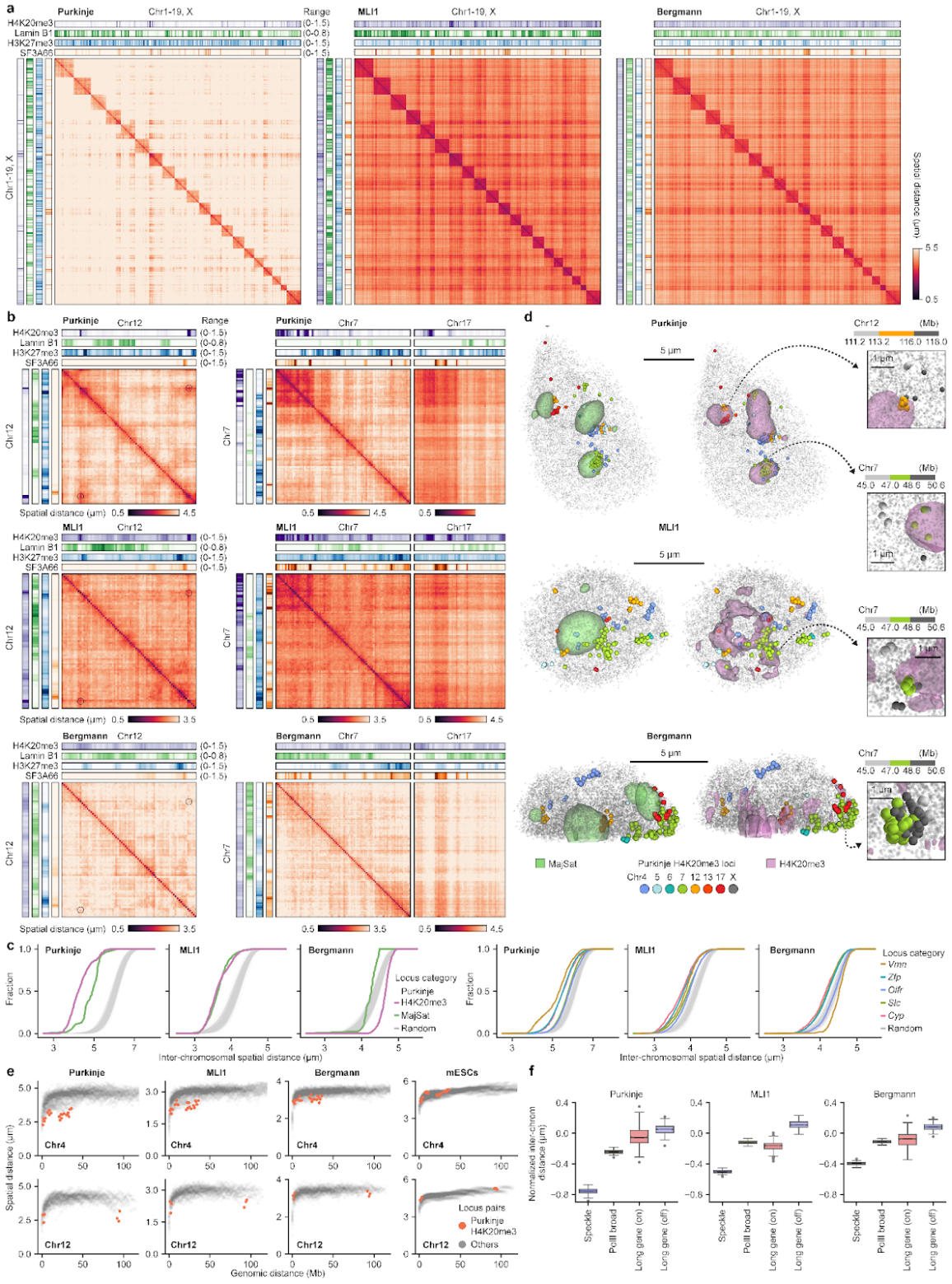


Extended Data Fig. 14| Additional visualization and characterization of H4K20me3-

associated regions.

a, Cartoon illustration of a mouse chromosome (top) and representative imaging-based chromatin profiling of repressive markers in Purkinje cells in chromosome 17 (bottom). **b**, Scatter plots of pairs of chromatin profiles (left), colored by the locus category of MajSat (orange), H4K20me3 (purple), H4K20me3 weak (light purple), or others (gray) and corresponding box plots showing H4K20me3 enrichment for loci in each category (right). **c**, Comparison of quantile of H4K20me3 and H4K20me3-weak category loci in **b** for each repressive marker, representing stronger enrichment of repressive markers around pericentromeric heterochromatin (MajSat, H4K20me3, H3K9me3, ATRX, H3K27me2, Fibrillarin, and mH2A1) at H4K20me3 loci (purple) over H4K20me3-weak loci (light purple). In contrast, two repressive markers (H3K27me3 and Lamin B1) are more depleted at H4K20me3 loci compared to H4K20me3-weak loci. **d**, Barplots comparing the locus characteristics such as mCH desert(Lister et al. 2013) and SSDRs(Lilue et al. 2018; Lilue et al. 2019) between H4K20me3-weak loci (n = 455) and all 200-kb loci (n = 12,562) in Purkinje cells. **e**, Gene family characteristics either mCH desert (Lister et al. 2013), SSDRs (Lilue et al. 2018; Lilue et al. 2019), or both (left) and their radial positioning relative to nuclear periphery and H4K20me3 enrichment (middle), as well as their enrichments across chromosomes (right). Only a subset of genomic loci annotated with those categories with corresponding gene family names were included in this analysis. In box plots, the center lines for the median, boxes for the interquartile range, whiskers for values within 1.5 times the interquartile range, and points for outliers in **b**, **c**, **e**. **f**, H4K20me3 and H4K20me3-weak loci (n = 236, 455, respectively) distribution across chromosomes in Purkinje cells. **g**, GO term comparison between H4K20me3 and H4K20me3-weak associated genomic loci in Purkinje cells revealed enrichment of distinct gene families at each category. **h**, Additional visualization of H4K20me3-enriched *Vmn* and *Olfr* gene family loci overlaid with H4K20me3 staining with a maximum z-projection of four sections in Purkinje cells. **i**, Additional examples of ensemble-averaged and single allele chromatin profiles sorted by H4K20me3 enrichment from bottom to top in Purkinje cells. **j**, Comparison of ensemble-averaged radial positioning of genome-wide DNA loci in the nucleus across cell types separates neurons and glial cells. **k**, Comparison between RNAPIISer5-P enrichment and

spatial distance from nuclear periphery for individual loci. Top 5% H4K20me3-enriched loci (purple, *n* = 277) in each cell type tend to localize in the nuclear interiors yet excluded from RNAPIISer5-P in neurons (Purkinje, MLI1, and MLI2/PLI) but not in Bergmann glia cells. Those neuron-specific outlier loci found here were not observed in human cell culture(Chen et al. 2018). **l**, Transition of H4K20me3-enriched loci from Purkinje to MLI2/PLI cells showed largely conserved H4K20me3-associated loci between neuronal cell types. *n* = 252 loci. **m**, The H4K20me3-enriched regions in Purkinje cells shown with radial positioning of genomic loci (top) or H4K20me3 and Lamin B1 chromatin profiles (bottom) across cell types. Cell-type specific changes of radial positioning of genomic loci can be observed at the H4K20me3 enriched regions identified in Purkinje cells, highlighted as gray shaded regions (top). **n**, Comparison of locus identity overlap between H4K20me3-associated loci in this study and radial down loci changing their localization from the nuclear periphery to interior in forebrain neurons during postnatal development (Tan et al. 2021), showing a high degree of overlap (60.5% and 33.6%, top and middle) relative to the control loci without H4K20me3 enrichment (7.8%, bottom) despite using different methods applied to distinct mouse brain regions. 200 kb binning (*n* = 12,562 loci in total) was used for the analysis in **a**-**g**, **i**-**n**. *n* = 4,015 cells from two biological replicates of the adult mouse brain cerebellum.

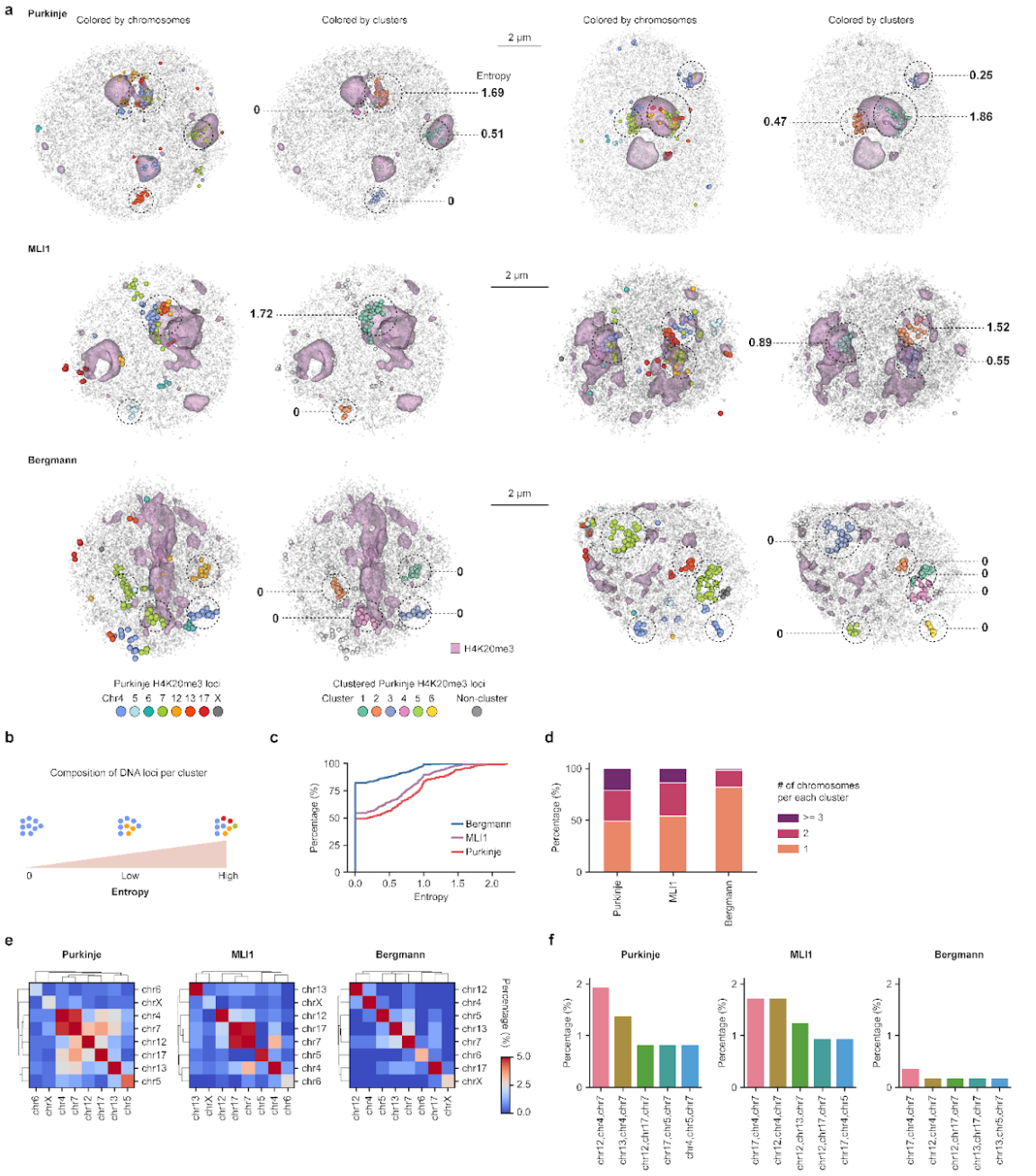


Extended Data Fig. 15| Additional visualization and characterization of 3D genome

organization associated with subnuclear structures.

a, Ensemble-averaged spatial distances between pairs of genomic loci along with chromatin profiling from each cell type across chromosomes (Chr1-19, X). **b**, Additional examples of ensemble-averaged spatial distances between pairs of genomic loci along with chromatin profiling from each cell type at specific chromosomes. Locations of pairs of H4K20me3-associated loci in Purkinje cells in Fig. 5c are highlighted as black circles in chromosome 12. **c**, Cumulative distribution of inter-chromosomal distances between pairs of loci with top 5% association to MajSat or Purkinje-specific H4K20me3 characterized in Fig. 5c (left) or from loci associated with specific gene families (right) compared to random pairs of loci (n = 1,000 trials). **d**, Representative 3D images of the nucleus with MajSat or H4K20me3 staining and chromosomal loci in each cell type. Identified Purkinje H4K20me3 loci highlighted with colors tend to colocalize at H4K20me3 territories in Purkinje and MLI1 cells, but localize at the nuclear periphery in Bergmann glia cells. Bergmann glia cells are shown as a side view for visual clarity. Additional panels (right) represent a zoom-in view of the specific H4K20me3-associated genomic regions identified in Purkinje cells and their neighboring genomic regions (2 Mb upstream and downstream regions). The view points are rotated from the original 3D images of the nucleus for visual clarity. **e**, Ensemble-averaged spatial distance between pairs of genomic loci as a function of genomic distance in chromosome 4 (top) and 12 (bottom). Pairs of H4K20me3 loci (orange) identified in Purkinje cells showed closer spatial distances compared to the similar genomic distance pairs in Purkinje, MLI1, and Bergmann glia cells, but not in mESCs. To evaluate this more quantitatively, we performed 1,000 rounds of bootstrap iterations for each genomic distance and calculated 95% confidence intervals (CI). For chromosome 4, in Purkinje and MLI1 cells, all of the intra-chromosomal locus pairs had pairwise distances falling below the lower bound of 95% CI. For Bergmann glia, 95.2% of genomic loci pairs were below the lower bound and only 30.0% of pairs in mESCs had pairwise distance smaller than lower bound. Similarly, for chromosome 12, all intra-chromosomal genomic loci pairs had pairwise distances smaller than the lower bound of 95% CI in Purkinje cells. In MLI1 and Bergmann glia, those percentages were 83.3% and 66.7% respectively, while none of the pairs in mESCs fell below the lower bound. **f**, Comparison of normalized inter-chromosomal distances of 1.5 Mb paint

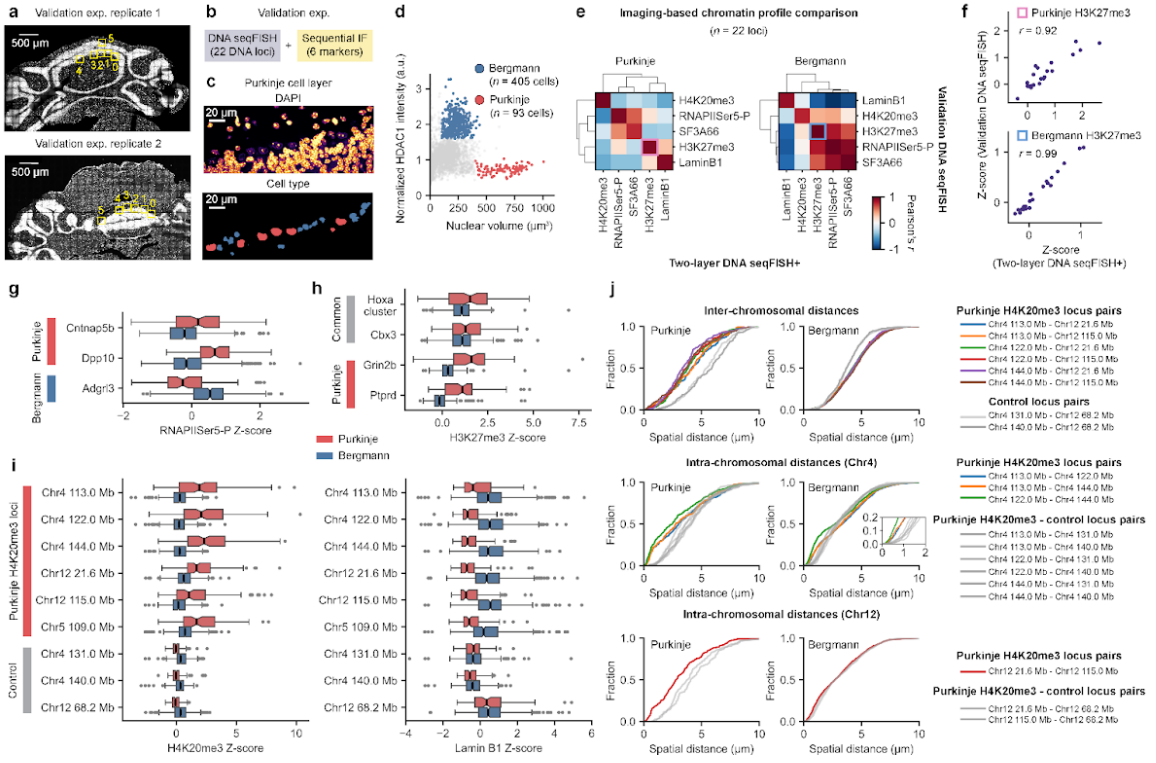
blocks that contain transcriptionally “on” or “off” long genes, as well as paint blocks that are located within nuclear speckle and RNAPIISer5-P broad peaks. Median distances from one thousand rounds of random bootstraps were used as expected inter-chromosomal distances. We note that we observed nuclear speckle-associated genomic loci are more enriched with inter-chromosomal interactions compared to the transcriptionally active long gene loci, consistent with a recent finding (Zhao et al. 2023). 1.5 Mb binning (n = 1,678 loci in total), grouped by the chromosome paint block barcodes, was used for the analysis in **a**-**c**, **e**, **f**. n = 128, 263, and 518 cells for Purkinje, MLI1, and Bergmann glia cells from two biological replicates of the adult mouse cerebellum. n = 1,076 cells from two biological replicates of mESCs.



Extended Data Fig. 16| Additional visualization and characterization of 3D genome organization around H4K20me3.

a, Representative 3D visualization of H4K20me3-associated genomic loci identified in Purkinje cells (left in each panel) and clustering results of their organization (right in each panel). Each cluster is circled out, and the corresponding inter-chromosomal entropy scores

are shown on the side. The entropy score is calculated based on chromosomal distribution within each cluster, with a score of 0 indicating that loci in a given cluster all come from a single chromosome. Higher entropy score reflected a higher level of inter-chromosomal interaction among the loci within a cluster. **b**, Illustration of the inter-chromosomal entropy score for each cluster formed by H4K20me3-associated genomic loci. The different colored spots represent DNA loci from different chromosomes. **c**, The cumulative distribution plot showing the differences of the cluster entropy scores across different cell types. The initial value at entropy zero represents the percentage of single-chromosome clusters. **d**, Stacked barplot comparing the cell-type specific differences of number of unique chromosomes in each cluster formed by H4K20me3-associated loci. In Purkinje cells, MLI1, and Bergmann glia, 49.3, 54.3, 82.1% of cluster formed by single chromosome, 29.6, 32.0, 16.5% of cluster formed by two chromosomes, 21.1, 13.7, 1.4% of cluster formed by three or more chromosomes, suggesting that a significantly larger percentage of clusters formed a higher-order chromosome organization with those loci in Purkinje cells and MLI1 compared to Bergmann glia. **e**, Heatmap of the percentage of chromosomal co-occurrence within H4K20me3-associated loci formed clusters, capturing the cell-type specific inter-chromosomal organization for Purkinje cells and MLI1 whereas Bergmann glia lacks such inter-chromosomal organization at those loci. **f**, Percentage of clusters formed by the H4K20me3-associated loci from three chromosomes. Top 5 unique chromosome compositions are shown for each cell type. Original 25-kb DNA loci without binning were used for the analysis in **c**-**f**. n = 361, 637, and 1,104 clusters from 128, 263, and 518 cells (Purkinje cells, MLI1, and Bergmann glia) from two biological replicates of the adult mouse cerebellum in **c**-**f**.



Extended Data Fig. 17| Validation imaging experiments with DNA seqFISH and sequential immunofluorescence in the adult mouse brain cerebellum.

a, Large-field DAPI images of the adult mouse brain cerebellum coronal sections from two biological replicates for validation imaging experiments. Yellow boxes represent unique fields of view (FOVs) imaged in each biological replicate. **b**, Summary of the design of the validation experiments. **c**, Example images of Purkinje cell layer by DAPI staining (top) and identified cell type identities (red, Purkinje cells and blue, Bergmann glia) (bottom). **d**, Identification of Purkinje cells and Bergmann glia using nuclear volume and HDAC1 intensity, a marker for Bergmann glia. HDAC1 immunofluorescence intensity was normalized by the mean intensity for each biological replicate. **e**, Degree of similarity of chromatin profiles between two-layer DNA seqFISH+ (x-axis) and validation DNA seqFISH (y-axis) evaluated by Pearson correlation for two cell types, representing the robustness of our chromatin profiling approach. **f**, Representative comparison of chromatin profiles of the identical marker (H3K27me3) between two-layer DNA seqFISH+ and validation DNA

seqFISH for two cell types. The systematic comparison is shown in **e**. **g**, Validation of cell-type specific RNAPIISer5-P enrichment at the identified long gene loci, confirming the enrichment in the expected cell types (Cntnap5b, Dpp10 in Purkinje cells and Adgrl3 in Bergmann glia). **h**, Validation of common and cell-type specific H3K27me3 enrichment at the identified gene loci, confirming the enrichment of Hoxa cluster and Cbx3 in both cell types and Grin2b and Ptprd only in Purkinje cells. **i**, Validation of enrichment of H4K20me3 (left) and Lamin B1 (right) at the identified H4K20me3-associated genomic loci, confirming the enrichment of H4K20me3 in Purkinje cells and Lamin B1 in Bergmann glia. Control loci from the same chromosomes are also shown for comparison. **j**, The cumulative distribution of spatial distances between pairs of genomic loci. The 3D genome organization of identified H4K20me3-associated loci in Purkinje cells was compared for inter-chromosomal pairs (top), intra-chromosomal pairs in chromosome 4 (middle) and chromosome 12 (bottom), confirming the closer spatial distances in pairs of H4K20me3-associated loci relative to control pairs in Purkinje cells. Similarly, we confirmed the larger spatial distances for inter-chromosomal pairs and shorter spatial distances in intra-chromosomal pairs in chromosome 4 in Bergmann glia cells as expected. We note that we did not separate homologous chromosomes in the intra-chromosomal analysis and thus half of the intra-chromosomal distances are expected to be computed between homologous chromosomes. n = 93 and 405 cells for Purkinje and Bergmann glia cells from two biological replicates of the adult mouse cerebellum for validation experiments in **a**-**j**. n = 128 and 518 cells for Purkinje cells and Bergmann glia from two biological replicates of the adult mouse cerebellum for two-layer DNA seqFISH+ experiments in **e**, **f**.

2.8 METHODS

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

RNA seqFISH+ encoding strategy

To spatially resolve mRNA profiles for 1,194 genes in the cell culture experiments (see 'Cell culture experiment'), we used a modified version of RNA seqFISH+ encoding scheme(Eng et al. 2019) consisting of 16-pseudocolor bases with 4 rounds of barcoding including one-error correction round, which can accommodate up to 4,096 (= 16^3) genes, in one fluorescent channel (635 nm) (Supplementary Table 1). Additional 64 genes for mRNA and non-coding RNA species were encoded as a non-barcoded seqFISH scheme (Shah et al. 2016; Shah et al. 2018; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) in one fluorescent channel (488 nm). In addition, to spatially resolve the nascent transcriptome (17,856 genes) typically at their transcription active sites in the nucleus (Shah et al. 2018), we applied a modified version of RNA seqFISH+ encoding scheme (Eng et al. 2019) consisting of 12-pseudocolor bases with 5 rounds of barcoding including one-error correction round, which can accommodate up to 20,736 (= 12^4) genes, in one fluorescent channel (561 nm). This intron seqFISH+ approach allows us to super-resolve subcellular localization of intronic RNAs at the sub-pixel resolution while original intron seqFISH decoding for 10,421 genes (Shah et al. 2018) was performed with pixel resolution using three fluorescent channels.

Similarly, for mouse brain cerebellum experiments (see 'Tissue slice experiment'), we profiled 60 mRNA species including cell-type marker genes in the adult mouse brain cerebellum (Kozareva et al. 2021) with the non-barcoded seqFISH strategy in one fluorescent channel (635 nm) as well as 17,856 genes for nascent transcripts by intron seqFISH+ in one fluorescent channel (561 nm) as described above.

Two-layer DNA seqFISH+ encoding strategy

To spatially resolve whole mouse chromosomes with 25-kb bin coverage, a two-layer barcoding strategy, consisting of diffraction limited locus imaging and chromosome paint imaging, was developed. First, mm10 mouse genome was divided into non-overlapping 25-kb loci and up to 60 loci were grouped together to make 1.5-Mb chromosome paint blocks. Those chromosome paint blocks were then separated into three groups according to their genomic coordinates in order to be encoded by three orthogonal fluorescent channels (n = 560, 559, and 559 blocks used in 635 nm, 561 nm, and 488 nm fluorescent channels). In total, 96 rounds of imaging were performed to decode the 100,049 loci encoded in two-layer DNA seqFISH+ (Supplementary Table 2). In the initial 60 rounds of imaging, 25-kb loci were sequentially read out one at a time for all chromosome paint blocks based on their genomic coordinates within each block in each fluorescent channel. These 60 rounds can resolve the identities of 25-kb loci within each chromosome paint block but cannot distinguish which specific chromosome paint block those loci belong to. In the subsequent 36 rounds of imaging, chromosome paint block identities were decoded by painting the individual 1.5-Mb blocks using a 9-pseudocolor base seqFISH+ coding scheme (Eng et al. 2019) with 4 rounds of barcoding in each fluorescent channel. This allows to resolve up to 729 (= 9^3) chromosome paint blocks in each fluorescent channel with one extra round for a stringent decoding. While original implementations of seqFISH+ (Eng et al. 2019; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) barcoded individual diffraction limited spots, this strategy barcodes individual chromosome paint blocks with unique pseudocolor combinations. This two-layer DNA seqFISH+ strategy, leveraging two layers of orthogonal barcoding (i.e. sequential barcoding of diffraction limited loci and scalable barcoding of chromosome paint blocks), can efficiently encode up to 131,220 (= 60 × 729 × 3) genomic loci within 96 rounds in three fluorescent channels, which are sufficient to accommodate all 25-kb loci in the mouse and human genome.

Primary probe design and synthesis

mRNA seqFISH+, intron seqFISH+, and non-barcoded RNA seqFISH probes were designed as described before (Shah et al. 2018; Eng et al. 2019; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). In brief, 35-nt RNA target binding sequences, 15-nt readout probe binding sites (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021), and a pair of 20-nt primer binding sites at 5' and 3' end of the primary probes were concatenated (up to 150-nt for mRNA seqFISH+ and non-barcoded RNA seqFISH; up to 170-nt for intron seqFISH+) to allow enzymatic probe amplification steps below. We used 32 primary probes per gene for mRNA seqFISH+, 12-25 primary probes (25 probes whenever possible) per gene for intron seqFISH+, and 8-50 primary probes per gene for non-barcoded RNA seqFISH.

For two-layer DNA seqFISH+ primary probes, the chromosomal DNA binding sequences were selected from mm10 newBalance DNA FISH probes at PaintSHOP resources (Hershberg et al. 2021)(https://github.com/beliveau-lab/PaintSHOP_resources). Specifically, primary probes with reported off-target scores less than or equal to 200 were initially selected and the total number of probes in each defined 25-kb locus was counted. Then the primary probes were sorted by the off-target score and 30 probes were selected from the smallest off-target score per genomic locus while the genomic locus with less than 30 probes were filtered out. After the selection, the primary probe sequences were assembled with readout probe and primer binding sites similarly to the DNA seqFISH+ study (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) with modified combinations of readout probe binding sites based on the two-layer DNA seqFISH+ coding scheme (see 'Two-layer DNA seqFISH+ encoding strategy'). At each 25-kb locus targeted, we used 30 primary probes selected with the criteria above to image individual loci as diffraction limited spots. Those primary probes (up to 170-nt) consist of the genomic region specific sequences (30-37-nt)(Hershberg et al. 2021) flanked by spacer sequences ("AA" and "A"), three identical 15-nt readout binding sites, corresponding to one of the 60 rounds of sequential diffraction limited spot imaging (hybridizations 1-60) in each fluorescent channel, three 15-nt readout binding sites, corresponding to three out of four chromosome paint barcoding rounds (hybridizations 61-96) in each fluorescent channel, and a pair of 20-nt primer binding sites at 5' and 3' end of the primary probes.

RNA seqFISH+ and two-layer DNA seqFISH+ primary probes were generated from oligo array pools (Twist Bioscience for 150-nt oligos; Agilent Technologies for 170-nt oligos) based on Oligopaint technologies (Beliveau et al. 2012) with enzymatic amplifications as described previously(Shah et al. 2018; Eng et al. 2019; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). Total of 3,001,470 two-layer DNA seqFISH+ primary probes were obtained from 13 different oligo array pools. Briefly, we performed PCR with Q5 Hot Start High-Fidelity (NEB M0494S), column purification with QIAquick PCR Purification Kit (Qiagen 28104), *in vitro* transcription (NEB E2040S) at 42°C for 8 hours in the presence of RNasin Ribonuclease Inhibitor (Promega N2111) and Pyrophosphatase (NEB M0361S), and reverse transcription (Thermo Scientific EP0751) at 50°C for 2 hours and then at 55°C for 2 hours followed by heat inactivation at 85°C for 5 minutes, RNA hydrolysis by 1 M NaOH at 65°C for 15 minutes, and neutralization by an equal amount of 1 M acetic acid. Then generated primary probes were concentrated by ethanol precipitation, pooled if necessary, then dried to a powder by speed-vac, and stored at -20°C. The primer pairs for the amplification were selected from previous studies (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021).

The DNA FISH probes for mouse repetitive elements (LINE1, SINEB1, Telomere, MinSat, MajSat, and rDNA) and 3632454L22Rik fiducial marker were designed and synthesized as described before(Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). Similarly, RNA FISH probes for mouse repetitive elements were designed and synthesized for this study. In brief, SINEB1 and two orthogonal sets of MERVL RNA FISH probes were generated using genomic DNA of E14 mESCs extracted with DNeasy Blood and Tissue Kits (Qiagen 69504) as a template for PCR, followed by in vitro transcription and reverse transcription. The PCR primers consisted of genomic DNA binding sites, originally designed for RT-qPCR (Zhang et al. 2019), and overhangs of readout probe binding sites. Single RNA FISH probes targeting MajSat (both sense and antisense strands), telomeric repeat-containing RNA (TERRA), and Nsmce2 intronic RNA repetitive regions used as internal fiducial markers were designed with overhangs of readout probe binding sites and purchased from Integrated DNA Technologies.

Readout probe design and synthesis

To implement the two-layer DNA seqFISH+ strategy, 96 unique readout probes were used in each fluorescent channel for a total of 288 unique readout probes for 3 fluorescent channels. The readouts probe sequences were obtained from our previous DNA seqFISH+ studies (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) as well as additional orthogonal readout probe sequences were generated and validated similarly to those previous studies. The RNA seqFISH+ readout probe sequences were selected from a subset of the two-layer DNA seqFISH+ readout probe sequences. The readout probe sequences (12-15-nt) for sequential immunofluorescence were selected from our previous studies (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) and further designed and validated with the same criteria for this study. The fluorescently-labeled readout probes (Integrated DNA Technologies) that can bind to the readout sequences on the primary probes or primary antibodies were conjugated in-house to Alexa Fluor 647–NHS ester (Invitrogen A20006), Cy3B–NHS ester (GE Healthcare PA63101), or Alexa Fluor 488–NHS ester (Invitrogen A20000) as described before (Eng et al. 2019) or directly purchased (Integrated DNA Technologies).

DNA-antibody conjugation

The oligonucleotide-conjugated antibodies were prepared similarly to those previously described (Gong et al. 2016; Agasti et al. 2017; Takei, Yun, et al. 2021). The BSA-free primary antibodies were purchased from commercial vendors whenever possible. For the BSA-free primary antibodies, we used the crosslinking of 5’ thiol-modified 18-nt DNA oligonucleotides (Integrated DNA Technologies) to lysine residues on antibodies via PEGylated SMCC cross-linker (SM(PEG)2) (Thermo Scientific Thermo Scientific 22102) (Agasti et al. 2017; Takei, Yun, et al. 2021). In addition, for some of the BSA-free primary antibodies, we used the crosslinking of 5’ azide-modified 18-nt DNA oligonucleotides (Integrated DNA Technologies) to lysine residues on antibodies via DBCO-PEG4-NHS (Sigma-Aldrich 764019) cross-linker with modifications from previous protocols(Gong et al. 2016). In brief, primary antibodies (90-100 μg) were buffer-exchanged to 1× PBS (Invitrogen AM9624) using 50KDa Amicon Ultra Centrifugal Filter Unit (Millipore,

UFC505096) and reacted with 10 equivalents of DBCO-PEG4-NHS at 4°C for 4–6 hours. Then the antibody solution was exchanged with 1× PBS using the 50KDa Amicon Ultra Centrifugal Filter Unit and reacted with 10 equivalents of the azide-modified 18-nt DNA oligonucleotides at 4°C for 48 hours. The DNA-primary antibody conjugates were washed with 1× PBS and concentrated using the 50KDa Amicon Ultra Centrifugal Filter Unit. For BSA-containing primary antibodies, we used SiteClick R-PE Antibody Labeling Kit (Life Technologies S10467) to crosslink 5' DBCO-modified 18-nt DNA oligonucleotides (Integrated DNA Technologies) to the specific sites on primary antibodies (Agasti et al. 2017; Takei, Yun, et al. 2021). In all conjugation strategies, excess oligonucleotides in the antibody solution were removed using the 50KDa Amicon Ultra Centrifugal Filter Unit at the final step. The oligonucleotide-conjugated primary antibodies were individually validated by immunofluorescence for their subcellular localization patterns and stored in 1× PBS at −80 °C as small aliquots. The oligonucleotide-conjugated primary antibodies are listed in Supplementary Table 3.

Cell culture and preparation

E14 mouse embryonic stem cells (E14Tg2a.4) (RRID:MMRRC_015890-UCD) from Mutant Mouse Regional Resource Centers were maintained on 0.1% gelatin (Sigma-Aldrich G1393) coated 6-well plates under the serum/LIF condition containing 15% ES-grade FBS (Gibco 16141061), 1,000 units/mL leukemia inhibitory factor (LIF) (Sigma-Aldrich ESG1106), 1× non-essential amino acids (Gibco 11140050), 1 mM sodium pyruvate (Gibco 11360070), 55 μM 2-mercaptoethanol (Gibco 21985023), 1× penicillin, and streptomycin (Gibco 15140122) in DMEM GlutaMAX (Gibco 10566016). NMuMG mammary gland cells (ATCC CRL-1636) were maintained on 6-well plates (Thermo Scientific 150687) in 10% FBS (Corning 35-010-CV), 10 μg/mL insulin (Gibco 12585014), 1× penicillin and streptomycin (Gibco 15140122), and DMEM (Corning 10-013-CV).

For live cell imaging of E14 cells with nuclear pore (Nup37) staining, the cells were transfected with a linearized Nup37-EGFP plasmid (Addgene plasmid # 104562) using FuGENE HD transfection reagent (Promega E2311). The polyclonal E14 cells were selected

with 1 μg/mL puromycin, and coverslips were subsequently prepared in the same manner as the seqFISH+ experiments as described below.

The coverslips were prepared as previously described (Takei, Yun, et al. 2021). In brief, E14 cells were plated on poly-D-lysine (Sigma-Aldrich P6407) and human laminin (BioLamina LN511) coated coverslips (25 mm x 60 mm) and incubated for 24 hours. Similarly, NMuMG cells were plated on poly-D-lysine (Sigma-Aldrich P7280) and human laminin (BioLamina LN511) coated coverslips and incubated for 24 hours. The cells were washed with 1× PBS (Invitrogen AM9624) once and fixed with freshly prepared 4% formaldehyde (Thermo Scientific 28908) in 1× PBS at room temperature for 10 minutes. The fixed cells were washed with 1× PBS a few times and stored in 70% ethanol at -20°C (Finn et al. 2019; Takei, Yun, et al. 2021) until the cell culture experiment below.

Cell culture experiment

The samples for cell culture experiments were prepared similarly to our previous study (Takei, Yun, et al. 2021) with some modifications. First, the samples were permeabilized and prepared for sequential immunofluorescence. The fixed cells on the coverslips were dried and permeabilized with 0.5% Triton-X (Sigma-Aldrich 93443) in 1× PBS at room temperature for 15 minutes using a sterilized custom-made chamber with a silicon plate (McMASTER-CARR 86915K16) attached on each coverslip. The samples were then washed three times with 1× PBS and blocked at room temperature for 15 minutes with a blocking solution consisting of 1× PBS, 10 mg/mL UltraPure BSA (Invitrogen AM2616), 0.3% Triton-X, 0.1% dextran sulfate (Sigma D4911) and 0.5 mg/mL sheared Salmon Sperm DNA (Invitrogen AM9680). Then DNA oligo-conjugated primary antibodies were incubated in the blocking solution with 100-fold diluted SUPERase In RNase Inhibitor (Invitrogen AM2694) at 4°C overnight. The typical estimate of the final concentration of each primary antibody in the blocking solution was 5-10 ng/μL. After DNA oligo-conjugated primary antibody incubation, the samples were washed with 1× PBS three times and incubated at room temperature for 15 minutes. The samples were then post-fixed with freshly made 4% formaldehyde in 1× PBS at room temperature for 5 minutes, washed with 1× PBS six times,

and incubated at room temperature for 15 minutes. The samples were further post-fixed with 1.5 mM BS(PEG)5 (PEGylated bis(sulfosuccinimidyl)suberate) (Thermo Scientific A35396) in 1× PBS at room temperature for 20 minutes and quenched with 100 mM Tris-HCl pH7.4 (Alfa Aesar J62848) at room temperature for 5 minutes. We note these post-fixation steps allow the stabilization of antibodies and samples during the heating step required for DNA FISH preparation. After the post-fixation steps, the samples were washed with 1xPBS three times and air dried upon removing the custom silicon chamber.

After the sequential immunofluorescence preparation above, the samples were prepared for RNA seqFISH steps. The custom-made flow cells (fluidic volume ~30 μl), made from glass slide (25 x 75 mm) with 1 mm thickness and 1 mm diameter holes and a PET film coated on both sides with an acrylic adhesive with total thickness 0.25 mm (Grace Bio-Labs RD481902), were attached to the prepared coverslips. The samples were rinsed with 4× SSC and hybridized with RNA seqFISH primary probe pools (1-10 nM per probe), 4 nM Nsmce2 fiducial marker probe, and 10 nM polyT LNA oligo (Qiagen) in a 50% hybridization buffer consisting of 50% formamide (Invitrogen AM9342), 2× SSC, and 10% (w/v) dextran sulfate (Millipore 3710-OP). The hybridization was performed at 37°C for 48-72 hours in a humid chamber. After primary probe hybridization, the samples were washed with a 55% wash buffer consisting of 55% formamide, 2× SSC, and 0.1% Triton X-100 at room temperature for 30 minutes and rinsed three times with 4× SSC. Then the samples were imaged for RNA seqFISH as described below (see ‘Sequential imaging’).

After RNA seqFISH imaging, the samples were prepared for DNA seqFISH+ steps. The samples were taken from the microscope, rinsed with 1× PBS, and incubated with 100-fold diluted RNase A/T1 Mix (Thermo Fisher EN0551) in 1× PBS at 37°C for 1 hour to digest RNA species and remove RNA seqFISH+ primary probes. The samples were then rinsed three times with 1× PBS and three times with a 50% denaturation buffer consisting of 50% formamide and 2× SSC and incubated at room temperature for 15 minutes. Following the incubation, the samples were heated on the heat block at 90°C for 5 minutes in the 50% denaturation buffer with aluminum sealing tapes (Thermo Scientific 232698) on the inlet and

outlet of the custom chamber. Immediately after heating, the samples were rinsed with 4× SSC and hybridized with a DNA seqFISH+ primary hybridization buffer consisting of two-layer DNA seqFISH+ probes (~0.2 nM per probe), rDNA probes (~10 nM per probe), ~1 μM LINE1 probe, ~1 μM SINEB1 probe, 100 nM 3632454L22Rik fiducial marker probe (Integrated DNA Technologies), 35% formamide, 2× SSC, and 10% (w/v) dextran sulfate (Millipore 3710-OP) at 37°C for 5-7 days in a humid chamber. After the hybridization step, the samples were washed with a 35% wash buffer, consisting of 35% formamide, 2× SSC, and 0.1% Triton X-100, for six times and then incubated at room temperature for 15 minutes, followed by rinsing for three times with 4× SSC.

After DNA seqFISH+ hybridization, the samples were further processed to stably maintain DNA seqFISH+ primary probes on the chromosomal DNA during imaging routines. First, 250 nM global ligation bridge oligo(Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021)(Integrated DNA Technologies) was hybridized in a 20% hybridization buffer consisting of 20% formamide, dextran sulfate (Sigma D4911), and 4xSSC at 37°C for 2 hours. The samples were then washed three times with a 12.5% wash buffer consisting of 12.5% formamide, 2× SSC, and 0.1% Triton X-100, incubated at room temperature for 5 minutes, and rinsed three times with 1× PBS. Next, to perform ligation between 5'- and 3'-end of the DNA seqFISH+ primary probes with the hybridized ligation bridge oligo, the samples were incubated with 20-fold diluted Quick Ligase in 1× Quick Ligase Reaction Buffer from Quick Ligation Kit (NEB M2200) supplemented with additional 1 mM ATP (NEB P0756) at room temperature for 1 hour. The samples were then washed with the 12.5% wash buffer and rinsed three times with 1× PBS. After the ligation steps above, the samples were further processed with amine modification and post-fixation to synergistically stabilize the probes as we demonstrated before (Takei, Yun, et al. 2021). For the amine modification, the samples were rinsed with 1× Labeling Buffer A and incubated with 10-fold diluted Label IT Amine Modifying Reagent in 1× Labeling Buffer A from Label IT Nucleic Acid Modifying Reagent (Mirus Bio MIR 3900) at room temperature for 45 minutes. After three rinses with 1× PBS, the samples were post-fixed with 1.5 mM BS(PEG)5 in 1× PBS at room temperature for 30 minutes and quenched with 100 mM Tris-HCl pH7.4 at room temperature

for 5 minutes. Then the samples were washed with the 55% wash buffer at room temperature for 5 minutes, rinsed with 4× SSC for three times, and stored in 4× SSC at 4°C until the imaging for DNA seqFISH+ and sequential immunofluorescence (see 'Sequential imaging').

Tissue slice experiment

All animal care and experiments were carried out in accordance with Caltech Institutional Animal Care and Use Committee (IACUC) and NIH guidelines. 6-week-old C57BL/6J female mice from The Jackson Laboratory (Stock No: 000664 | B6) were used for the cerebellum tissue slice experiments. The brain samples and coverslips with 15-20 μm coronal sections of the cerebellum were prepared similarly to those described before(Takei, Zheng, et al. 2021).

The tissue slice experiments were performed similarly to the cell culture experiment (see 'Cell culture experiment') and our previous mouse cortex experiment (Takei, Zheng, et al. 2021) with some modifications. In brief, the permeabilization and sequential immunofluorescence were performed as described before (Takei, Zheng, et al. 2021). After the sequential immunofluorescence preparation, custom-made flow cells (fluidic volume about 40 μl) were attached to the coverslips. Then the RNA seqFISH preparation and imaging were performed similarly to the cell culture experiment with a different set of non-barcoded mRNA seqFISH primary probes including mouse cerebellum marker genes (Kozareva et al. 2021), intron seqFISH+ probes, and polyT LNA oligo (Qiagen). After RNA seqFISH imaging, the samples were prepared for DNA seqFISH+ steps similarly to the cell culture experiment except the extended 90°C heating time to 6 minutes, followed by DNA seqFISH+ and sequential immunofluorescence imaging as described below (see 'Sequential imaging').

Automated microscope setup

All imaging experiments were performed with the confocal fluorescence imaging platform and fluidics delivery system as described before (Shah et al. 2018; Eng et al. 2019; Takei,

Yun, et al. 2021; Takei, Zheng, et al. 2021). In brief, the microscope (Leica DMi8) was equipped with a confocal scanner unit (Yokogawa CSU-W1), a sCMOS camera (Andor Zyla 4.2 Plus), a 63× oil objective (NA = 1.40, Leica 11506349), Borealis beam conditioning unit (Andor), a motorized stage (ASI MS2000), fiber coupled lasers (635, 561, 488, and 405 nm) from CNI and Shanghai Dream Lasers Technology, and filter sets from Semrock. In addition, the custom-made automated sampler was set up for automated buffer exchange coupled with hybridization and imaging routines (see 'Sequential imaging') to move to the well of the designated hybridization buffer corresponding to each hybridization round from a 2.0-mL 96-well plate (Corning 3960). The hybridization buffer and other buffers were moved through a multichannel fluidic valve (IDEX Health & Science EZ1213-820-4) to the custom-made flow cell with a syringe pump (Hamilton Company 63133-01). The automated fluidics delivery and imaging were controlled by a custom-written script in μManager (Edelstein et al. 2010).

Sequential imaging

The seqFISH hybridization and imaging routines were performed as described previously (Shah et al. 2018; Eng et al. 2019; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). In brief, the sample with the custom-made flow cell was connected to the automated fluidics system on the microscope. The field of views (FOVs) for the images were registered based on the DAPI-based nuclear staining. Then imaging of RNA seqFISH+ as well as sequential immunofluorescence for some targets was performed with the sequential hybridization and imaging routines described below. The samples were then disconnected from the fluidics system and proceeded to the DNA seqFISH+ preparation (see 'Cell culture experiment' and 'Tissue slice experiment'). Next, the registered FOVs for RNA seqFISH+ were loaded and manually shifted to find the same cells in the original FOVs, followed by DNA seqFISH+ and sequential immunofluorescence imaging using the sequential hybridization and imaging routines described below.

The sequential hybridization and imaging routines were performed at room temperature on the automated confocal microscope. Briefly, for the sequential hybridization routine, the serial hybridization buffer, which consisted of a mixture of two or three unique readout

probes (10-250 nM) with different fluorophores (Alexa Fluor 647, Cy3B or Alexa Fluor 488) and 10% EC buffer (10% ethylene carbonate (Sigma E26258), 10% dextran sulfate (Sigma D4911) and 4× SSC), was picked up from a 96-well plate and incubated in the flow cell for 20 minutes. The samples were then washed with 1 mL of a 4× SSCT buffer (4× SSC and 0.1% Triton-X), 330 μL of the 12.5% wash buffer, and 200 μL of 4× SSC, followed by a staining with about 200 μL of the DAPI solution consisting of 5 μg/mL DAPI (Sigma D8417) and 4× SSC for 30 seconds. The sample was then imaged with an anti-bleaching buffer consisting of 50 mM Tris-HCl pH 8.0 (Invitrogen 15568025), 4× SSC, 3 mM Trolox (Sigma 238813), 10% D-glucose (Sigma G7528), 100-fold diluted catalase (Sigma C3155), 1 mg/mL glucose oxidase (Sigma G2133). After the image acquisition detailed below, the sample was washed with 1 mL of the 55% wash buffer for 1 minute to strip off readout probes, followed by an additional incubation for 1 minute and rinsing with 4× SSC. Those serial hybridization, imaging, and signal extinguishing routines were repeated until the completion of all designated rounds.

The imaging conditions were determined based on the previous studies (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). Briefly, snapshots were acquired per fluorescent channel per field of view with 250 nm z-steps over 6 μm for cell culture experiments and 12 μm for tissue slice experiments. The pixel size for x and y is 103 nm. RNA seqFISH+ imaging was performed with 635 nm, 561 nm, and 488 nm fluorescent channels by omitting a 405 nm fluorescent channel to prevent a potential damage on the nuclei prior to DNA FISH (Takei, Zheng, et al. 2021) except for a DAPI alignment hybridization round in the end. The readout probes for fiducial markers were also included in the first 2 fluorescent channels to allow image registration at the subpixel resolution. For the tissue slice experiments, polyA staining was performed in the 488 nm fluorescent channel. DNA seqFISH+ imaging was performed with 635 nm, 561 nm, 488 nm, and 405 nm fluorescent channels with DNA seqFISH+ targets in the first 3 fluorescent channels and DAPI staining in the 405 nm fluorescent channel. The readout probes for fiducial markers were also included in the first 3 fluorescent channels to allow image registration at the subpixel resolution. Imaging for sequential immunofluorescence was similarly performed with 635 nm, 561 nm, 488 nm, and 405 nm

fluorescent channels, staining primary antibody targets in the first 2 fluorescent channels, fiducial markers in the third fluorescent channel, and DAPI in the last 405 nm fluorescent channel. At the beginning and end of the imaging rounds, fiducial marker images only with fiducial markers were obtained in the first 3 fluorescent channels for image registration. Furthermore, at the end of all imaging routines, images were manually checked and problematic imaging rounds such as off-focus and intensity saturation were repeated.

CUT&Tag

CUT&Tag(Kaya-Okur et al. 2019) experiments were performed according to the protocol of Epicypher CUTANA Direct-to-PCR CUT&Tag Protocol (v1.7) with minor modifications. In brief, for each condition, approximately 100,000 cells were harvested from mESCs or NMuMG cells by centrifugation at 600 x g for 3 minutes at room temperature. Then nuclear extraction was performed with nuclear extraction buffer, consisting of 20 mM HEPES pH 7.4 (Teknova H1030), 10 mM KCl (Invitrogen AM9640G), 0.1% Triton X-100 (Sigma-Aldrich 93443), 0.5 mM spermidine (Epicypher 21-1005), and 1× cOmplete EDTA-free Protease Inhibitor Cocktail (Sigma 11873580001) in nuclease-free water (Invitrogen 10977015), on ice for 10 minutes, followed by centrifugation at 600 x g for 3 minutes. Isolated nuclei were aliquoted to 8-strip PCR tubes containing Concanavalin A (ConA) coated magnetic beads (Epicypher 21-1401) activated by bead activation buffer, consisting of 20 mM HEPES pH 7.4, 10 mM KCl, 1 mM $CaCl_2$ (Sigma-Aldrich 21115), and 1 mM $MnCl_2$ (Sigma-Aldrich M1787), and incubated at room temperature for 10 minutes. Nuclei-bound ConA beads were buffer exchanged by using a magnetic rack to 50 µl of cold antibody 150 buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM spermidine, 1× cOmplete EDTA-free Protease Inhibitor Cocktail, 0.01% digitonin, and 2 mM EDTA (Invitrogen AM9260G)) containing primary antibodies (RNAPIISer5-P (Cell Signaling 13523S), H3K27me3 (Cell Signaling 9733S), or IgG (Epicypher 13-0042)) at a 50-fold (RNAPIISer5-P, H3K27me3) or 100-fold dilution (IgG), and incubated overnight at 4°C. The samples were then incubated with 0.5 µg of anti-rabbit IgG secondary antibody (Invitrogen 31212) in 50 µl of digitonin 150 buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM spermidine, 1× cOmplete EDTA-free Protease Inhibitor Cocktail, and 0.01% digitonin) at room temperature for 30 minutes,

followed by two washes with cold digitonin 150 buffer. The samples were then incubated with 50 µL of digitonin 300 buffer (20 mM HEPES pH 7.4, 300 mM NaCl, 0.5 mM spermidine, 1× cOmplete EDTA-free Protease Inhibitor Cocktail, and 0.01% digitonin) containing 2.5 µL of pAG-Tn5 (EpiCypher 15-1017) at room temperature for 1 hour in each tube, followed by two washes with cold digitonin 300 buffer. The samples were then incubated with 50 µL of tagmentation buffer (10 mM $MgCl_2$ (Invitrogen AM9530G) in digitonin 300 buffer) in each tube at 37°C for 1 hour, followed by a wash with TAPS buffer, consisting of 10 mM TAPS buffer pH 8.5 (Boston BioProducts BB-2375) and 0.2 mM EDTA. Then 5 µL of SDS release buffer (10 mM TAPS buffer pH 8.5 and 0.1% SDS (Invitrogen 24730020)) was added to each tube, followed by an incubation at 58°C for 1 hour and a subsequent addition of 15 µL of SDS quench buffer (0.67% Triton X-100). Then sequencing library was prepared by adding 2 µL of i5 and i7 primers (Integrated DNA technologies, 10 µM stock each) and 25 µL of Q5 High-Fidelity 2× Master Mix (NEB M0492S) to each tube, followed by an amplification for 15 (RNAPIISer5-P and H3K27me3) or 17 (IgG control) PCR cycles. The libraries were then purified by using 1.2× magnetic beads (Omega Bio-Tek M1378) according to the manufacturer's protocol. After the libraries were quantified and pooled, paired-end sequencing of CUT&Tag libraries was performed on the Illumina NextSeq 2000 sequencing systems using NextSeq 1000/2000 P2 Reagents (100 Cycles) v3.

CUT&RUN

CUT&RUN(Skene & Henikoff 2017; Skene et al. 2018) experiments were performed according to the protocol of CUTANA ChIC/CUT&RUN kit (EpiCypher 14-1048) with modifications of buffer compositions based on the Genome Organization using CUT and RUN Technology (GO-CaRT)(Ahanger et al. 2021), which successfully generated Lamin B1 profiles in differentiated cells. In brief, nuclear extraction was performed with approximately 100,000 cells from E14-mESCs or NMuMG cells for each condition, similarly to the CUT&Tag protocol described above, with modified nuclear isolation buffer(Ahanger et al. 2021), consisting of 10 mM HEPES pH7.4, 10 mM KCl, 0.1% IGEPAL CA-630 (Sigma-Aldrich I8896), 0.5 mM Spermidine, 1× cOmplete EDTA-free

Protease Inhibitor Cocktail. Isolated nuclei were aliquoted to 8-strip PCR tubes containing activated ConA beads and incubated overnight at 4°C with primary antibodies (H3K4me3 (Cell Signaling 9751T), H3K9me3 (Diagenode C15410193), Lamin B1 (Abcam ab16048), IgG (Epicypher 13-0042)) at a 50-fold dilution (H3K4me3, H3K9me3, Lamin B1) or 100-fold dilution (IgG) in a 50 µL of blocking buffer, consisting of a wash buffer (20 mM HEPES pH7.4, 150 mM NaCl, 0.1% BSA (Invitrogen AM2616), 0.5 mM Spermidine, 1× cOmplete EDTA-free Protease Inhibitor Cocktail) with 2 mM EDTA. The samples were then washed twice with cold wash buffer and incubated with 50 µL of the wash buffer containing 2.5 µL of pAG-MNase (EpiCypher 15-1016) at room temperature for 10 minutes in each tube, followed by two washes with the cold wash buffer. The samples were then incubated in the wash buffer supplemented with 2 mM $CaCl_2$ at 4°C for 2 hours, followed by an incubation with a stop buffer (Epicypher 21-1003) at 37°C for 10 minutes. Then DNA fragments were purified from supernatant using the CUTANA DNA Purification Kit (EpiCypher 14-0050), followed by sequencing library preparation using CUTANA CUT&RUN Library Prep Kit (EpiCypher 14-1002) according to the manufacturer's protocol with 12 PCR cycles. After the libraries were quantified and pooled, paired-end sequencing of CUT&RUN libraries was performed on the Illumina NextSeq 2000 sequencing systems using NextSeq 1000/2000 P2 Reagents (100 Cycles) v3.

Nuclear and cytoplasmic segmentation

The 3D nuclear and 2D cytoplasmic segmentation for individual cells were performed with a generalist, deep learning-based segmentation method, Cellpose (Stringer et al. 2021). First, for the nuclear segmentation, aligned and scaled images with Lamin B1, H3K27me3, H4K20me3, and DAPI in cell culture experiments or with BRG1, mH2A1, and DAPI in cerebellum tissue slice experiments were combined and used as 3D segmentation inputs for the Cellpose with specific parameters (flow_threshold=0.5, cellprob_threshold=0.5 in cell culture and flow_threshold=0.8, cellprob_threshold=0.5 in tissues). The original 3D nuclear segmentation labels were then eroded by 2 pixels if different labels were located at adjacent pixels to avoid potential misassignment. In addition, the labels whose centroids were located within 20 pixels from the edge of the images in x or y dimensions were filtered out. Second,

for the cytoplasmic segmentation in cell culture experiments, aligned polyA images with a maximum intensity z-projection by ImageJ were used as 2D segmentation inputs for the Cellpose. The obtained cytoplasmic labels were then converted to ImageJ ROIs, eroded by 4 pixels, and manually corrected if necessary by using an ImageJ plugin, LabelsToROIs (Waisman et al. 2021). For cerebellum tissue slice experiments, we only computed nuclear labels and did not create cytoplasmic labels, similarly to our previous tissue slice experiments (Takei, Zheng, et al. 2021). Finally, the obtained nuclear labels and cytoplasmic ROIs were loaded to MATLAB, and then were matched and renumbered by comparing the centroid location of nuclear labels with cytoplasmic ROIs. In this step, any matches without unique pairs were filtered out. The nuclear properties (e.g., volume, centroid) were then extracted from segmented nuclear labels using regionprops3 function in MATLAB. The final nuclear and cytoplasmic labels for individual cells were stored as labeled images.

Evaluation of nuclear preservation in live and fixed cells

To evaluate the nuclear height differences in live and fixed cells, we performed 3D confocal imaging of the same nuclei in live cells as well as in fixed cells after a mock seqFISH+ protocol. Specifically, we used a polyclonal E14 line stably expressing a fluorescent-tagged nuclear pore (Nup37-EGFP). After the mock seqFISH+ protocol by following the cell culture experiment protocol up to the probe hybridization step, we imaged the same nuclei using Lamin B1 immunofluorescence using Lamin B1 primary antibody (Abcam ab16048) and anti-rabbit secondary antibody (Invitrogen A32732). We also note that the minimal effect of DNA seqFISH+ heating on nuclear structures was previously validated by comparing immunofluorescence signals before and after heating at voxel levels (Takei, Yun, et al. 2021).

We first used custom trained Cellpose models to find 3D segmentations for the nuclei in all image stacks (Stringer et al. 2021; Pachitariu & Stringer 2022). We obtained 2D masks by first finding 3D binary masks for each nucleus from labeled images computed by Cellpose, then taking the maximum projection of the nuclear binary mask along the z-axis. We used

the maximum projected nuclear masks to find cropping bound for each individual nucleus and to mask out neighboring nuclei and other unwanted signals before fitting.

We analyzed the image stacks of nuclear pore GFP fluorescence in live nuclei and corresponding images of Lamin B1 stain in the same nuclei post mock protocol using gaussian fitting functions modified from a published single molecule microscopy fitting package, ADCG (Alternating Descent Conditional Gradient), to localize the fluorescence puncta in both the Nup37-GFP and Lamin B1 IF images (Boyd et al. 2017). Our package containing the modified code, seqFISH_ADCG.jl version 0.1.3, is hosted at https://github.com/CaiGroup/SeqFISH_ADCG.jl. The ADCG algorithm iteratively builds models of images by finding local maxima in the convolution of the image-model residuals, adding a new PSF to the model at the coordinates where the convolution is maximal, then refining the coordinate and brightness parameters of each PSF by gradient descent in each iteration.

For the Nup37-GFP images acquired in live cells, we used our modified 3D tiled implementation of ADCG to fit nuclear pore puncta to obtain their xyz localization coordinates. Then we filtered out punctas from outside of the nuclei. Tiles were 13x13x8 voxels (xyz) and overlapped by 3 voxels with neighbors on all sides. We aimed to set the manual z-threshold just outside the z-slice where the highest or lowest puncta peaks in intensity. For the fixed images of Lamin B1 stain, we found points on the nuclear periphery using only the first step of ADCG without tiling, by finding the local maxima above a threshold of convolution of the images with the point spread function. Performing gradient descent to optimize the fit coordinates and brightness of PSFs is not feasible for the Lamin B1 stain images because the region of Lamin B1 fluorescence is continuous. However, the Lamin B1 signals were bright and there was no difficulty in measuring the boundary of the nuclei. While the localization accuracy of the punctas are less than one z-section (250 nm), there is 1 z-section uncertainty on where the top of the nuclei are. Then the degree of difference in nuclear heights in live cells and after the mock seqFISH+ protocol was evaluated by using a two-sided paired t-test.

We note that we used laminin-coated coverslips to seed E14 cells, a condition known to promote monolayer formation rather than 3D colonies in mESCs. This condition has been used for live cell imaging of mESCs with effective cell state transitions and differentiation (Hormoz et al. 2016; Gu et al. 2018). We also note that the nuclear morphology of mESCs under this condition could differ from the nuclei extracted in suspension for single-cell Hi-C studies (Nagano et al. 2017; Stevens et al. 2017).

seqFISH image processing

The seqFISH image processing steps were performed based on previous studies(Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) with modifications. In brief, the preprocessing of the images were first performed by applying a flat field correction, a chromatic aberration correction considering full affine shifts in the X and Y dimensions (scaling, rotation, translation and shearing) but only translation in the Z dimension, and a background subtraction using the ImageJ rolling ball background subtraction algorithm with a radius of 3 pixels. Second, pixel locations for seqFISH spots were identified by using a Laplacian of Gaussians filter and a 3D local maxima finder with thresholding values obtained from semi-manual steps (Takei, Yun, et al. 2021). The identified seqFISH spot locations were further super-resolved at a sub-pixel resolution using a 3D radial center algorithm(Parthasarathy 2012; Liu et al. 2013). Third, images from different hybridization rounds were aligned to the initial hybridization (hybridization 1) image in DNA seqFISH+ at a subpixel resolution by computing the translation of identified fiducial markers in each fluorescent channel as detailed below. To align RNA and DNA images, the alignment to correct any rotation computed from DAPI staining images was further applied.

Using the aligned spots, seqFISH decoding was performed. The decoding of non-barcoded RNA seqFISH was performed based on the previous studies (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). In addition, the decoding of mRNA and intron seqFISH+ was performed similarly to those used for 1-Mb resolution DNA seqFISH+ experiments (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) with appropriate pseudocolor numbers and barcode keys

with a voxel search radius of square root of 3 and one round of the error correction. We then used seed information generated during the barcode finding in seqFISH experiments for stringency (Shah et al. 2018; Eng et al. 2019; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). For mRNA seqFISH+ for cell culture experiments, we compared seed 3 and seed 4 results of mESCs, and filtered out top 150 barcodes, including on and off targets, that are differentially identified between those. Then the results of at least 3 out of 4 seeds from remaining barcodes were used for the downstream analysis. For intron seqFISH+ experiments, we used the results of at least 4 out of 5 seeds as before (Shah et al. 2018). For the intron seqFISH+ experiments in cell culture, the distance to the nearest chromosome territory defined by the DNA seqFISH+ result was computed for each intron spot. Then intron spots within 500 nm from their own chromosome territories were considered to be transcription active sites (TASs), and other spots, which include introns outside TASs as well as false positives, were filtered out from the downstream analysis.

The decoding of two-layer DNA seqFISH+ was newly developed by leveraging the previous 25-kb resolution DNA seqFISH+ decoding approach (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). At each rounded pixel location where the spots were identified in the first 60 hybridization rounds, z-scored chromosome paint intensities for each cell and each hybridization round were computed for the next 36 hybridization rounds to provide 36 z-scored chromosome paint intensity values on each spot, corresponding to 9 pseudocolors for 4 barcoding rounds described in ‘Two-layer DNA seqFISH+ encoding strategy’. The barcodes for each spot were identified by matching the pseudocolors with the largest chromosome paint intensity z-score values in each barcoding round and compared with codewords identifying their genomic loci. To avoid false assignments, we dropped loci whose lowest chromosome paint intensity z-score in any barcoding round was above 0, or whose highest was below 0.5.

The alignment accuracy for the images was evaluated by calculating the Euclidean distances of individual aligned fiducial markers between a fiducial marker reference image and each serial round image in 3D or given dimensions at each field of view. We note that we

computed x-y-z translations of matched fiducial marker spots using a subset of fiducial markers that were successfully aligned with a given alignment setting, similarly to our previous studies (Eng et al. 2019; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). The offset of each image for the alignment was then calculated by taking an average of the translation of the identified fiducial markers in each fluorescent channel at each field of view. For the fiducial markers, we used the endogenous DNA FISH spots targeting a repetitive region at 3632454L22Rik loci in ChrX (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021), which labeled 1-2 spots per nucleus.

Sequential immunofluorescence and DNA FISH image processing

To obtain subnuclear structures for individual cells, we used sequential immunofluorescence and repetitive DNA FISH images. Similarly to image processing for the spot detection described above, the images were corrected for a chromatic aberration shift and aligned to the initial hybridization image in DNA seqFISH+ by computing and propagating the translation of identified fiducial markers in an orthogonal fluorescent channel (488 nm) that was not used for sequential immunofluorescence or repetitive DNA FISH. In contrast to image processing for the spot detection, the background subtraction processing was not applied to the images except for the subnuclear foci analysis below. The mean intensities for each nuclear label were computed for all markers using the aligned images.

Subnuclear foci analysis

The foci detection for subnuclear markers was performed as described previously (Takei, Zheng, et al. 2021). In brief, aligned sequential immunofluorescence and DNA FISH images were background subtracted using the ImageJ rolling ball background subtraction algorithm with a radius of 9 pixels and binarized by Yen's auto threshold method in ImageJ for each slice in the z-stack, followed by filling and opening steps. The subnuclear foci properties were then extracted from each of the labeled objects using regionprops3 function in MATLAB after removing objects smaller than 20 voxels or greater than 100,000 voxels and assigned to corresponding cells using nuclear labels. To compare the cell-type specific

subnuclear foci organization, the distributions of foci properties such as foci number per cell, individual foci volume, and total foci volume per cell were compared across cell types.

Conversion of voxel information to physical size

After image processing steps above, we converted the voxel information of the images or decoded spots to physical size (0.103 μm for x and y and 0.250 μm for z) for the downstream analysis below.

Separation of homologous chromosomes

In our previous works (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021), we used DBSCAN to cluster genomic loci into their separate chromosome copies based on their spatial location. This worked well in most cases, but failed in cases where two copies of a chromosome were close to each other. To address these cases, we developed a package, DNASeqFISHChromosomeAssignment.jl, that clusters loci first with DBSCAN, then refines the clustering by taking each locus's spatial and genomic coordinates into account. It checks whether each DBSCAN cluster has a proportion of unique genomic loci below a user-defined threshold and whether the average spatial distance between subsequent loci is above another user-defined threshold. If so, it splits the cluster using an algorithm that we call Longest Disjoint Paths (LDP). This algorithm conceptualizes chromosomes as invisible strings on which we can image beads at selected genomic loci with DNA seqFISH. We can infer which loci are on each copy of the chromosome by stringing together subsequent genomic loci to find a distinct path for each chromosome copy. The LDP algorithm is stringent and does not assign all valid loci to a chromosome copy on its own. To assign loci that are not included in one of the LDPs to a chromosome copy, we cluster each unassigned locus with the LDP that has the most loci within a user-specified search radius of the locus.

Finding the multiple longest disjoint paths is an integer programming problem similar to a maximum flow problem. To set up the problem, the LDP algorithm first constructs a directed graph where nodes represent loci. Loci within a user-specified spatial radius (different for different datasets) and different genomic coordinates are connected by directed edges

pointing towards the locus of larger genomic coordinates. Edges are weighted according to the formula

$$\frac{1}{g_c - g_p} e^{-\frac{||r_c - r_p||^2}{2\sigma^2}}$$

where g_p and g_c are genomic coordinates of the parent and child nodes of the directed edge, r_p and r_c are the position vectors of parent and child nodes, and σ is a user set parameter. We then reduce the number of edges in the graph by finding its transitive reduction, then removing any outgoing edge from a node that has lower weight than that node's lowest weighted outgoing edge in the transitive reduction. To finish framing the problem, we add imaginary source and destination nodes to the graph that respectively have outgoing or incoming edges of no weight connecting them to every node that represents a genomic locus.

Our integer programming problem is tasked to find the maximally weighted edges that comprise one or more disjoint paths. To enforce that all chosen edges are part of disjoint paths, we add the following constraints: the solution can choose either exactly one incoming edge and one outgoing edge or no edges from each node representing a genomic locus and no more outgoing edges from the source than the maximum allowed number of chromosomes that can be chosen.

2D and 3D visualization of DNA seqFISH+ data

The 2D visualization of DNA seqFISH+ spots was performed as described before(Bhat et al. 2023). Because raw DNA seqFISH+ images are barcoded and the identities of DNA spots are indistinguishable without decoding, we reconstructed images after decoding from rounded voxel location of the decoded DNA seqFISH+ spots by applying a multidimensional Gaussian filter (sigma = 1) with scipy.ndimage.gaussian_filter package in python and overlaid with aligned raw immunofluorescence images by using ImageJ. Those composite images are displayed in 2D either as a single z-section or a maximum z-projection of multiple z-sections as specified in each figure caption.

The 3D visualization of DNA seqFISH+ or intron seqFISH+ data was performed by using python API Mayavi2(Ramachandran & Varoquaux 2011). Physical coordinates of all detected DNA spots or decoded intron spots were stored as x, y and z numpy arrays and then visualized as spheres. DNA spheres were colored differently to represent different chromosomes or gene families, or displayed as transparent small spheres when they were shown as background. Intron spheres were colored by chromosome identities of each gene. Genomically adjacent DNA spots were connected using tubes to show the organization of a single chromosome. H4K20me3 immunofluorescence and MajSat DNAFISH signals were visualized by displaying a surface around x, y, and z coordinates with intensity z-score values above 2.

Sequencing feature analysis in the mouse genome

To compute GC content per 25 kb genomic bins, we created a bed file for the 25 kb bins with unmasked mouse mm10 reference genome and performed the computation within each bin using 'bedtools nuc'(Quinlan & Hall 2010). To compute the gene density and gene length features per 25 kb genomic bins, we downloaded GRCm38/mm10 refGene database, retrieved the longest gene annotation for each gene, and mapped back the gene name and corresponding gene length to each 25 kb genomic bin. We then further computed the GC content, gene density per 100 kb and 200 kb bins. Meidan gene length per 100 kb and 200 kb bins were defined by the median gene lengths of genes falling in those bins.

Sequencing-based data processing

CUT&Tag and CUT&RUN sequencing data produced in this study was processed using the nf-core/cutandrun pipeline (Ewels et al. 2020) (version 3.1) with Nextflow (Di Tommaso et al. 2017) (version 24.04.1). Briefly, reads were aligned to the mm10 genome using bowtie2 (Langmead & Salzberg 2012) (version 2.4.2), keeping only reads with a minimum alignment q score of 20. Then, reads were masked using the mm10 blacklist from ENCODE (Amemiya et al. 2019). Finally, bigWig files were generated from masked alignments with 100 kb bins using the bamCoverage function of deeptools (Ramírez et al. 2016) (version 3.5.1).

A/B compartment score file from the Hi-C matrix was downloaded from the original study at the 100 kb bin size for mESCs (Bonev et al. 2017) or kindly provided by Sandra Peiró at the 25 kb bin size for NMuMG cells (Pascual-Reguant et al. 2018). Compartment scores of NMuMG cells were further binned using a mean score of 25 kb bins in each 100 kb bin.

Processed bulk RNA-seq data for cell culture were obtained from NCBI GEO (accession GSE98674 for mESCs (Antebi et al. 2017) and accession GSE96033 for NMuMG cells(Pascual-Reguant et al. 2018)). For the GRO-seq analysis in mESCs (Jonkers et al. 2014), single-end FASTQ files were downloaded from NCBI GEO (accession GSE48895) and first trimmed to keep the first 32 bases using the cutadapt package (version 1.18)(Martin 2011). Trimmed reads were aligned to the mm10 genome using bowtie2 (Langmead & Salzberg 2012). Uniquely mapped reads were filtered and sorted before duplicate reads were removed using samtools (Li et al. 2009). Strand-specific RPKM bigwig signal files were generated using bamCoverage with filterRNAstrand option employed and step size of 10. Forward and reverse strand aligned gene body bed files (generated from GENCODE vM25 annotation) were used to compute GRO-seq signal intensity within gene body regions using computeMatrix command of deeptools. GRO-seq intensity within ±500bp around gene transcription start sites was calculated using a similar approach.

The processed pA-DamID data for Lamin B1 (log2-normalized LaminB1 pA-DamID scores) in mESCs was obtained from NCBI GEO (accession GSE181693)(van Schaik et al. 2022) and binned with 100 kb. The different categories lamina-associated domains (cLADs, fLADs, and ciLADs) from Lamin B1 DamID datasets were obtained from NCBI GEO (accession GSE17051) (Peric-Hupkes et al. 2010; Meuleman et al. 2013) and mm9 genomic coordinates of the obtained files were converted to mm10 using the UCSC Genome Browser program liftover. Similarly, the LADs in mESCs were obtained from the original study(Peric-Hupkes et al. 2010) and mm9 genomic coordinates were converted to mm10.

The RNA & DNA SPRITE (RD-SPRITE) data was obtained from NCBI GEO (accession GSE151515)(Quinodoz et al. 2021). To map the distribution of specific RNA molecules across the genome, the frequency of interactions between each RNA transcript and distinct genomic regions was quantified, as detailed previously. Briefly, we partitioned the genome into 100 kb bins and the raw interaction frequencies were determined by counting the instances in which an RNA transcript and a genomic bin co-occurred within SPRITE clusters. We focused on SPRITE clusters containing 2-1,000 reads for our analyses, as previously described. For the weighted bedgraphs, raw interaction counts were applied to a scaling factor based on the size of the corresponding SPRITE cluster. This normalization involved calculating all possible pairwise interactions within a cluster and assigning each interaction a weight of 2/n, where 'n' represents the total number of reads within that cluster.

Differential gene expression analysis

For cell culture study, single cell intron count matrices and metadata from mESC and NMuMG intron seqFISH+ experiments were loaded to Seurat (Stuart et al. 2019) (version 3.2.0). Intron counts of two biological replicates of mESC sample and one biological replicate of NMuMG sample were aggregated to generate normalized ensemble count data for mESC and NMuMG respectively. Intron differential expression analysis between mESC and NMuMG were performed using DESeq2 (Love et al. 2014) (version 1.26.0). Genes with the adjusted p-value < 0.01 and absolute value of log2 fold change > 2 were selected as significantly differential expressed genes.

The adult mouse cerebellum single-nucleus RNA-seq (snRNA-seq) data (Kozareva et al. 2021) was obtained from NCBI GEO (accession GSE165371). Single-cell cell-gene count matrices and sample meta tables were loaded to Seurat (version 3.2.0) for cell-type specific analysis. Gene counts of each cell cluster were aggregated for every replicate sample as ensemble count data for differential expression analysis using DESeq2. Differentially expressed genes were selected using the same criteria described above. Transcriptional activities across the mm10 reference genome were computed by summing averaged DEseq2 normalized counts of identified differentially expressed genes located within each 200 kb

genomic bins within cell types. The resultant average gene counts at 200 kb bins were used to calculate log2-fold expression change between pairs of cell types. The same analysis was performed by using intron seqFISH+ measurements with cell type information defined by mRNA seqFISH clusters.

For the cerebellum study, to compare the relationships between cell-type specific gene expression and immunofluorescence marker enrichment, differentially expressed 200 kb bins between pairs of cell types obtained above were selected, and the ensemble-averaged immunofluorescence marker differences between the same cell type pair were further calculated, similarly to our previous study (Takei, Zheng, et al. 2021). Box plots were then used to show the immunofluorescence marker differences between those differentially expressed 200 kb bins computed from intron seqFISH+ measurements.

The mouse primary visual cortex scRNA-seq data(Tasic et al. 2016) were obtained from NCBI GEO (accession GSE71585) with the cell-type annotation and processed TPM files and normalized as described before (Eng et al. 2019; Takei, Zheng, et al. 2021). The obtained cell-type specific gene expression levels were compared with the processed chromatin profiles (Zenodo doi: 10.5281/zenodo.4708111) from the adult mouse cerebral cortex (Takei, Zheng, et al. 2021).

Imaging-based transcriptomic data analysis

The similarity of ensemble-averaged mature or nascent transcriptome profiles by mRNA seqFISH+ or intron seqFISH+ in cell culture was compared to those by bulk RNA-seq in mESCs (Antebi et al. 2017) and NMuMG cells (Pascual-Reguant et al. 2018) or GRO-seq in mESCs (Jonkers et al. 2014) by using Spearman correlation, confirming high consistency of the datasets.

The preprocessing, clustering, and visualization of the imaging-based transcriptomic data (mRNA and intron seqFISH+) for the adult mouse brain cerebellum were performed with a Scanpy toolkit (Wolf et al. 2018) in Python. For mRNA seqFISH analysis, we chose genes

based on the max copy number of at least 20 in any cells, which yielded 49 genes. The gene count matrix with those genes was then used for the further clustering analysis. We then applied total count normalization and log(counts + 1) to transform the matrix, and performed the dimensional reduction with principal component analysis (PCA), followed by batch correction with Scanorama (Hie et al. 2019). We then constructed a nearest-neighbor graph with k = 40 neighbors using the top 30 principal components, followed by clustering of cells with Leiden clustering (Traag et al. 2019) and embedding using uniform manifold approximation and projection (UMAP) (Becht et al. 2018). Similarly, cells were clustered with Scanpy (Wolf et al. 2018) using intron seqFISH+ datasets. Briefly, we applied total count normalization, log(counts + 1), and scaling to transform the matrix, and performed the dimensional reduction with PCA and batch correction with Scanorama (Hie et al. 2019). We then performed construction of a nearest-neighbor graph with k = 25 neighbors using the top 30 principal components, Leiden clustering (Traag et al. 2019), and UMAP embedding (Becht et al. 2018). During these steps we filtered out cells based on the nuclear volume. Specifically, we filtered out cells with nuclear volume less than 100 μm^3 as a potential mis-segmentation. In addition, after Leiden clustering, we filtered out cells with nuclear volume outside the interquartile range in each cluster to minimize the doublets. In the end, we obtained n = 4,015 cells, consisting of n = 1,504, 832, 518, 357, 263, 164, 113, 88, 76, 56, 29, 15 cells from Leiden cluster 0 to 11, in two biological replicates of adult mouse brain cerebellum datasets (Supplementary Table 4).

The clustering results for individual cell types of the mouse brain cerebellum were compared to the normalized pseudo-bulk snRNA-seq expression profiles of each cell type in the adult mouse cerebellum, computed in the original study (Kozareva et al. 2021). The gene expression profiles of 49 genes after the filtering described above were used for the comparison. The degree of similarity for each cell type between imaging and sequencing datasets was evaluated by using the Pearson correlation and cell-type identity was annotated to each mRNA seqFISH cluster. Those clusters were annotated as Granule (cluster 0, 1), Bergmann glia (cluster 2), MLI1 (cluster 4), Purkinje cells (cluster 6, 11), MLI2/PLI (cluster 7), Endothelial (cluster 8), Astrocyte (cluster 9), and OPC/ODC (cluster 10). Clusters 3 and

5 were filtered out due to the mixture of cell types. Similarly, nascent transcriptional profiles of the four major cell types (Granule, Bergmann glia, MLI1, and Purkinje cells) in the adult mouse brain cerebellum by intron seqFISH+ were compared to RNA expression profiles of each cell type obtained from snRNA-seq datasets (Kozareva et al. 2021), using a subset of genes with detected average copy number of more than 0.01 per cell per each cell type by intron seqFISH+.

Single-cell global chromatin state analysis

Cells were clustered by averaged intensity profiles of individual immunofluorescence markers in each nucleus as performed before (Takei, Zheng, et al. 2021) by using a Scanpy toolkit (Wolf et al. 2018) in Python. Briefly, we used the mean intensity profiles of each immunofluorescence marker (n = 27 markers) within each nuclear label. Similarly to the intron seqFISH+ analysis, we then applied total intensity normalization, log(intensity + 1), and scaling to transform the matrix, and then performed the dimensional reduction with PCA and batch correction with Scanorama (Hie et al. 2019). Then we performed construction of a nearest-neighbor graph with k = 40 neighbors using the top 30 principal components, Leiden clustering (Traag et al. 2019), and UMAP embedding (Becht et al. 2018). The similarity of obtained clusters was compared to those from mRNA seqFISH by computing the overlapped fraction between a given pair of clusters.

Estimation of detection efficiency and false positive rates for two-layer DNA seqFISH+

The estimation of detection efficiency by two-layer DNA seqFISH+ for cycling mESCs as well as post-mitotic diploid cells in the female mouse brain was performed similarly to our previous studies (Takei et al. 2017; Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). Briefly, by considering the cell cycle distribution, we estimated detection efficiency as 21.1 ± 6.8% (median ± s.d.) from 63,466 ± 20,525 (median ± s.d.) DNA spots per cell for 100,049 loci in single male mESCs (n = 1,076 cells), while 5.0 ± 3.5% (median ± s.d.) from 9,912.0 ± 6,932.0 (median ± s.d.) counts per cell for 100,049 loci in post-mitotic Purkinje cells (n = 113 cells) in the female mouse brain cerebellum. We note that the lower detection efficiency of the brain samples could be caused by incomplete coverage of entire nuclei in z direction

for some cells, optical clouding of spots due to smaller nuclear sizes in tissues, as well as reduced hybridization efficiency due to the tissue thickness and different fixation conditions.

The estimation of false positive rates from two-layer DNA seqFISH+ was adapted from the previous DNA seqFISH+ scheme (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). To compute the false positive rates, both on-target barcodes (n = 100,049 barcodes) and blank barcodes, consisting of all the remaining error-correctable barcodes (n = 31,171 barcodes) in the codebook (Supplementary Table 2), were run simultaneously. The false positive rates were then computed by mean off-target barcode counts per barcode per cell divided by the sum of mean on- and off-target barcode counts per barcode per cell, which provided false rates of 1.6%, 1.8%, and 4.6% in mESCs (n = 1,076 cells), NMuMG cells (n = 384 cells), and cells in the adult mouse brain cerebellum (n = 4,015 cells).

Estimation of detection efficiency and false positive rates for mRNA and intron seqFISH+

The detection efficiency for mRNA seqFISH+ and intron seqFISH+ was estimated relative to the non-barcoded sequential single-molecule FISH (smFISH) detection efficiency, similarly to our previous studies (Shah et al. 2018; Eng et al. 2019). Briefly, the slope of seqFISH+ relative to smFISH was computed either directly (intron seqFISH+) or indirectly with RNA-seq (mRNA seqFISH+). From the slope values, we estimated the detection efficiency of mRNA seqFISH+ as 78.9% and 43.0% in mESCs and NMuMG cells and the detection efficiency of intron seqFISH+ after removing non-transcription active site spots as 29.1% in mESCs. For mRNA seqFISH+, 150 barcodes, which were identified as outliers during seed 3 and seed 4 barcode comparison (Extended Data Fig. 2) and included both on- and off-target barcodes, were filtered out from the analysis.

The estimation of false positive rates for mRNA seqFISH+ measurements was performed similarly to DNA seqFISH+ measurements (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) by using all on- and off-target barcodes from the codebook (Supplementary Table 1) simultaneously during the barcode decoding step. The false positive rate was computed with

on-target (n = 1,163) and off-target (n = 2,783) barcodes (n = 150 barcodes filtered out as described above) for mRNA seqFISH+. The false positive rates were computed by mean off-target barcode counts per barcode per cell divided by the sum of mean on- and off-target barcode counts per barcode per cell. With this calculation, the false positive rates for mRNA seqFISH+ were estimated as 7.9% and 10.2% in mESCs and NMuMG cells. We note that the different stringency of decoding results can be chosen by using different seed values (see 'seqFISH image processing'). In comparison with the decoding results of at least 3 out of 4 seeds above, more stringent results only from 4 seeds were estimated as 54.5% and 29.8% for detection efficiency and 3.2% and 4.8% for false positive rates in mESCs and NMuMG cells.

For intron seqFISH+, we implemented an additional filtering step of decoded spots by leveraging spatial locations of intron seqFISH+ and two-layer DNA seqFISH+ spots, given the fact that transcription active sites are at the chromosome territories (Shah et al. 2018). Specifically, we filtered out intron seqFISH+ spots that are more than 0.5 μm away from their chromosome territories computed from DNA seqFISH+ homologous chromosomes. To evaluate false positive rates from the filtered spots, we pre-assigned chromosome identity to each of the off-target barcodes (n = 2,880) with the same composition of chromosome identities as on-target barcodes (n = 17,856). Then the false positive rates for intron seqFISH+ were similarly calculated as mRNA seqFISH+ and estimated as 7.7% and 10.0% in mESCs and NMuMG cells. We used the filtered-out intron seqFISH+ data for downstream analysis in mESCs and NMuMG cells.

Pairwise spatial distance analysis for DNA loci

We calculated mean pairwise distances between genomic loci similarly to our previous studies(Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021) with modifications. When computing pairwise distances for one chromosome, we only considered distances between pairs of loci that we found to be on the same homologous chromosome. To calculate mean distances, we count co-detections and sum pairwise distances for each loci pair in separated chromosomes in each cell. After evaluating for all cells, we divide the sums of pairwise

distances by the number of co-detections. The pairwise spatial distance maps without binning (25-kb) or with 200-kb or 1.5-Mb (paint barcode) binning were used as specified in each figure.

The 1-Mb and 25-kb DNA seqFISH+ datasets (n = 2,460, 1,200 loci, respectively) from two biological replicates of mESCs (Takei, Yun, et al. 2021) were obtained from the 4D Nucleome data portal (https://data.4dnucleome.org/) under accession number 4DNESL2AY9CM. The median spatial distances were then compared to those from this study by using commonly profiled genomic loci. Characterization of cell-type specific inter-chromosomal association was conducted on the pairwise spatial distance matrix at the 1.5-Mb bin size (paint barcode). Spatial distances between top 5% genomic loci associated with each subnuclear marker or all loci annotated with specific gene families were selected in each cell type. In this analysis, spatial distances from pairs of intra-chromosomal loci were filtered out to only access spatial distances of pairs of inter-chromosomal loci. Random 5% of the genomic loci were selected by bootstrap 1,000 times and inter-chromosomal spatial distances from random selected genomic loci were extracted to compare with marker specific spatial distances in each cell type, similarly to previously described(Arrastia et al. 2021). Within cerebellum cell types, inter-chromosomal distances of 1.5-Mb paint blocks containing speckle peaks, broad RNAPIISer5-P peaks, transcriptionally active and inactive long genes were calculated respectively. For each category of paint blocks, expected inter-chromosomal distance was calculated as the median of 1,000 rounds of bootstrapping of randomly sampled paint blocks. Inter-chromosomal distances within each category minus the expected inter-chromosomal distance was used as the normalized inter-chromosomal distances.

Normalization for pairwise spatial distance analysis

The two-layer DNA seqFISH+ spatial distance maps typically contain boundaries between loci encoded in different chromosome blocks in different fluorescent channels (Extended Data Fig. 3l-t, Supplementary Notes). We note that these boundaries do not affect all the analysis performed in the paper, such as A/B compartment analysis, spatial chromatin

profiling, and spatial distance analysis (>= 1 Mb bin sizes), which were further validated by orthogonal datasets in mESCs and NMuMG cells.

To remove these block patterns from two-layer DNA seqFISH+ spatial distance maps at 25 kb bins, we corrected the ensemble-averaged spatial distance distribution shift due to the block barcoding scheme as discussed in Supplementary Notes. According to the two-layer barcoding scheme, each genomic locus was assigned two coordinates, genomic coordinate (1D genomic location) and block coordinate (block identity). Thus, for a given genomic locus pair i,j, there are three distances: genomic distance G_{ij}, block distance B_{ij} and spatial distance. The pairwise distances of loci pairs with identical genomic and block distances at single allele level followed a gamma distribution. Consider all loci pairs i and j with a genomic distance G_{ij}=n, for such pairs, two block distances were possible: B_{ij} = b1 or b2. Two gamma distributions were fitted to the distribution of pairwise distances at the single-cell level, represented as $P(i,j|G_{ij} =n, B_{ij} =b_1)$ and $P(i,j|G_{ij} =n, B_{ij} =b_2)$. These two distributions were shifted to the generalized distance distribution $P(i,j|G_{ij} =n)$ by percentile mapping using probability density functions. The median spatial distance map between pairs of intra-chromosomal loci using shifted distances were calculated as the normalized distance maps.

Radial organization analysis for DNA loci and subnuclear foci within the nucleus

The radial organization of DNA loci within the nucleus was evaluated by computing the convex hull surface of the DNA loci similarly to previous studies (Liu et al. 2020; Su et al. 2020; Payne et al. 2021). To compute the radial positioning of DNA loci within the nucleus, we constructed a 3D convex hull for each nucleus using the DNA seqFISH+ spots per cell using the SciPy spatial library in Python. At this step, cells with less than 100 DNA seqFISH+ spots were filtered out. We then calculated the spatial distance of individual spots from the nuclear periphery by calculating the distance of intersection from the convex hull surface to the centroid of the nucleus through the individual spots. The radial scores of genomic loci were similarly computed by scaling the spatial distance from the nuclear center to nuclear periphery as 0 to 1. The computed median distance profiles from the nuclear periphery in each cell type were compared by Pearson or Spearman correlation with Lamin B1 chromatin

profiles of the corresponding cell type, which are independently measured and typically enriched at the nuclear periphery, with 100-kb binning for the cell culture datasets and with 200-kb binning for the adult mouse brain cerebellum datasets. Similarly, the radial organization of subnuclear foci within the nucleus was computed using the centroid voxel location of each of the foci.

The similarity of radial positioning of genomic loci among cell types in the adult mouse brain cerebellum was computed by Pearson correlation of median radial scores across cell types with 200-kb binned loci. The top 5% variable loci in radial positioning between neurons and glial cells were computed by using the radial profiles of neuronal cells (Purkinje cells, MLI1, MLI2/PLI) and glial cells (Bergmann glia, Astrocytes) as input. Out of 629 variable loci, we identified 409 loci moved interior of the nucleus in neurons versus 220 loci in glial cells. We then compared the degree of overlap between those variable loci and Purkinje H4K20me3-associated loci or mCH desert loci (Lister et al. 2013).

Imaging-based chromatin profiling analysis

We used z-score normalization of each chromatin marker (antibodies and ncRNAs) in each 3D nuclear label to compute voxel resolution chromatin profiling, similarly to those previously described (Takei, Yun, et al. 2021; Takei, Zheng, et al. 2021). Briefly, the intensity values for each marker obtained from aligned sequential immunofluorescence or RNA FISH or DNA FISH images were normalized by computing intensity z-scores at individual voxels per each nuclear label and then the z-scored intensity profiles for each marker were computed at rounded voxel locations where the final decoded DNA spots were identified by DNA seqFISH+. This approach allows us to investigate matrices consisting of DNA loci by subnuclear markers in individual cells. Specifically, each single cell was presented as DNA loci to subnuclear marker matrix, where each detected 25-kb DNA locus is represented as a vector with features including the normalized z-scores of subnuclear marker intensities, as well as radial positioning inside the nucleus.

Then DNA loci and their associated features were binned into 100-kb, 200-kb, or 1.5-Mb (paint barcode). Single cells were then grouped according to cell type to calculate the ensemble-averaged level of subnuclear marker scores per each cell type. For cell culture study, median scores of each feature associated with DNA bins from mESCs or NMuMG cells were calculated to get the ensemble-averaged marker intensity z-score at 25-kb, 100-kb, 200-kb or 1.5-Mb bins. For the cerebellum data, single cells assigned to different cell types were grouped and median scores of each feature were calculated per each cell type at 200-kb and 1.5-Mb bins. For both cell culture and cerebellum studies, the rare case of DNA bins which were filtered out during radial score calculation were also filtered out in this analysis.

The fraction of loci chromatin profiles were also calculated at 25 kb, 100 kb, and 200 kb bin sizes. For each cell type and genomic bin, the fraction of loci with the z-score higher than 1.5 for each immunofluorescence marker or spatial distance from the nuclear periphery smaller than 500 nm was calculated as a fraction score. Both median and fraction of loci ensemble profiles were used to assess the practical resolution for the cell culture study. We note that without special indication, ensemble chromatin profiles refer to median z-score profile in this paper.

We also performed imaging-based chromatin profiling by calculating the closest spatial distances between segmented subnuclear foci (see ‘Subnuclear foci analysis’) and DNA loci. This analysis showed a high degree of similarity between z-score-based SF3A66 chromatin profiles and physical distances from SF3A66 foci across cell types in the adult mouse cerebellum. We also confirmed that our DNA seqFISH+ protocol preserved the subnuclear structures in the tissue sections by showing the high degree of similarity of DNA locus organization around nuclear speckle markers (Rnu2 by RNA FISH and SF3A66 by immunofluorescence), which were imaged before and after DNA seqFISH+ preparation, respectively. In addition, as the nucleolar marker Fibrillarin had a low signal-to-noise ratio in tissue experiments, we used ITS1 RNA, another nucleolar marker, for the chromatin profiling and visualization of the nucleolus in tissue experiments. We note this low signal-

to-noise ratio was not observed in cell culture experiments.

To validate the imaging-based chromatin profiles in this study, we compared those with 1-Mb resolution imaging-based chromatin profiles by DNA seqFISH+ obtained from the original study (Takei, Yun, et al. 2021) and other sequencing-based datasets such as CUT&Tag, CUT&RUN, and pA-DamID (van Schaik et al. 2022). The comparison of chromatin profiles between DNA seqFISH+ (n = 2,460 loci) and two-layer DNA seqFISH+ was performed at commonly profiled 25-kb genomic loci that are on average ~1 Mb apart, while those between two-layer DNA seqFISH+ and sequencing datasets were with 100 kb binning. The degree of similarity between datasets was evaluated by Pearson or Spearman correlation. In addition, to further validate the genomic loci to the nuclear lamina association, we compared the enrichments of imaging-based Lamin B1 profiles from mESCs across previously identified lamina-associated domains (Peric-Hupkes et al. 2010; Meuleman et al. 2013). See 'Sequencing-based data processing' for the details of sequencing-based data source and processing.

To evaluate the practical resolution of imaging-based chromatin profiles in this study, we used both ensemble fraction and median z-score profiles in mESCs. For the nuclear speckle associations, the SF3A66 imaging profiles were compared with the speckle marker RD-SPRITE (Quinodoz et al. 2021) profiles with 100 kb bins. Local maxima in the imaging-based SF3A66 fraction and median z-score profiles were detected using the scipy.find_peaks() function and were compared to the local maxima in the RD-SPRITE Rn7sk profiles. The distance from each detected local maxima in the imaging track to the closest local maxima in the RD-SPRITE Rn7sk track was calculated to evaluate the similarity between the imaging and sequencing profiles. The peak similarity within the sequencing method was calculated using the above method for RD-SPRITE Rn7sk and Malat1 profiles as a positive control. The closest distance of randomized peaks in both fraction or median imaging profiles to the detected peaks in RD-SPRITE Rn7sk profiles were calculated as negative controls.

For three chromatin markers (RNAPIISer5-P, H3K4me3, and H3K27me3), we compared the imaging-based fraction and median z-score immunofluorescence profiles with the CUT&Tag profiles in mESCs generated in this study. Peaks were identified using scipy.find_peaks() function in the 25-kb binned imaging profiles for these chromatin markers. Due to the higher noise level in the 25-kb binned chromatin profiles, we assessed the reliability of the detected peaks using a downsampling bootstrap analysis. In this procedure, 80% of mESCs were randomly selected, and both fraction and median z-score profiles were calculated from the selected 80% of cells, as previously described. This downsampling was repeated 500 times, and peaks were detected in each iteration's fraction and median z-score profiles. The percentage of times each 25-kb bin was identified as a peak was then calculated. Bins detected as peaks in at least 80% of the bootstrap iterations were considered true peaks for median z-score and 50% threshold was used for fraction profile, and their peak widths were calculated using the function scipy.signal.peak_widths(). These confident peak widths typically ranged from 50-kb to 150-kb sizes across markers. Chromosome X was excluded from this analysis due to its high noise. Furthermore, the CUT&Tag profiles were smoothed using various window sizes of 2–48 (corresponding to 50–1200 kb), and the Spearman correlation between the smoothed sequencing and the imaging profiles was computed for each chromosome. The elbow changing point of the correlation curve was decided using the KneeLocator function in the python package Kneed.

Lamina associated domains (LADs) were detected using scipy.signal.find_peaks() function in both fraction and median z-score chromatin profiles in mESCs (25 kb and 100 kb bins) and NMuMG cells (100 kb bins) for Lamin B1 marker as well as for spatial distance from the nuclear periphery. In both mESCs and NMuMG cells, Spearman correlation of fractional distance from the nuclear periphery profiles with pA-DamID profile(van Schaik et al. 2022) were calculated with 100 kb bins. For spatial distance profiles to the nuclear periphery distance in mESCs with 25 kb bins, the Spearman correlation of smoothed pA-DamID 25-kb bin profiles from bin window sizes of 2–48 (corresponding to 50–1200 kb) were calculated in each chromosome, and the elbow changing point of the correlation curve was detected using python package Kneed as described above. LAD peaks and peak width were

calculated with 100-kb binned Lamin B1 and spatial distance to nuclear periphery profiles in mESCs as described in "Peak definition of chromatin profiles". For each detected LAD in imaging-based chromatin profiles, the percentage of bins overlapped with DamID (Peric-Hupkes et al. 2010) LADs in mESCs was calculated.

For nucleolar associated domains (NADs), the fraction and median z-score imaging-based chromatin profiles of Fibrillarin marker were compared to the nucleolar ITS1 enrichment profiles from RD-SPRITE (Quinodoz et al. 2021) as well as the enrichment scores from NAD-seq (Bizhanova et al. 2020) in mESCs. Pearson correlation coefficients were calculated between these profiles across the genome as well as within individual chromosomes.

Downsampling of mESCs was performed to assess the robustness of the fraction and median z-score chromatin profiles. Measured mESCs were randomly downsampled to 25, 50, 100, 250, and 500 cells, with 30 repetitions for each group. The 100-kb binned fraction and median z-score chromatin profiles were calculated for the chromatin markers SF3A66, RNAPIISer5-P, Lamin B1, H3K27me3, H3K9me3, and Fibrillarin. For each iteration of downsampling, the Pearson correlation between the downsampled chromatin profile and the original ensemble chromatin profile from all 1,076 cells measured was calculated for each chromatin marker. The robustness of the chromatin profiles was evaluated based on the changes of Pearson correlation coefficients as a function of the number of cells.

We note that the dynamic range of fluorescence intensities for a given marker can influence the interpretation of the marker enriched DNA loci. For example, H3K9me3 staining intensities are predominantly enriched in the chromocenter in mESCs (Extended Data Fig. 4g), and therefore the marker enrichment on DNA loci reflects their spatial proximity to the chromocenter (Fig. 1g, Extended Data Fig. 4j). Conversely, for markers with a more diffuse nuclear localization (e.g., H3K9ac, RNAPIISer5-P), the marker enrichment on DNA loci may indicate the relative level of enrichment across genomic regions. In addition, we note that local enrichment or depletion of a marker on a DNA locus could be masked out if the DNA locus is spatially close to bright foci of the marker (e.g., H3K9me3 staining at the

chromocenter) in single cells. Furthermore, we note that in our imaging-based chromatin profiling, the fluorescence intensity z-scores were determined from total voxels per marker in each nucleus and the baseline of the values could be different from the other orthogonal datasets, depending on their normalization methods.

In addition, while the fraction of loci processing improved the similarity with orthogonal sequencing-based datasets for certain markers (e.g. Lamin B1, Fibrillarin), there are several advantages to using the median z-score or average spatial distances. First, the median z-score chromatin profiling is more robust when fewer spots are detected (Extended Data Fig. 10f, g). For example, in some cerebellar cell types, chromatin profiles are more robustly reproduced using the median z-score based on the number of spots detected per defined bin. Second, the median z-score and average spatial distance calculations capture information even when loci are far from subnuclear markers. For certain markers (e.g., Lamin B1 or the nuclear periphery), these profiles can provide qualitatively different information. For example, in the adult mouse brain cerebellum datasets, using average spatial distances from the nuclear periphery enables us to identify DNA loci closer to the nuclear interior. The fraction of loci data cannot distinguish loci that are more toward the nuclear interior if they do not contact with the nuclear periphery. The average positioning has also been used previously to identify 3D genome reorganization during postnatal development in the mouse brain (Tan et al. 2021). Given these factors, as well as the overall consistency in top enriched loci regardless of the processing methods, we used the z-score-based chromatin profiling for the adult mouse brain cerebellum.

The similarities of genome-wide chromatin profiles between cell types in the adult mouse cerebellum were evaluated by computing the Pearson correlation coefficient of each marker between given pairs of cell types with 200 kb binning. In addition, the similarities of chromatin profiles between pairs of markers in each cell type were evaluated by comparing the overlap of top 5% genomic loci associated with each marker in each cell type.

To compare differential association between two markers in each cell type, the top 5% differentially associated bins were selected. The ensemble-averaged z-score intensity of two markers were plotted along the x- and y- axis, respectively. We then moved a slope = 1 line along the x or y axis until 5% of the bins above or lower this line were identified. Similarly, SF3A66 z-score change and RNAPIISer5-P z-score change between two cell types were plotted along the x- and y-axis, respectively. A slope = 4 line was moved along the SF3A66 axis until 5% of the bins were on the left or right side of this line to define the top 5% SF3A66 differentially associated bins. A slope = 0 line was moved along the RNAPIISer5-P axis until 5% of the bins were above or below the line to define the 5% RNAPIISer5-P differentially associated bins.

We also computed compartment scores (Nichols & Corces 2021) by using selected immunofluorescence markers including H3K27ac, H3K27me3, H3K9me3, and Lamin B1 in mESCs and NMuMG cells. For each chromosome, correlation matrices of mESC and NMuMG 100 kb ensemble-averaged immunofluorescence z-scores for the above 4 markers were computed. The first eigenvectors of the above correlation matrices were calculated as immunofluorescence A/B compartment scores. The sign of the compartment scores was checked for each chromosome and multiplied by -1 to flip the signs if necessary to correct the direction of the eigenvector of the entire chromosome. The similarity of Hi-C and imaging datasets were then compared by computing Spearman correlation using 100-kb bins.

To compare the enrichment score of immunofluorescence markers between mouse cell lines, the enrichment scores for each dataset were first normalized by percentile to avoid the difference across different methods. Differential scores were calculated by subtraction of percentile scores between mESCs and NMuMG cells within sequencing-based or imaging-based data, and then Pearson's correlation was calculated to examine the similarity of differential association between sequencing-based and imaging-based methods. The changes of A/B compartment scores between mouse cell lines obtained from Hi-C or two-layer DNA seqFISH+ were similarly computed and Pearson’s correlation of differential A/B

compartment scores between Hi-C and two-layer DNA seqFISH+ was calculated to compare the similarity of the biological changes captured by those orthogonal methods.

Peak definition of chromatin profiles

To characterize genome fragments associated with various subnuclear markers, peak calling was performed on ensemble-averaged 200-kb binned z-score tracks for both cell culture and cerebellum studies. One dimension ensemble-averaged z-scores of each subnuclear marker were mean-centered within chromosomes first, then scipy.signal.find_peaks function was performed with manually decided "prominence" and "height threshold" parameters to call out the peaks. Overlapping peaks were merged to one broader peak. For the nuclear periphery distance, negative mean centered distance was used due to its negative correlation with lamina-associated domains. Genomic bins characterized as Purkinje H4K20me3-associated peaks (see 'Characterization of H4K20me3-associated genomic loci') were compared to the peak definition in other cell types for H4K20me3 and nuclear periphery. Then the transition of H4K20me3 bins into other type of repressive bins in other cell types were visualized using Sankey plot using pySankey package.

The LAD calling of imaging-based Lamin B1 profiles was similarly performed using 100 kb bins with slightly optimized parameters for mESCs and NMuMG cells. Specifically, for the ensemble-averaged Lamin B1 z-scores, the baseline was first elevated by 0.3, and peaks were called using python scipy.signal.find_peaks package, with parameters as follows: prominence of 0.1, height threshold of 0.1, relative height to be 0.4. For ensemble-averaged spatial distances from the nuclear periphery, the ensemble track was mean centered within the chromosome, flipped and baseline elevated by 0.3. Peaks were called using python scipy.signal.find_peaks package, with parameters as follows: prominence of 0.1, height threshold of 0.1, relative height to be 0.35.

To classify active subnuclear organization, RNAPIISer5-P peaks spanning more than 2 Mb were defined to be broad peaks and rest peaks were defined to be sharp peaks. SF3A66 peaks overlapped with RNAPIISer5-P peaks were assigned to be speckle peaks. Similarities of

speckle peaks, broad and sharp RNAPIISer5-P peaks across cerebellum cell types were calculated by defined overlapping score: the sum of the fractions of the number of overlapping peaks in all peaks in both cell types and then divided by 2. To compare to the active subnuclear markers, we computed repressive peaks using peaks at least in one of the following repressive markers (nuclear periphery, H4K20me3, or H3K27me3). We note that we arbitrarily chose 2 Mb as a size of cut-off of the RNPIISer5-P broad and sharp domain. The differences of the GC content between broad and sharp domains were well conserved regardless of the cut-off values of 1 Mb, 2 Mb, and 4 Mb.

Characterization of cell-type-specific long genes

To characterize long gene specific association with subnuclear markers, genes with > 200 kb were defined as long genes and DNA bins containing long genes were selected for further analysis. Pseudo-bulk normalized mRNA counts of different cell types in the adult mouse brain cerebellum provided by snRNA-seq study(Kozareva et al. 2021) were used to represent mRNA expression level. For each long gene, the maximum RNAPIISer5-P normalized z-scores among the gene spanned 200kb DNA bins were calculated as long gene RNAPIISer5-P level. Pearson correlation between normalized mRNA counts and RNAPIISer5-P z-scores of each long gene across cell types was calculated. Among 867 defined long genes, 132 genes were identified with a Pearson correlation coefficient higher than 0.8. To further evaluate other subnuclear marker profiles in long gene regions, jaccard index between peaks of other subnuclear markers and RNAPIISer5-P peaks in long gene DNA bins were calculated using scipy.spatial.distance. Specific immunofluorescence markers (SOX2, HDAC1, ATRX, Fibrillarin, RNAPIISer2-P) were filtered out from this analysis due to the cell-type specific staining or low signal-to-noise ratio.

To compare the relationship between RNAPIISer5-P enrichment and open chromatin region at the long gene loci, the processed file of genome length for the open chromatin regions by ATAC-seq in Purkinje cells was obtained from the original study(Kwak et al. 2021). Then the Purkinje RNAPIISer5-P enrichment at the cell-type specific RNAPIISer5-P long gene loci (Pearson's $r > 0.8$) with or without a highly open conformation containing >20,000-bp

peaks (Kwak et al. 2021) (n = 17, 263 genes, respectively) were compared by two-sided Wilcoxon's signed rank-sum test.

The relative spatial distances of the cell-type specific RNAPIISer5-P long gene loci (n = 132) and neighboring loci from the nuclear periphery were calculated by comparing their ensemble-averaged spatial distances from the nuclear periphery in each cell type. Specifically, 1 Mb genomic spans centered by each of the long gene loci were decided as the center regions first. Then flanking regions were defined as 1 Mb up-stream and down-stream of the center regions. Relative spatial distance from the nuclear periphery was calculated by using the mean distance to the nuclear periphery of the flanking regions minus the corresponding mean distance to the nuclear periphery of the gene center regions. The computed relative spatial distance from the nuclear periphery versus scaled snRNA expression (Kozareva et al. 2021) were compared for the long genes in each cell type in the adult mouse cerebellum. The long genes with scaled expression larger than 1 were considered as transcriptionally active in each cell type, and the relative spatial distance from the nuclear periphery between transcriptionally active and inactive genes were compared by two-sided Wilcoxon's signed rank-sum test. In addition, radial positioning of transcriptionally active and inactive long gene loci were also compared together with that of speckle peaks and broad RNAPIISer5-P peaks.

To identify Purkinje-specific H3K27me3-associated long genes, the cell-type-specific H3K27me3-associated loci were first identified by comparing H3K27me3 profiles between pairs of cell types in the adult mouse brain cerebellum (see 'Imaging-based chromatin profiling analysis'). Then Purkinje-specific H3K27me3-associated long genes (n = 116 genes) were obtained by finding long genes (>200 kb) annotated in the Purkinje-specific H3K27me3-associated loci relative to MLI1, MLI2/PLI, or Bergmann glia cells. The developmental gene expression profiles of those long genes were further examined by using a differential gene expression analysis with a provided processed data (Stoyanova et al. 2021), identifying 36 out of 46 developmental differentially expressed genes, which is a

subset of Purkinje-specific H3K27me3-associated long genes, are down-regulated from P0 to adult Purkinje cells.

Characterization of H4K20me3-associated genomic loci

To examine H4K20me3-associated genomic locus, 200-kb DNA bins identified as H4K20me3-associated peaks that are not assigned as MajSat peaks in the analysis under 'Peak definition of chromatin profiles' were further processed. H4K20me3-associated DNA bins with a normalized z-score >= 1 in Purkinje cells were defined to be H4K20me3-associated bins, others are defined to be weak H4K20me3-associated bins. To compare the other repressive marker enrichments at the H4K20me3-associated loci in Purkinje cells, quantile-normalized immunofluorescence marker z-scores were compared.

To compare the gene expression levels of these H4K20me3-associated genomic loci, the genes with the whole gene body inside the H4K20me3-associated genomic loci were selected. The expression levels of those genes measured by snRNA-seq (Kozareva et al. 2021) were shown as box plots along with those inside the RNAPIISer5-P and speckle peaks for comparison in each cell type.

To characterize gene families and functions associated with the H4K20me3 marker, gene prefixes were extracted by cropping gene name symbols before any digit or symbol character. Recurrent gene prefixes were considered to be gene family, and then enrichment of gene families in H4K20me3-associated locus were computed at 200kb bins. Number of H4K20me3-associated bins in Purkinje cells that overlap with DNA bins containing SSDR genes (Lilue et al. 2018; Lilue et al. 2019) or DNA bins previously identified as mCH desert (Lister et al. 2013) were further counted by converting mm9 genomic coordinates provided by the original studies to mm10 using the UCSC Genome Browser program liftover.

Individual H4K20me3 profiles in Purkinje cells were further investigated to characterize the chromatin organization of individual alleles at strong and weak H4K20me3-associated DNA bins. H4K20me3-associated peaks enriched with *Vmn* and *Olfr* gene families were selected

to represent strong and weak H4K20me3-associate genomic locus respectively. The genomic region spanning up to 10 Mb upstream or downstream of the beginning or end *Vmn* or *Olfr* H4K20me3-associated peaks were selected to separate homologous chromosomes. H4K20me3 and Lamin B1 immunofluorescence intensity z-scores in those regions from all alleles detected in Purkinje cells were collected, then visualized as heatmaps.

The individual spatial clusters formed by these H4K20me3-associated loci were profiled directly in single cells. The H4K20me3-associated loci were extracted in each cell, and then each spatial cluster was defined as a group of H4K20me3-associated loci with pairwise spatial distance smaller than 0.45 μm. Specifically, a spatial pairwise distance matrix of H4K20me3-associated loci was calculated for each single cell. Each of the DNA loci was considered as a node, and nodes with spatial distance smaller than 0.45um were considered as spatially connected, forming a graph in each single cell. Each connected component in this graph was called using the function "networkx.connected_components", and was considered as a cluster. Clusters with less than 10 DNA loci were filtered out to reduce noise. The thresholding of 0.45 μm was determined since 0.45 μm is the peak of intra-chromosomal distance across all cell types, filtering out clusters with less than 10 loci was equivalent to maintaining the top 10% of clusters across cell types. To examine the inter-chromosomal association within each spatial cluster, the entropy of chromosomal distribution was calculated as an indicator of inter-chromosomal mixture score using the "scipy.stats.entropy" function with base = 2. Note that if a cluster contains only one chromosome, the entropy is 0, and as the entropy increases, the cluster becomes more mixed with different chromosomes. The co-occurrence frequency of different chromosomes was calculated within different cell types. We also calculated the number and the composition of chromosomes for each cluster in each cell type. Specifically, the percentage of clusters formed by single chromosomes, two chromosomes, and three or more chromosomes among all clusters were calculated in each cell type. Then top 5 triplet clusters that occurred most frequently in each cell type were further visualized to show cell-type specific higher-order genome organization with the H4K20me3-associated loci.

Gene ontology analysis

Gene ontology (GO) analysis was performed using ClusterProfiler (3.18.1) (Yu et al. 2012) under R (4.0.3) platform. Specific pairs of inputs below were formed into a gene list with corresponding annotations, and comparison between different annotations were conducted using "compareCluster" function using default settings, then gene ontology terms were further combined using "simplify" function at 0.7 cutoff. Visualizations of pathway enrichment were plotted by calling the "dotplot" function built in ClusterProfiler.

To compare the pathway enrichment between mESC and NMuMG cells, GO analysis was performed on the differentially expressed genes identified by intron seqFISH+. To compare the speckle and RNAPII associated gene pathways, genes overlapping with top 5% of SF3A66 or RNAPIISer5-P genomic loci in each major cell type in the adult mouse cerebellum were selected to compare the pathway enrichment. To obtain pathway enrichment associated with H3K27me3, genes overlapping with top 5% genomic loci associated with H3K27me3 or Lamin B1 in each major cell type in the adult mouse cerebellum were used for GO comparison. Similarly, genes located in Purkinje H4K20me3 or H4K20me3 weak genomic loci were extracted and compared for the GO enrichment. 200 kb binned genomic loci were used in this analysis for the tissue experiments.

Gene category annotations

The list of housekeeping genes that are conserved between mouse and human was obtained from the Housekeeping and Reference Transcript Atlas database (Hounkpe et al. 2021). The list of synaptic genes in human was obtained from the SynGO database (Koopmans et al. 2019) and converted to mouse orthologous genes by using Ensembl BioMarts (Kinsella et al. 2011). Similarly, the list of autism spectrum disorder (ASD) candidate genes in human was obtained from the SFARI Gene database(Abrahams et al. 2013) (gene score 1, 2) and converted to mouse orthologous genes.

Validation tissue slice experiment and analysis

To validate the robustness of the two-layer DNA seqFISH+ results, the adult brain cerebellum tissue slice experiments were repeated with less multiplexed DNA seqFISH.

Specifically, we designed an increased number of 100 primary probes per 25 kb locus for a higher sensitivity for selected 22 genomic regions (Supplementary Table 5) using mm10 newBalance DNA FISH probes at PaintSHOP resources (Hershberg et al. 2021) and then those sequences were attached with four identical readout probe binding sequences shared within each of the 25 kb loci to sequentially image each locus at a time. Those primary probes were generated from an oligo array (Twist Bioscience), as described above (see 'Primary probe design and synthesis'). The validation tissue slice experiments were performed similarly to the protocol above (see 'Tissue slice experiment') using a different set of primary probes as described above as well as a subset of primary antibodies (Supplementary Table 4) by omitting RNA seqFISH steps. The images were taken under the same condition as the two-layer DNA seqFISH+ datasets.

The image processing was performed similarly to those described above (see 'Nuclear and cytoplasmic segmentation' and 'seqFISH image processing) except that DAPI and Lamin B1 images were used as 3D nuclear segmentation inputs for the Cellpose (Stringer et al. 2021). In addition, to distinguish Purkinje cells and Bergmann glia in the Purkinje cell layer, the Purkinje cell layers were manually segmented as a ROI in each image using ImageJ and a subset of nuclear labels whose xy centroids are within the ROI were considered for further processing. The cell type identities in this layer were then determined by comparing obtained mean normalized HDAC1 intensities per replicate and nuclear volumes.

The detection efficiency of the validation experiments was estimated as 60.0% for 22 genomic loci (n = 128 and 518 cells for Purkinje cells and Bergmann glia from two biological replicates), which is an order of magnitude higher compared to the ~5% efficiency in the two-layer DNA seqFISH+ tissue experiments (see 'Estimation of detection efficiency and false positive rates for two-layer DNA seqFISH+').

The imaging-based chromatin profiles were generated by using the z-score normalization (see 'Imaging-based chromatin profiling analysis'). The degree of similarity between the validation experiment datasets and two-layer DNA seqFISH+ datasets (200 kb bins) were

compared by computing Pearson correlation coefficients. The distributions of z-score values for each marker for each locus were shown as boxplots for each cell type. In addition, the distributions of spatial distances for pairs of genomic loci, including control locus pairs with similar genomic distances, were shown for both inter-chromosomal and intra-chromosomal associations. We note that computed intra-chromosomal spatial distances contain those between homologous chromosomes as we did not separate homologous chromosomes within a cell.

Statistics and reproducibility

Cells shown in Fig. 1 and Extended Data Figs. 2-7 are representatives of 1,076 cells from two biological replicates of mESCs or 384 cells from one biological replicate of NMuMG cells. Cells shown in Figs. 2-6 and Extended Data Figs. 6-16 are representative of 4,015 cells from two biological replicates of the adult mouse brain cerebellum.

2.9 Appendix

Analysis and Normalization of Technical Artifacts in Two-layer DNA seqFISH+

In this appendix, we will discuss the potential mechanism and normalization of the technical noise by two-layer DNA seqFISH+, which is observed as block patterns in median spatial distance maps between pairs of intra-chromosomal loci. At the same time, we show that the technical noise does not affect the chromatin profiling analysis that we performed.

In **Section 1**, we show that the block pattern results from the down-sampling of genomic loci near block boundaries, rather than distortions or aberrations in the data. In **Section 2**, we show that the chromatin profiles are not affected by the technical noise using simulated down-sampling (**Section 2.1**), and two-layer DNA seqFISH+ datasets (**Section 2.2**). Finally, we introduce a normalization method to remove the technical noise in two-layer DNA seqFISH+ median distance maps (**Section 3**). We then discuss the limitations of the proposed normalization method (**Limitations**), and discuss potential solutions to recover or

avoid down-sampled genomic loci in the future when using two-layer barcoded DNA seqFISH+ in section **Discussions**.

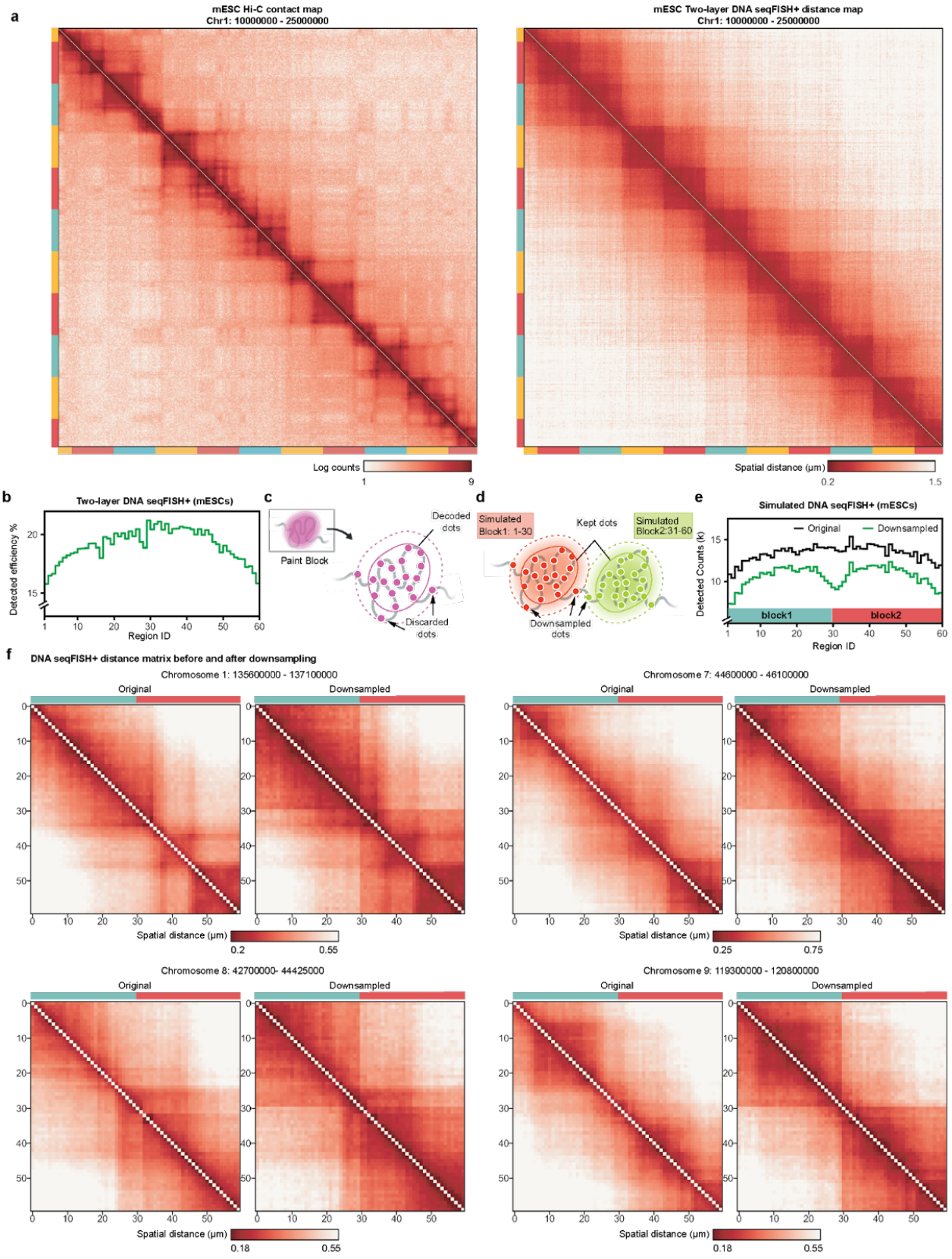
Down-sampling of genomic loci reproduces the technical noise in spatial distance maps
The two-layer DNA seqFISH+ intra-chromosomal pairwise distance maps showed a block pattern of squares, matching the 60-loci (1.5-Mb) chromosome blocks used for the two-layer barcoding (Appendix Fig. 1a). Here, we provide multiple lines of evidence that these block patterns result from the down-sampling of genomic loci near block boundaries, rather than from distortions or aberrations in the data.

Consistent with down-sampling, the genomic loci near the beginning and end of each 60-loci block have lower detection rates on average compared to those in the center of the blocks (Appendix Fig. 1b). In the two-layer barcoding scheme, the identity of each 60 contiguous 25-kb DNA loci were decoded with the fluorescence intensity of the chromosomal block paint. In this barcoding scheme, it is possible that the genomic loci in higher paint intensity regions are more likely to be decoded than the loci in lower paint intensity regions in 3D spaces in the images. This could potentially lead to lower detection efficiency of genomic loci near the beginning and end of each chromosomal block, which are on-average more likely to be at the periphery of each chromosomal paint in the images (Appendix Fig. 1c). As a consequence, the down-sampling of DNA bins near the block boundary in 3D space led to enrichment of apparent increase in interactions between loci in the center of the paint blocks in the pairwise interaction maps.

To confirm this mechanism of block pattern formation, we used the previously published 25-kb DNA seqFISH+ data(Takei et al. 2021) as ground truth for a down-sampling simulation. The original DNA seqFISH+ dataset profiled 60 continuous 25-kb genomic loci for each target genomic region in mESCs. We divided those 60 genomic loci into two groups consisting of two contiguous 30-loci blocks and calculated the spatial distance maps of pairs of genomic loci. In the down-sampling simulation, we dropped 20% of detected DNA spots furthest from the centroids of each of the computed blocks to simulate the lowered detection

efficiency at the edge of paint signals in the two-layer barcoding scheme (Appendix Fig. 1d). The down-sampled data recapitulated the technical noise features that were observed in the two-layer DNA seqFISH+ data. First, we observed a decrease in the number of detected DNA bins at the start and end segments of artificial 30-loci blocks after the down-sampling DNA spots based on the spatial location relative to the chromosomal paint center (Appendix Fig. 1e), similarly to the two-layer DNA seqFISH+ datasets. Second, we found that the spatial distance maps calculated from the down-sampled data show a block pattern in accordance with the genomic location of the simulated down-sampling (Appendix Fig. 1f).

Future works will be focused on ameliorating the down-sampling effect by developing deep-learning image processing tools to allow recovery of genomic loci near boundaries with lower paint intensities.


a
mESC Hi-C contact map
Chr1: 10000000 - 25000000
mESC Two-layer DNA seqFISH+ distance map
Chr1: 10000000 - 25000000
Log counts
1
9
Spatial distance (μm)
0.2
1.5
b
Two-layer DNA seqFISH+ (mESCs)
Detected efficiency %
Region ID
c
Paint Block
Decoded dots
Discarded dots
d
Simulated Block1: 1-30
Kept dots
Simulated Block2:31-60
Downsampled dots
e
Simulated DNA seqFISH+ (mESCs)
Original
Downsampled
Detected Counts (k)
block1
block2
Region ID
f
DNA seqFISH+ distance matrix before and after downsampling
Chromosome 1: 135600000 - 137100000
Chromosome 7: 44600000 - 46100000
Chromosome 8: 42700000- 44425000
Chromosome 9: 119300000 - 120800000
Original
Downsampled
Spatial distance (μm)
0.2
0.55
0.25
0.75
0.18
0.55
0.18
0.55

Appendix Fig. 1| The down-sampling simulation of ground truth data can recapitulate technical noise features observed in spatial distance maps by two-layer DNA seqFISH+.

a, Representative Hi-C map (Bonev et al. 2017) (left) and median spatial distance map between pairs of intra-chromosomal loci by the two-layer DNA seqFISH+ in mESCs (right). The left and bottom lines with green and red colors mark block boundaries used for the two-layer DNA seqFISH+ barcoding. **b**, Average detection efficiency of 25-kb genomic loci with locus identities from 1-60 within 1.5-Mb blocks used in the two-layer DNA seqFISH+ barcoding. **c**, Illustration showing the decrease of fluorescence intensity of a paint block (left), and DNA spots located at the periphery of the paint block are dropped (right). **d**, Illustration of down-sampling simulation on 60 genomic loci. The 60 continuous 25-kb genomic loci are splitted into the first and second 30-loci blocks, and identified genomic loci furthest from block centroid are dropped in simulated down-sampling. **e**, Total counts of detected 25-kb genomic loci with identities from 1-60 across 20 chromosomes before and after down-sampling in DNA seqFISH+ data (Takei et al. 2021). The green and red bars mark simulated block boundaries. **f**, Representative median spatial distance maps between pairs of intra-chromosomal loci before and after down-sampling in DNA seqFISH+ data (Takei et al. 2021). The green and red bars above distance matrices mark simulated block boundaries. n = 1,076 cells from two biological replicates of mESCs by two-layer DNA seqFISH+ in this study and n = 446 cells from two biological replicates of mESCs by original DNA seqFISH+ (Takei et al. 2021).

Down-sampling does not affect chromatin profiling analysis

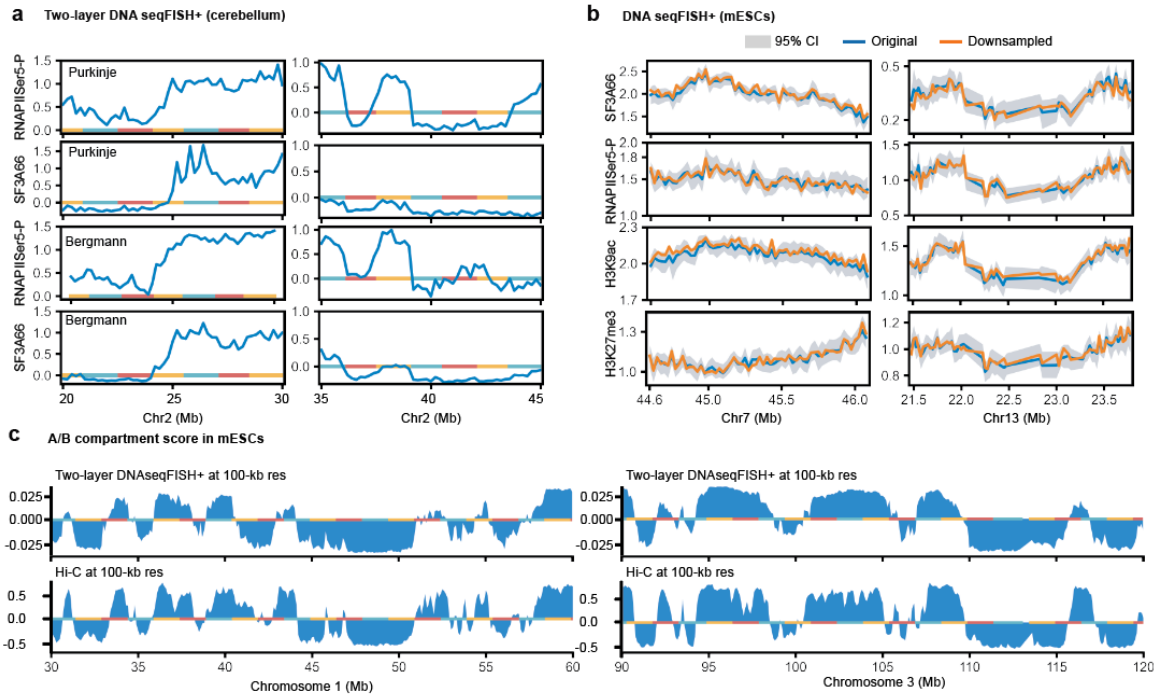
In order to test the effect of the down-sampling of genomic loci at the paint block boundaries on the imaging-based chromatin profiles, we examined the changes of chromatin profiles by two-layer DNA seqFISH+ at the block boundaries. Specifically, we plotted RNAPIISer5-P and SF3A66 profiles in Purkinje cells and Bergmann glia at 200-kb resolution, a resolution we used in this study. Notably, those tracks showed sub-megabase patterns with no discontinuities bias at the block border (Appendix Fig. 2a), confirming that the down-sampling of genomic loci does not have a significant effect on the chromatin profiling analysis in our two DNA seqFISH+ data.

Furthermore, we used the published DNA seqFISH+ datasets along with sequential immunofluorescence(Takei et al. 2021) to simulate the effect of down-sampling on chromatin profiling analysis, similarly to Section 1. First, we computed z-score normalized fluorescence intensities of major immunofluorescence markers (SF3A66, RNAPIISer5-P, H3K9ac and H3K27me3) of the published DNA seqFISH+ datasets as we performed in the two-layer DNA seqFISH+ analysis. Median ensemble z-scores of selected immunofluorescence markers for each genomic loci were calculated for original and down-sampled data at 25-kb, 50-kb, and 100-kb resolutions. Simulated down-sampled immunofluorescence z-scores of most genomic loci fell within the 95% confidence interval (98.85% for 25-kb resolution, 96.32% for 100-kb resolution, 94.66% for 200-kb resolution) (Appendix Fig. 2b). Those results further support that the down-sampling of genomic loci does not have a significant effect on the chromatin profiling analysis in DNA seqFISH+ data.

Down-sampling does not affect A/B compartment analysis

We further confirmed that A/B compartments were not affected by the down-sampling by comparing our results with published Hi-C datasets in mESCs and NMuMG cells(Bonev et al. 2017; Pascual-Reguant et al. 2018). Specifically, we computed the A/B compartment scores using selected immunofluorescence markers including H3K27ac, H3K27me3, H3K9me3, and LaminB1 as described in Methods (see “Imaging-based chromatin profiling analysis” section). We then examined the changes of computed A/B compartment scores at

the block boundaries used for the two-layer DNA seqFISH+ barcoding. Notably, we did not observe the bias at the paint borders, suggesting that the down-sampling does not affect the A/B compartment scores we obtained.



Appendix Fig. 2| Representative examples showing that down-sampling does not affect chromatin profiling and A/B compartment analysis.

a, Representative examples of RNAPIISer5-P and SF3A66 chromatin profiles by two-layer DNA seqFISH+ in Purkinje and Bergman cells in 10 Mb regions. Colored lines show different 60-loci blocks. **b**, Representative examples of chromatin profiling for each marker before and after down-sampling computed from mESC 25-kb DNA seqFISH+ data(Takei et al. 2021). **c**, Representative examples of A/B compartment scores (y-axis) computed from two-layer barcode DNA seqFISH+ chromatin profile, compared with Hi-C computed A/B compartment scores(Bonev et al. 2017; Pascual-Reguant et al. 2018). Colored lines show different 60-loci blocks. n = 1,076 cells from two biological replicates of mESCs and n = 384 cells from one biological replicate of NMuMG cells by two-layer DNA seqFISH+ in

this study and n = 446 cells from two biological replicates of mESCs by original DNA seqFISH+(Takei et al. 2021).

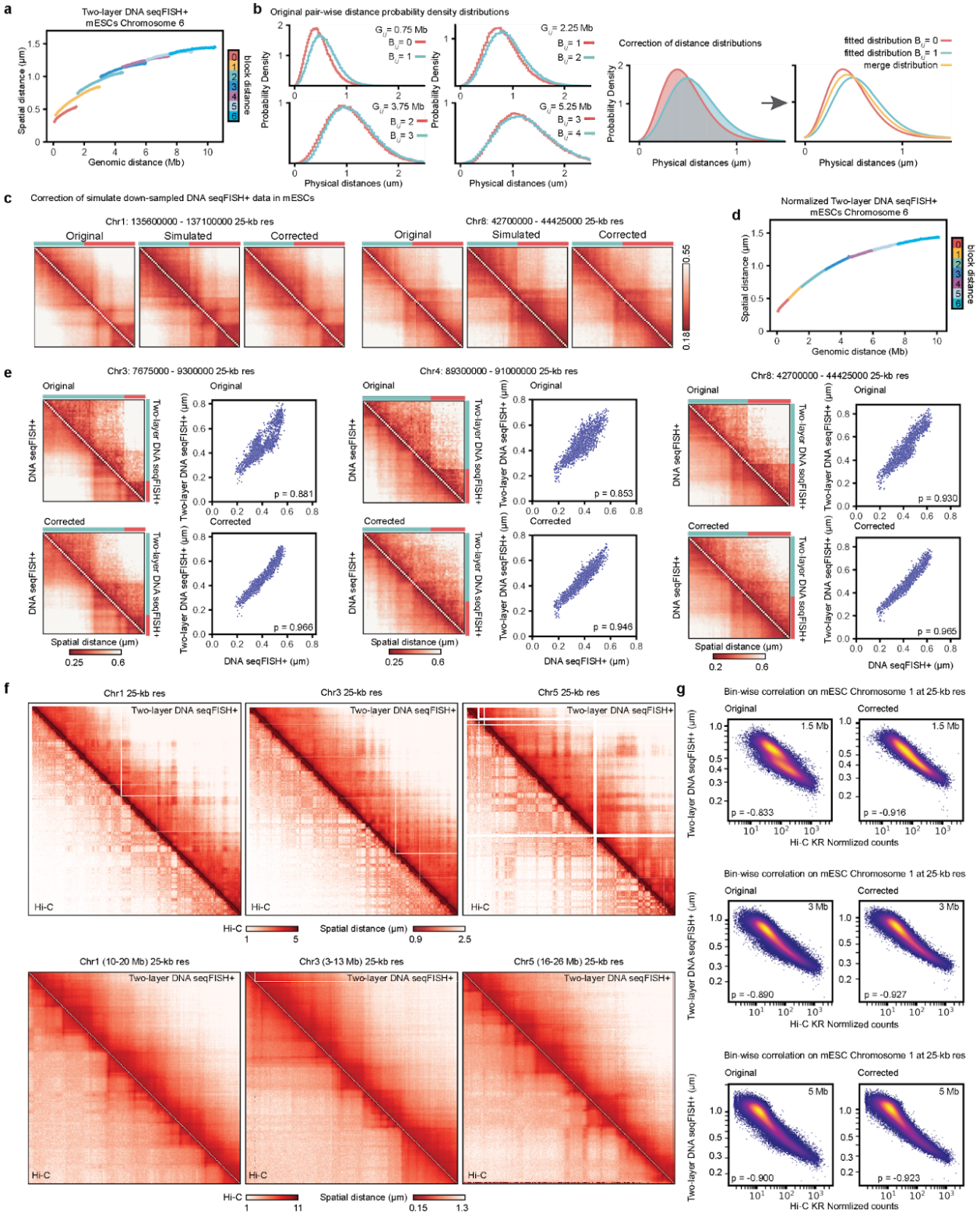
Normalization to remove block patterns in the spatial distance maps

Since the block pattern in the spatial distance maps we described in Section 1 can be quantified as a pairwise distance shift related to block distance, we tested whether correcting distance shifts will remove the block pattern in spatial distance maps. According to the two-layer barcoding scheme, each genomic locus was assigned two coordinates: genomic coordinate (1D genomic location) and block coordinate (block identity). Thus, for a given genomic locus pair i,j, there are three distances: genomic distance G_{ij}, block distance B_{ij}, and spatial distance. By plotting spatial distance of pairs of genomic loci as a function of genomic distance, we found that the spatial distance is smaller for any given locus pair i,j with a fixed genomic distance when the pair loci are located at closer blocks(Appendix Fig. 3a). We considered that this distance shift results from the down-sampling of genomic loci due to the two-layer barcoding scheme as discussed in Section 1.

To correct the spatial distance shift, we corrected the distribution of pairwise distances at the single chromosome level. Genomic locus pairs separated by a fixed genomic distance could be located in genomically closer blocks (or the same block) or genomically further blocks. We shifted both the distance distributions for pairs located at closer blocks and further blocks to an intermediate distribution (Appendix Fig. 3b), and recalculated a normalized distance map as detailed in Methods (see “Pairwise spatial distance analysis for DNA loci” section).

By applying this normalization method, we showed that the block patterns were successfully removed both from simulated down-sampled DNA seqFISH+ data and two-layer barcoded DNA seqFISH+ data in mESCs. In the published DNA seqFISH+ data, simulated down-sampled distance matrix was recovered after correction (Appendix Fig. 3c). In the two-layer DNA seqFISH+ data, block patterns disappeared in normalized distance matrices (Appendix Fig. 3d), reflected in a better bin-wise correlation with corresponding distance maps computed by DNA seqFISH+ data (Appendix Fig. 3e), and Hi-C contact matrices in diagonal

plot (Appendix Fig. 3f). The bin-wise correlation of Hi-C contact matrices and corrected distance matrices improved after normalization (from -0.833 to -0.910 for genomic locus closer than 1.5 Mb, from -0.885 and -0.919 for genomic locus closer than 3 Mb, Chromosome 1 is shown in Appendix Fig. 3g). Genome-wide bin-wise correlation with Hi-C did not change significantly, since the block patterns are originally only notable for close genomic locus pairs (<5 Mb).


a
Two-layer DNA seqFISH+
mESCs Chromosome 6
Spatial distance (μm)
Genomic distance (Mb)
block distance
b
Original pair-wise distance probability density distributions
Probability Density
Physical distances (um)
Correction of distance distributions
fitted distribution B_U = 0
fitted distribution B_U = 1
merge distribution
Physical distances (μm)
c
Correction of simulate down-sampled DNA seqFISH+ data in mESCs
Chr1: 135600000 - 137100000 25-kb res
Chr8: 42700000 - 44425000 25-kb res
Original
Simulated
Corrected
d
Normalized Two-layer DNA seqFISH+
mESCs Chromosome 6
e
Chr3: 7675000 - 9300000 25-kb res
Chr4: 89300000 - 91000000 25-kb res
Chr8: 42700000 - 44425000 25-kb res
DNA seqFISH+
Two-layer DNA seqFISH+ (μm)
DNA seqFISH+ (μm)
Spatial distance (μm)
p = 0.881
p = 0.966
p = 0.853
p = 0.946
p = 0.930
p = 0.965
f
Chr1 25-kb res
Chr3 25-kb res
Chr5 25-kb res
Two-layer DNA seqFISH+
Hi-C
Chr1 (10-20 Mb) 25-kb res
Chr3 (3-13 Mb) 25-kb res
Chr5 (16-26 Mb) 25-kb res
g
Bin-wise correlation on mESC Chromosome 1 at 25-kb res
Two-layer DNA seqFISH+ (μm)
Hi-C KR Normlized counts
1.5 Mb
3 Mb
5 Mb
p = -0.833
p = -0.916
p = -0.890
p = -0.927
p = -0.900
p = -0.923

Appendix Fig. 3| Normalization of block patterns can correct ensemble-averaged spatial distance maps.

a, Median physical distance computed from two-layer DNA seqFISH+ as a function of genomic distance at 25-kb resolution, separated by different block distances between genome loci pairs for chromosome 6 in mESCs. **b**, Pairwise distance probability density function of genomic loci pairs with genomic distance ($G_{i,j}$) at 0.75 Mb, 2.25 Mb, 3.75 Mb, 5.35 Mb separated by different paint block distances ($B_{i,j}$) (left panel). Right panel illustrates the fitted gamma distribution (left) and merged gamma distribution (right). **c**, Normalization of simulated data. Median spatial distance between pairs of intra-chromosomal loci from simulation of DNA seqFISH+ at 25-kb resolution, original (left), down-sampled (middle) and normalized down-sampled data (right). Red and green bars above distance matrices mark block boundaries. **d**, Physical distance as a function of genomic distance at 25-kb resolution after normalization of two-layer DNA seqFISH+ data. **e**, Concordance of DNA seqFISH+ distance map with original (upper panel) and corrected (lower panel) two-layer DNA seqFISH+ distance map at 25-kb resolution. For each panel, bottom left shows DNA seqFISH+ distance map, top right shows original or normalized two-layer DNA seqFISH+ map. **f**. Concordance between corrected two-layer DNA seqFISH+ (top right) and Hi-C maps (bottom left) at 25-kb resolution. Upper panel shows entire chromosomes, bottom panel shows 10 Mb genomic regions. **g**, Bin-wise correlation of Hi-C maps to original (left) and corrected (right) two-layer DNA seqFISH+ distance matrices on chromosome 1 at 25-kb resolution in mESCs. Upper panels show the correlation of genomic loci separated within 1.5-Mb, the center panels use genomic loci separated up to 3 Mb, and the lower panels use genomic loci separated up to 5 Mb. n = 1,076 cells from two biological replicates of mESCs by two-layer DNA seqFISH+ in this study and n = 446 cells from two biological replicates of mESCs by original DNA seqFISH+ (Takei et al. 2021).

Limitations

For cerebellum data, due to the difficulty in separating alleles compared to cultured cell data, the distance distribution across genomic locus pairs with a fixed genomic distance appeared as a mixture of gamma and gaussian distributions. Thus, the normalization approach could

not be applied to cerebellum data. While the normalization procedure can remove the block effects in the ensemble-averaged pairwise distance matrices, to recover the loci near block boundaries in single cells, future works can use deep-learning image processing tools to ameliorate the effect of the lower paint intensities near the boundaries in the two-layer barcoding schemes.

Discussions

The block pattern from down-sampling of genomic loci near block boundaries is the main reason that we were not able to use the median spatial distance maps for downstream analysis like loop detection. The normalized spatial distance maps were not accurate enough for reconstructing A/B compartments and chromosome loops, and we were not able to apply the normalization method to cerebellum tissue. If we can recover the down-sampled genomic loci, we may be able to reconstruct more accurate spatial distance maps without the block pattern, fully leveraging the comprehensive information in the dataset.

One solution to recover the down-sampled genomic loci is to segment the aggregated blob fluorescence signals of each chromosome block, aligning the masks of block signals from all rounds of hybridizations, and decode the block identities at block level. In this way, by changing the segmentation threshold, we are able to control to a certain extent how much of the boundary is included. Another solution is to use machine learning or deep learning to train a classifier on a down-sampling simulation of ground truth data, like in section 1, and apply it to the two-layer barcoded data.

Lastly, we observed that the down-sampling of genomic loci near block boundaries has minimal impact on the chromatin profiling. Since distance matrix calculations require simultaneous detection of genomic locus pairs, they are highly sensitive to detection efficiency drops (impact scales quadratically). In contrast, chromatin profiling is the z-score normalized fluorescence intensities for each genomic locus, thus is relatively resistant to moderate down-sampling effects, explaining why chromatin profiling data remains robust despite these technical artifacts.

Appendix references

Bonev, Boyan, Netta Mendelson Cohen, Quentin Szabo, Lauriane Fritsch, Giorgio L. Papadopoulos, Yaniv Lubling, Xiaole Xu, et al. 2017. "Multiscale 3D Genome Rewiring during Mouse Neural Development." Cell 171 (3): 557–72.e24.

Pascual-Reguant, Laura, Enrique Blanco, Silvia Galan, François Le Dily, Yasmina Cuartero, Gemma Serra-Bardenys, Valerio Di Carlo, et al. 2018. "Lamin B1 Mapping Reveals the Existence of Dynamic and Functional Euchromatin Lamin B1 Domains." Nature Communications 9 (1): 3420.

Takei, Yodai, Jina Yun, Shiwei Zheng, Noah Ollikainen, Nico Pierson, Jonathan White, Sheel Shah, et al. 2021. "Integrated Spatial Genomics Reveals Global Architecture of Single Nuclei." Nature 590 (7845): 344–50.

Chapter 3

AN ATLAS OF HUMAN FETAL KIDNEY ORGANOGENESIS INTEGRATING SINGLE-CELL MULTIOME SEQUENCING AND SPATIAL TRANSCRIPTOMICS

3.1 Introduction

The mammalian kidney plays a crucial role in maintaining body fluid homeostasis through its epithelial networks: the nephron and the collecting duct (Nielsen et al. 2012). The nephron eliminates toxic compounds, while also recovering molecules, such as proteins, metabolites, and ions. The collecting system regulates water, salt, and pH balance. Loss of kidney function, as in end-stage renal disease (ESRD), results in high morbidity and mortality, with dialysis and transplantation as primary treatments. A comprehensive understanding of kidney development is vital for developing accurate disease models and synthetic kidney surrogates (Oxburgh et al. 2017). While most current studies on kidney development utilize mouse models, it is important to note that the human kidney has unique features, including the completion of nephrogenesis before birth, which differs from mice where nephrogenesis continues postnatally, approximately 1 million nephrons (Hughson et al. 2003) compared to about 18,000 in mice, a multilobar structure versus the unilobar mouse kidney with a single lobe (Lindström, McMahon, et al. 2018), and distinct regulatory programs and gene activity during development (Kim et al. 2024). Therefore, constructing a complete atlas of human kidney development is essential. A complete atlas would serve as a reference for normal human kidney development and highlight the cellular and molecular features absent in current kidney organoids (Khoshdel-Rad, Ahmadi, and Moghadasali 2022; Fausto et al. 2024); these insights would guide efforts to extend organoid maturation to more advanced developmental stages and help identify novel therapeutic targets (Tran et al. 2022).

The development of the human kidney is a complex and spatially orchestrated process, with nephrogenesis being the essential aspect, where progenitor cells differentiate into functional nephrons (McMahon 2016; Cullen-McEwen, Sutherland, and Black 2016). Spatial cues,

including morphogen gradients and localized cell-cell interactions, play a crucial role in directing cell fate and organization (Lindström et al. 2021; Wu et al. 2022). The human fetal kidney at 17- to 18-weeks of gestation is particularly well-suited for studying kidney development, as it represents a period of active nephrogenesis. During this period, the fetal kidney contains a variety of cell types, including SIX2+ nephron progenitors and maturing nephrons (Lindström, De Sena Brandine, et al. 2018), each located in spatially distinct regions . Thus, examining one section of the human fetal kidney at this stage provides a comprehensive view of nephrons at various developmental stages, offering valuable insights into the process of nephrogenesis.

To fully understand the molecular and spatial complexities of kidney development, advanced genomic techniques are essential. Single-cell multiome sequencing, which allows single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) simultaneously from the same cell, provides insights into advanced cell type identification and gene regulation landscape (Ma et al. 2020; Guyer et al. 2023; Chopp et al. 2023). Spatial transcriptomics reveals the cell organization and cell-cell interactions in the kidney tissue (Polonsky et al. 2024; Abedini et al. 2024). By integrating both technologies, this thesis aims to construct a detailed atlas of the human fetal kidney at 17-to 18-week of gestation, serving as a valuable resource for the community.

3.2 Results

Jointly profiling of gene expression and chromatin accessibility of human fetal kidney

To generate a comprehensive atlas of human fetal kidney development that integrates single-cell molecular profiles with spatial organization, we employed 10X Genomics multiome sequencing and spatial transcriptomics by seqFISH on human 17.1- to 18.6-week kidneys (N = 4) (Fig. 1a). For each fetus, one kidney was processed with the 10X Genomics multiome workflow to capture both single-nucleus RNA and ATAC profiles simultaneously, providing detailed transcriptomic and epigenetic data at the single-cell level. Additionally, for the 17.1-week fetus, the remaining kidney was sectioned and analyzed with seqFISH to map spatial gene expression patterns, preserving the tissue's anatomical context. This approach yielded

high-quality paired transcriptomic and epigenomic measurements from individual cells via multiome sequencing, as well as spatially resolved gene expression data from seqFISH, enabling the integration of cellular identity with tissue architecture.

The transcriptomic data from the multiome dataset served as an anchor for integrating the spatial transcriptomics and single-cell multiome sequencing profiles. Therefore, we first analyzed the single-nucleus RNA sequencing (snRNA-seq) data from the multiome sequencing results using Seurat v5 pipeline (Hao et al. 2024). From this snRNA-seq data, we obtained 37,784 high-quality nuclei after performing rigorous quality control and removing potential doublets. Each nucleus had a median of 5,393 unique molecular identifiers (UMIs), ensuring sufficient sequencing depth for accurate cell type classification (You et al. 2021). We used Harmony from Seurat v5 to remove batch effects across different samples (Korsunsky et al. 2019). Subsequently, unsupervised uniform manifold approximation and projection (UMAP) clustering identified 23 distinct clusters corresponding to coarse cell types (Fig. 1b). Based on the expression of well-known marker genes, these clusters were grouped into five major lineages: nephrogenic cells (NPC), ureteric cells (UPC), interstitial cells (INT), vascular cells (VAS), and immune cells (IMM) (Fig. 1d) (Combes et al. 2019; Lindström, McMahon, et al. 2018; Kim et al. 2024). Violin plots illustrate the expression levels of key marker genes across these lineages, confirming their identities (Fig. 1d).

To identify fine-grained cell types within each lineage, we performed sub-clustering analysis on each lineage separately. For the nephrogenic lineage, which consists of the major components forming the nephron (the basic functional unit of the kidney), we identified 28 sub-clusters (Fig. 1c). Each sub-cluster was characterized by specific marker gene expression patterns, as shown in the violin plots (Fig. 1e), allowing us to map the developmental stages of these cells. Similarly, sub-clustering of the other lineages revealed a rich diversity of cell types (Fig. 1f), with their marker genes detected and visualized for the ureteric, interstitial, vascular, and immune compartments (Figs. 1g).

After confirming that we captured the major lineages and fine-grained cell types, we curated a set of 224 marker genes from the snRNA-seq data to probe in RNA seqFISH experiments. These genes were selected from the top marker genes of sub-clusters identified in the snRNA-seq data across all five lineages in the human fetal kidney. This selection enables the transcriptomic data to act as anchors for integrating the single-cell multiome sequencing and spatial transcriptomics data, allowing the projection of jointly profiled epigenetic information onto the spatial profiling based on cell identities determined from snRNA-seq and seqFISH. To enhance coverage, we also included expert-curated conventional marker genes that were not detected as marker genes in the snRNA-seq analysis (Fig. 1h). Among the 224 genes, 200 were probed using the barcoded RNA seqFISH+ strategy, with 6 pseudocolors and 4 rounds of hybridization on the Cy3b channel, while the remaining 24 genes were probed using a non-barcoded seqFISH approach on the 488 nm channel.

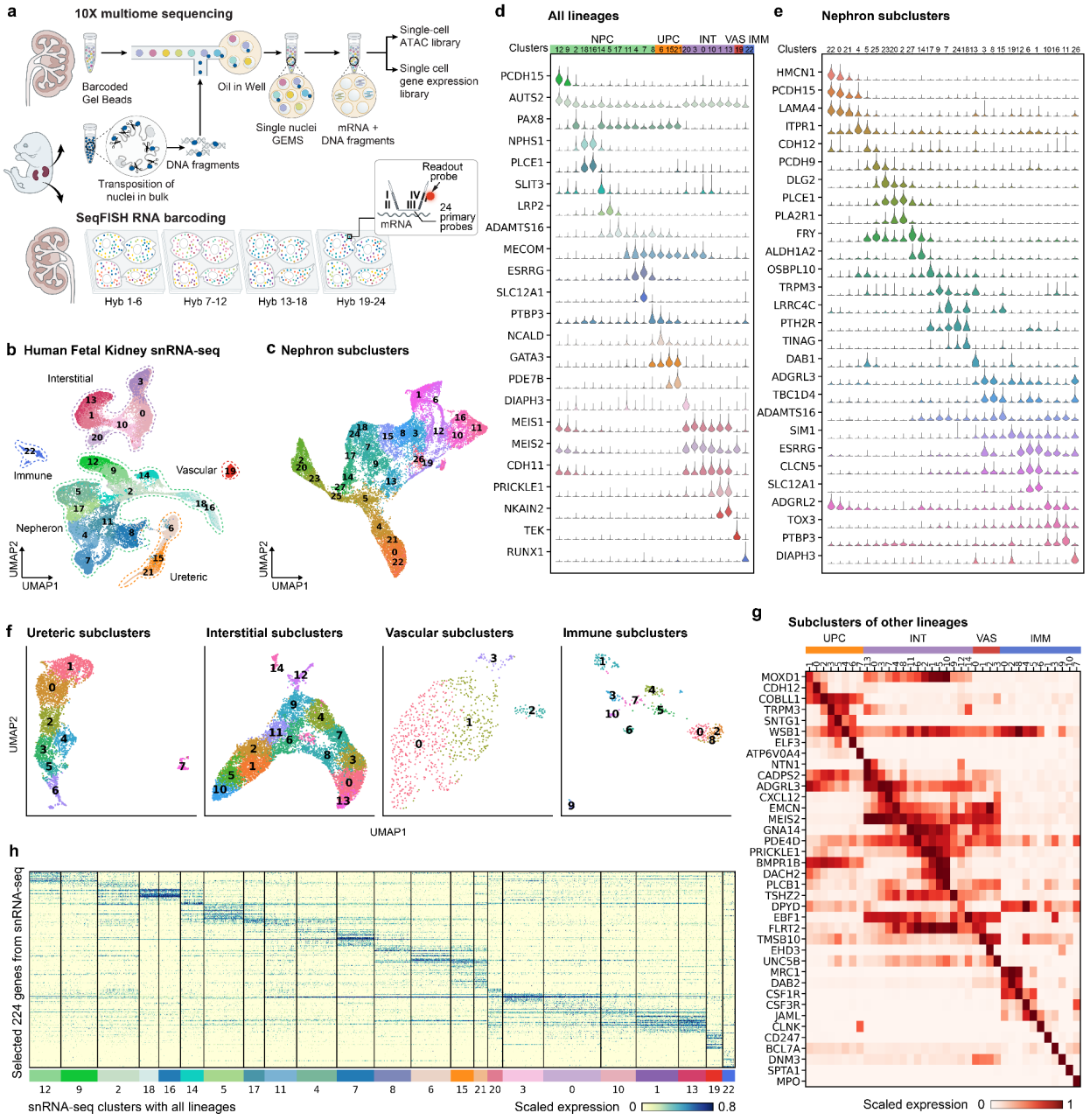


Fig. 1 | Identification of coarse and fine-grained cell types in human fetal kidneys using snRNA-seq

a, Schematic of the experimental methodology: From four human fetuses aged 17.1- to 18.6-week, one kidney per fetus was subjected to 10X multiome sequencing. Additionally, the remaining 17.1-week female kidney was used for spatial transcriptomics profiling via seqFISH. **b,** UMAP embedding of single-nucleus RNA sequencing data from four human

fetal kidney samples, comprising 37,784 cells clustered into 23 groups (0–22), colored by cluster assignments. Five main lineage clusters are circled and labeled: nephrogenic cells (NPC), ureteric cells (UPC), interstitial cells (INT), vascular cells (VAS), and immune cells (IMM). Harmony batch correction was applied to integrate the four samples. **c,** UMAP embedding of sub-clustered nephrogenic cells from four human fetal kidney samples, colored by sub-cluster assignments from 0-27, comprising 19247 cells. **d,** Violin plot showing the expression levels of marker genes for 23 major cell types (clusters 0–22) across all lineages, ordered by the developmental stages of nephrogenic cells (NPC), ureteric cells (UPC), interstitial cells (INT), vascular cells (VAS), and immune cells (IMM). The cluster 0-22 definition follows the cluster assignment in a. **e,** Violin plots showing marker gene expression for 28 nephrogenic sub-clusters (0–27), ordered by their developmental stages. The cluster definition follows cell assignment in c. f, UMAP embeddings of sub-clustered cells from the remaining four lineages: 3928 ureteric cells (UPC) with subclusters from 0-7, 10272 interstitial cells (INT) with subclusters 0-14, 816 vascular cells (VAS) with subclusters 0-3, and 435 immune cells with subclusters 0-10 (IMM). g, Heatmap showing scaled marker gene expression levels for sub-clusters within the four lineages described in panel **f,** ordered by the developmental stages of ureteric cells (UPC), interstitial cells (INT), vascular cells (VAS), and immune cells (IMM). h, Heatmap of single-nucleus expression levels for 224 selected marker genes for seqFISH profiling, across all 37,784 cells, clustered as in panel b, ordered by developmental stages in the same sequence as panel d.

seqFISH profiling of human fetal kidney

We performed RNA seqFISH+ on a 15 μm thick coronal section from the 17.1 week old human fetal kidney, confirming that the midline section includes the cortex with the nephrogenic zone, medulla with renal pyramids, and renal pelvis, encompassing all major anatomical structures of the fetal kidney (Fig. 2a). To direct cluster the seqFISH data, scTransform normalization were applied on the seqFISH data to remove the detection efficiency bias of barcoded (563 nm) and non-barcoded (488 nm) channels (Hafemeister and Satija 2019), and UMAP dimension reduction was performed for visualizing different cell types (Fig. 2b). Using 224 measured genes, we identified all five major lineages involved in

kidney organogenesis and their subtypes through direct clustering of the seqFISH+ data (Fig. 2c). Notably, rare cell types, including RENIN+ and FOXD1+ cells, were detected in the coronal section, as shown in the UMAP plot of clustering. The spatial distribution of marker genes representing key nephrogenesis structures and lineage specific cell types corresponded closely with their expected locations (Fig. 2g).

In addition to the recovery of rare cell populations, we identified a previously unreported cell type: SLC12A3+ ureteric cells. Typically, SLC12A3 is expressed in distal convoluted tubule (DCT) cells of the nephron (Miao et al. 2021; Li and Gu 2022). However, seqFISH data revealed low but detectable SLC12A3 expression in a subset of ureteric cells located in the deep medullary region (Fig. 2h). To confirm that this represents a genuine cell type and not a technical artifact, we integrated snRNA-seq and seqFISH+ data by projecting nephron progenitor (NPC) and ureteric epithelium (UE) cells into a shared latent space using harmony (Fig. 2i). This further confirmed that SLC12A3+ cells form a distinct subpopulation within the ureteric lineage, present in both sequencing and spatial datasets, validating the authenticity of the SLC12A3 signal (Fig. 2j). Notably, this subpopulation was not detected when clustering snRNA-seq data alone but emerged only upon co-clustering with spatial data, underscoring the ability of spatial transcriptomics to identify novel cell types using a small panel of 224 genes. This unexpected expression pattern suggests a potential specialized role for SLC12A3 during the 17-week stage of human kidney development.

While nephron, ureteric, and interstitial lineages were accurately identified, vascular endothelial and immune cells suffered from contaminated signal from neighbor cells. Subclustering analysis revealed significant contamination of their clusters with marker genes from other lineages, such as PDGFRA, PDGFRB, and MEIS1 from interstitial cells. This contamination likely resulted from inaccurate cell segmentation, leading to misassignment of transcripts from adjacent cells. The elongated morphology of vascular endothelial cells, characterized by small nuclei, and the small size and variable shapes of immune cells made them particularly susceptible to segmentation errors in 2D spatial experiments. Precise cell segmentation in spatial transcriptomics, especially for cell types with complex morphologies, is necessary for future experiments.

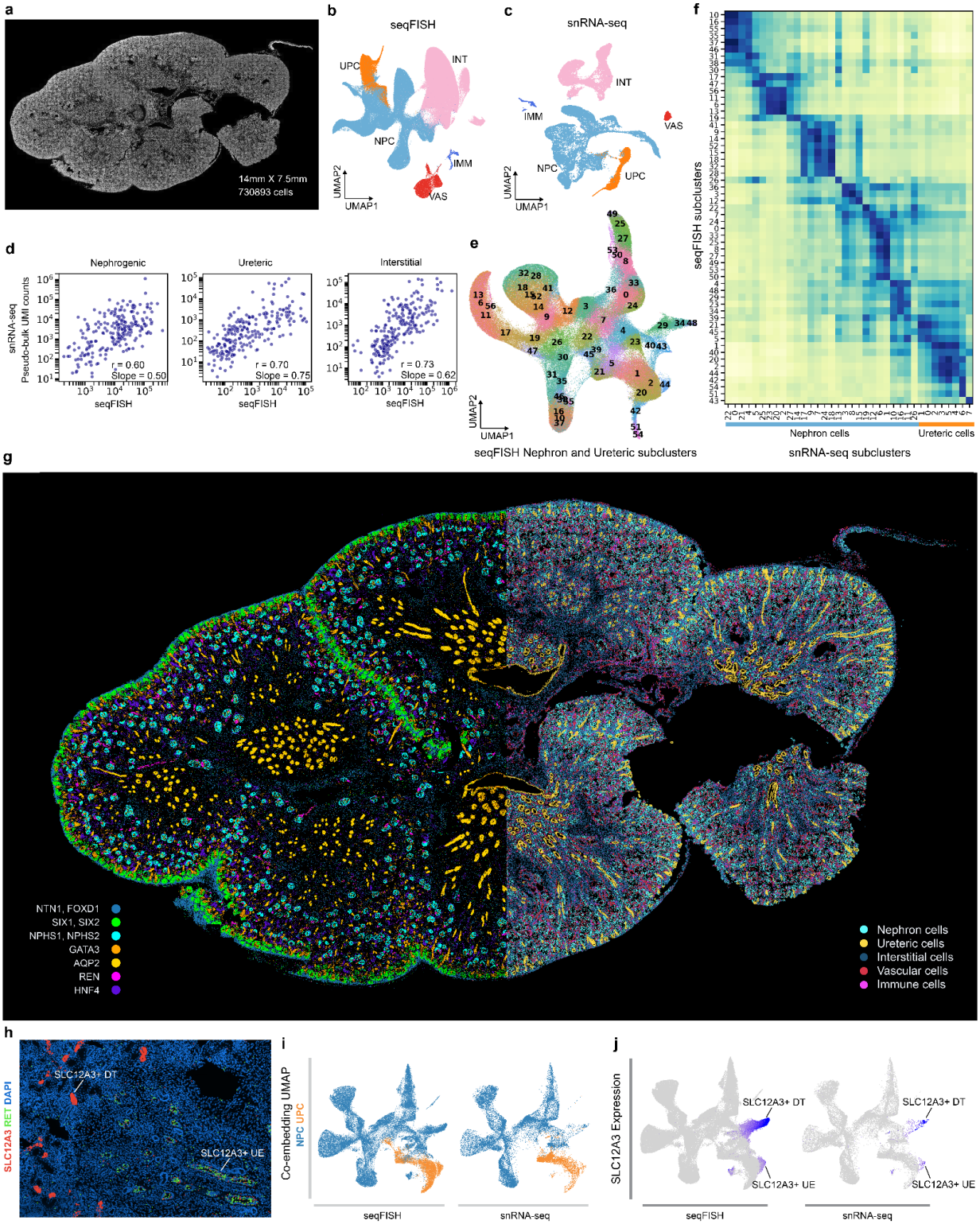

a
14mm X 7.5mm
730893 cells
b
seqFISH
UPC
INT
NPC
IMM
VAS
UMAP2
UMAP1
c
snRNA-seq
INT
IMM
VAS
NPC
UPC
UMAP2
UMAP1
f
seqFISH subclusters
Nephron cells
Ureteric cells
snRNA-seq subclusters
d
Nephrogenic
Ureteric
Interstitial
snRNA-seq
Pseudo-bulk UMI counts
r = 0.60
Slope = 0.50
r = 0.70
Slope = 0.75
r = 0.73
Slope = 0.62
seqFISH
e
UMAP2
UMAP1
seqFISH Nephron and Ureteric subclusters
g
NTN1, FOXD1
SIX1, SIX2
NPHS1, NPHS2
GATA3
AQP2
REN
HNF4
Nephron cells
Ureteric cells
Interstitial cells
Vascular cells
Immune cells
h
SLC12A3 RET DAPI
SLC12A3+ DT
SLC12A3+ UE
i
Co-embedding UMAP
NPC UPC
seqFISH
snRNA-seq
j
SLC12A3 Expression
SLC12A3+ DT
SLC12A3+ UE
SLC12A3+ DT
SLC12A3+ UE
seqFISH
snRNA-seq

Fig. 2 | Validation and characterization of seqFISH on human fetal kidney

a, Whole-slide DAPI image of the 17 week female human fetal kidney coronal section. **b,** UMAP embedding of direct seqFISH clustering of cells, colored by lineage assignments (NPC: nephrogenic cells, UPC: ureteric cells, INT: interstitial cells, VAS: vascular cells, IMM: immune cells). **c,** UMAP embedding of single-nucleus RNA sequencing clustering of cells from 4 human fetal kidney samples, colored by lineage assignments. **d,** Pearson correlation of pseudo-bulk counts between seqFISH and single-nucleus RNA sequencing for the 224 spatially profiled mRNAs within nephrogenic, ureteric, and interstitial lineages. **e,** UMAP embedding of direct SeqFISH clustering of cells from nephrogenic and ureteric lineages, colored by subcluster identification (as detailed in f). **f,** Comparison of cell clusters (0-54) defined by seqFISH and cell types identified by single-nucleus RNA sequencing. Cells from nephrogenic and ureteric lineages are ranked first by the developmental stage within the nephron lineage and then within the ureteric lineage, based on single-nucleus RNA sequencing. **g,** Overview of whole-slide seqFISH profiling, with the left half displaying detected marker genes for key cell types and the right half showing each detected cell colored by its lineage assignment. **h,** Zoom-in of a section showing detected SLC12A3 and RET mRNAs. RET is a marker gene for ureteric cells, illustrating the reliable detection of SLC12A3+ nephron cells in the distal tubule, as well as ureteric cells. **i,** UMAP embedding of Harmony batch-corrected co-embedding of SeqFISH and single-nucleus RNA sequencing cells from both nephrogenic and ureteric lineages, colored by lineage. **j,** Feature plot displaying the expression levels of SLC12A3 from both seqFISH and single-nucleus RNA sequencing, revealing two distinct SLC12A3+ populations in the single-nucleus RNA sequencing data when co-embedded with spatial data.

Imputation of spatially unmeasured gene expression

Having validated that both snRNA-seq and seqFISH+ profiles faithfully captured major and rare cell types in the 17-week and 18-week human fetal kidneys, we next explored whether genes not directly probed in seqFISH could be inferred from snRNA-seq gene expression

profiles. Because vascular and immune cells were sparse in our snRNA-seq dataset (816 vascular cells and 435 immune cells), we focused on the major nephron, ureteric, and interstitial lineages, selecting 194 highly variable genes within those lineages from the 224 genes profiled by seqFISH+. Notably, the seqFISH+ dataset profiled a substantially larger number of cells (730,893) compared to snRNA-seq (37,784 cells). After subdividing the whole fetal kidney slide into five anatomical lobes, we applied tangram imputation independently to each lobe (see Methods) (Fig. 3a) (Biancalani et al. 2021). Tangram mapping revealed a high correlation between imputed and measured seqFISH+ gene expressions for most genes (Fig. 3b, left panel). Leave-one-out cross-validation (LOOCV) further showed that 82.99 % of genes maintained correlation coefficients above 0.2 (Fig. 3b left panel). As expected, the few genes like with LOOCV scores below 0.1, such as unique cluster markers REN and TUBB3, showed lower accuracy when left out, presumably because seqFISH+ lacked alternative markers for those specialized cell populations. Encouragingly, rare populations such as FOXI1+ intercalated cells were clearly recovered.

A key challenge in mapping single-nucleus RNA sequencing (snRNA-seq) gene expression data to spatial coordinates is evaluating the reliability of the imputed gene expressions. Since Tangram employs a probabilistic approach, results can vary across independent runs. We hypothesize that genes which cannot be accurately imputed also exhibit lower consistency across multiple imputation runs. To test this hypothesis, we performed two rounds of Leave-One-Out Cross-Validation (LOOCV) for all 194 genes. We found that the correlation between the LOOCV-imputed gene expressions and the seqFISH+ ground truth was positively associated with the Pearson correlation of imputation across independent Tangram runs (Fig. 3c). Genes that showed high consistency across runs but lower correlations with the seqFISH+ ground truth often exhibited sparse expression in seqFISH+, even though they had clear expression patterns in snRNA-seq. These findings suggest that consistency across multiple Tangram runs is an effective predictor of imputation accuracy.

To assess the imputation consistency for genes not probed in seqFISH+, we performed five independent runs of Tangram and calculated the median correlation between every pair of

runs for each gene. A total of 4,875 genes demonstrated high consistency (> 0.4) across these runs. We further validated this approach using five known glomerulus podocyte markers: NPHS1, CLIC5, NTNG1 (Menon et al. 2018), TCF21 (Kliewe et al. 2021), and CLIC3 (Cheval et al. 2011), with median correlations of 0.96, 0.90, 0.86, and 0.10, respectively. For four of these five genes, the imputed expressions agreed well with the expected spatial patterns. However, CLIC3 exhibited a lower median correlation (0.10) and less accurate imputation judging by its imputed gene expression pattern(Fig. 3g). Additionally, the gene TCF21, which is expressed in both nephron and interstitial lineages according to snRNA-seq (Fig. 3f), was accurately predicted, with higher imputation values in both lineages.

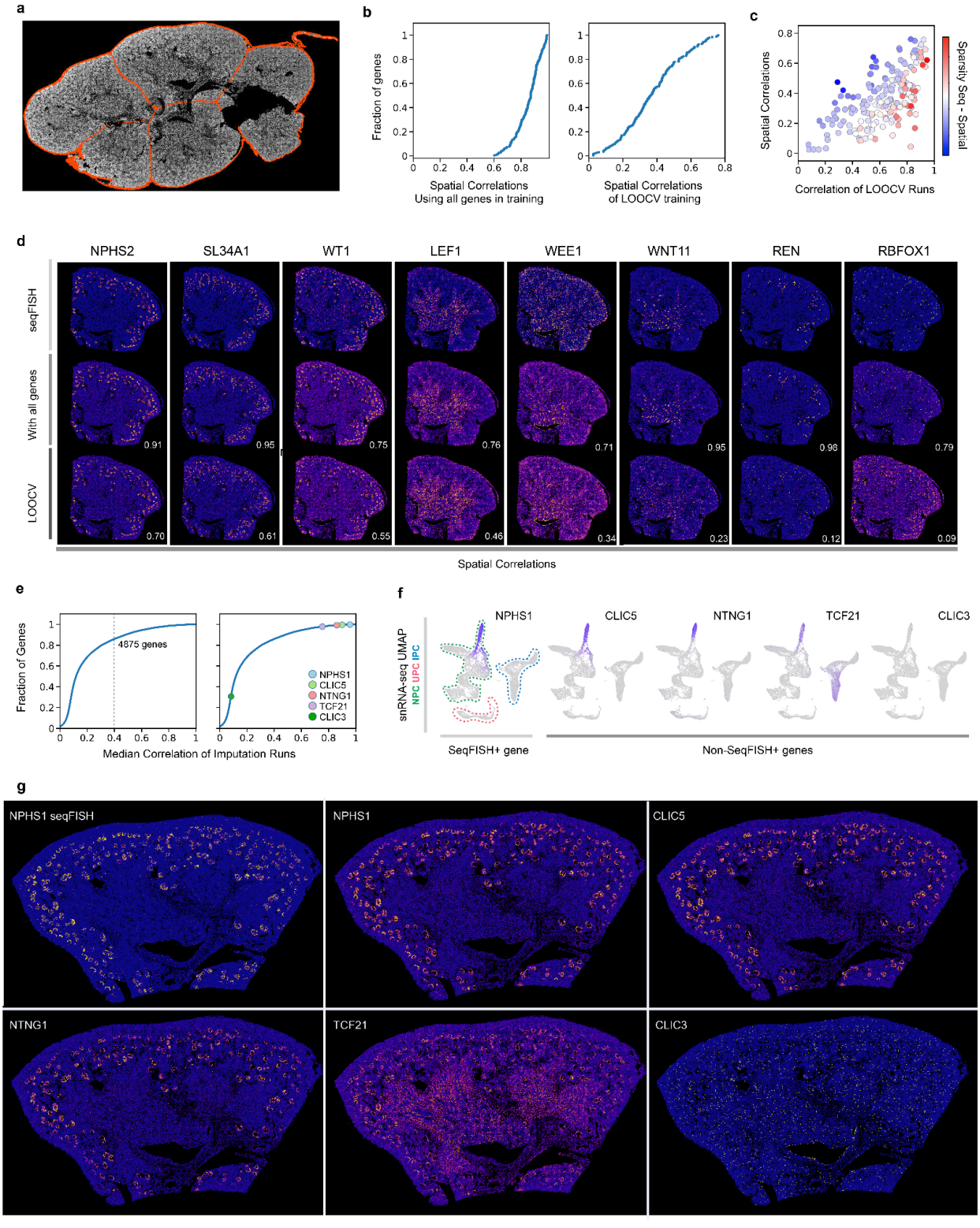

a
b
Fraction of genes
Spatial Correlations
Using all genes in training
Spatial Correlations
of LOOCV training
c
Spatial Correlations
Correlation of LOOCV Runs
Sparsity Seq - Spatial
d
NPHS2
SL34A1
WT1
LEF1
WEE1
WNT11
REN
RBFOX1
seqFISH
With all genes
LOOCV
0.91
0.95
0.75
0.76
0.71
0.95
0.98
0.79
0.70
0.61
0.55
0.46
0.34
0.23
0.12
0.09
Spatial Correlations
e
Fraction of Genes
4875 genes
NPHS1
CLIC5
NTNG1
TCF21
CLIC3
Median Correlation of Imputation Runs
f
snRNA-seq UMAP
NPC UPC IPC
NPHS1
CLIC5
NTNG1
TCF21
CLIC3
SeqFISH+ gene
Non-SeqFISH+ genes
g
NPHS1 seqFISH
NPHS1
CLIC5
NTNG1
TCF21
CLIC3

Fig. 3 | Imputation of unprofiled genes through integrating single-nucleus RNA sequencing and seqFISH

a, Whole-slide DAPI image of the coronal section from the 17-week-old female human fetal kidney. Red borders indicate manually defined anatomical lobes used for downsampling in Tangram computations. **b,** ECDF plots depicting the spatial correlation score, calculated as the cosine similarity between imputed gene counts and ground truth seqFISH gene counts, as defined in the Tangram paper [citation]. The left panel illustrates the imputation results trained using all 194 highly variable genes from cells in the nephrogenic, ureteric, and interstitial lineages. The right panel presents the leave-one-out cross-validation (LOOCV) spatial correlation for these 194 genes. **c,** Scatter plot where each point represents a gene, with the x-axis showing the Pearson's correlation coefficient between imputed expressions from two independent LOOCV rounds and the y-axis showing the spatial correlation score from the LOOCV test. Points are colored based on the difference in sparsity, defined as the sparsity of single-nucleus sequencing data minus that of SeqFISH data for each gene. **d, e,** Left: ECDF plot of imputation robustness scores for all genes, calculated as the median Pearson's correlation across all pairs from five independent Tangram imputation runs. A threshold of 0.4 on this median correlation classifies 4,875 genes as reliably imputed. Right: Highlight of five glomerulus marker genes with robustness scores ranging from high to low. **f,** Feature plots displaying the expression levels of the selected five glomerulus genes from single-nucleus RNA sequencing data. **g,** Spatial expression patterns comparing ground truth seqFISH and imputed expression for NPHS1 which marks glomerulus region, alongside imputed expressions for four additional glomerulus marker genes (shown in panels d and f) not detected in seqFISH. Notably, the gene CLIC3, with a low robustness score, exhibits a poor spatial pattern in its imputed expression.

3.3 Discussion

Through integration of single-nucleus RNA sequencing (snRNA-seq) and spatial transcriptomics profiling using seqFISH, we constructed a comprehensive human fetal kidney atlas. This atlas accurately captures all major cell types from snRNA-seq dissociated cells and their spatial distributions. By carefully selecting a compact set of 224 marker genes,

we successfully reconstructed the complete spatial organization of all cell types within the complex two-dimensional architecture of the developing human fetal kidney tissue. Remarkably, this small gene panel enabled the accurate prediction of at least ~4,700 gene expression patterns through the integration of snRNA-seq and seqFISH data, demonstrating the efficiency and predictive power of our approach.

The next step in our research is to infer neighborhood relationships of gene regulation networks from the imputed gene expression data, which would elucidate local co-expression patterns critical for understanding spatial gene interactions. However, a significant challenge lies in validating these inferred relationships, requiring us to rigorously confirm the accuracy of imputed gene expression patterns, potentially from seqFISH experiments targeting another set of genes for validation. In contrast to the challenges faced with imputation-based predictions, we achieved accurate classification of cell types at a fine-grained resolution directly from the seqFISH data. These classifications align closely with the snRNA-seq results and our established knowledge of kidney development, making neigborhood analysis based on cell-typing a solid future direction.

Although single-cell ATAC sequencing (scATAC-seq) data was not explored in this thesis, it needs to be closely investigated in the future, to fully use the power of single-cell multiome sequencing. Mapping scATAC-seq data to spatial transcriptomics profiles could provide insights into the epigenetic regulatory mechanisms governing gene expression in the spatial context of the developing kidney. However, this integration is likely to be challenging due to the inherent differences in data modalities, including the sparsity of scATAC-seq data, requiring advanced computational frameworks for alignment. Nevertheless, within the existing sequencing data, we can leverage tools like SCENIC+ to infer gene regulatory networks, identifying key transcription factors that drive cell differentiation (Bravo González-Blas et al. 2023). Combining these networks with trajectory analysis will enable us to map developmental lineages, further elucidating the cellular dynamics of kidney organogenesis.

This dataset is exceptionally information-rich and poised to significantly advance our understanding of human kidney development. Its high-resolution mapping of cell types, spatial organization, and predictive gene expression patterns provides a robust foundation for investigating the molecular mechanisms underlying kidney organogenesis. Furthermore, the atlas holds potential for transformative applications, including modeling congenital kidney disorders, informing the development of kidney organoids, and guiding regenerative medicine strategies. Future efforts to validate spatial gene relationships and integrate epigenetic data will enhance the utility of this resource, establishing it as a cornerstone for developmental biology and nephrology research.

3.4 Methods

Single-cell preparation for 10X multiome sequencing

Collection, preservation, and preparation of human fetal kidneys followed the protocol as previously described (Kim et al. 2024). Libraries were prepared from dissociated single nuclei using the 10X Genomics Chromium Next GEM Single-Cell Multiome ATAC + Gene Expression kit, according to the manufacturer's instructions.

Primary probe design and synthesis

To spatially profile 224 marker genes in the human fetal kidney, we categorized the gene targets into two groups: 200 genes were examined using barcoded RNA seqFISH+, and 24 genes were profiled using non-barcoded RNA seqFISH. The barcoded RNA seqFISH+ primary probes were developed based on a previously described, modified RNA seqFISH+ encoding strategy, employing 6-pseudocolor bases across 4 rounds of barcoding, including one round for error correction. This approach enables the detection of up to 216 genes in a single fluorescent channel at 561 nm. The remaining 24 genes were encoded using a non-barcoded seqFISH method in another fluorescent channel at 488 nm. The design of probes for both mRNA seqFISH+ and non-barcoded RNA seqFISH was adapted from previously reported protocols with minor modifications (Eng et al. 2019; Takei et al. 2021, 2025). Each primary probe was constructed with a 35-nucleotide sequence for binding to the RNA target, a 15-nucleotide region for readout probe attachment, and a pair of 20-nucleotide primer

binding sites at the 5’ and 3’ ends to enable enzymatic amplification]. For each gene, 24 primary probes were designed in both barcoded and non-barcoded RNA seqFISH methods.

Primary probes for mRNA seqFISH+ and non-barcoded RNA seqFISH were produced by Twist Bioscience, following methods outlined in prior studies with slight adjustments (Eng et al. 2019; Takei et al. 2021, 2025). The amplification startes with PCR amplification of DNA templates using Q5 Hot Start High-Fidelity polymerase (NEB M0494S), followed by purification with the QIAquick PCR Purification Kit (Qiagen 28104). In vitro transcription was conducted at 42°C for 16 hours using NEB’s E2040S kit, supplemented with RNasin Ribonuclease Inhibitor (Promega N2111) and Pyrophosphatase (NEB M0361S). Reverse transcription was performed using Thermo Scientific’s EP0751 at 50°C for 2 hours, then at 55°C for an additional 2 hours, followed by heat inactivation at 85°C for 5 minutes. RNA hydrolysis was carried out with 1 M NaOH at 80°C for 10 minutes, neutralized with an equivalent molecular amount of 10 M acetic acid. The resulting probes were purified using SPRI beads eluted with RNase free water (Liu et al. 2023), and their concentrations were measured to allow pooling of barcoded and non-barcoded probes. Amplification used primer set 7, selected from previous research (Eng et al. 2019; Takei et al. 2021, 2025).

Readout probe design and synthesis

To minimize non-specific binding of readout probes to human fetal kidney sections, we first conducted a control experiment to evaluate readout random sticking on human fetal kidney tissue. Specifically, we screened 60 distinct 15-nt readout DNA oligos using the established tissue seqFISH protocol, omitting the primary probe pools during the primary probe incubation step (see Methods section for “Tissue slice preparation”). Based on this screening, we selected the 48 readout probes that exhibited the least sticking, assigning 24 to the 561 nm fluorescent channel and the remaining 24 to the 488 nm channel. The 60 readout sequences used for screening were selected from our previous RNA seqFISH+ studies (Eng et al. 2019; Takei et al. 2021, 2025). To prepare the fluorescently labeled readout oligos, amine-modified 15-nt readout DNA oligos were ordered from Integrated DNA Technologies, designed to bind the corresponding readout sequences on the primary probes.

These oligos were conjugated in-house to either Cy3B–NHS ester (GE Healthcare PA63101) for the 561 nm channel or Alexa Fluor 488–NHS ester (Invitrogen A20000) for the 488 nm channel, following previously described conjugation protocols (Eng et al. 2019).

Tissue slice preparation

Tissue slice preparation were performed following previous publications (Eng et al. 2019; Takei et al. 2021, 2025). In short, coverslips were first functionalized with (APTES; Sigma-Aldrich, catalog no. A3648) and coated with Poly-D-lysine (Sigma-Aldrich, catalog no. P7280). Paraformaldehyde (PFA) fixed human fetal kidney tissue blocks were stored in -80°C until processing. Coronal sections of human fetal kidney were sectioned using a cryostat (Leica CM3050) to 15 to 20 μm and immediately attached onto the functionalized coverslips. The sections were immediately fixed with 4% PFA in PBS at room temperature for 10 minutes, followed by three washes with 1× PBS to remove remaining PFA. SeqFISH imaging rounds were performed as previously described (Eng et al. 2019; Takei et al. 2021, 2025), with 100ms exposure time under channel 561 and 300ms exposure time under channel 488.

Image analysis

Image analysis were performed using pyFISH pipeline as described in prior studies (Polonsky et al. 2024), published code are shared through github: https://github.com/CaiGroup/pyfish_tools.git. The pyFISH pipeline is designed for barcoded and non-barcoded seqFISH experiments, took raw images across rounds of hybridizations from seqFISH experiments and codebook, output gene-by-cell matrix for single-cell analysis.

Single cell multiome sequencing data processing

10X Single-cell multiome sequencing data were processed using the CellRanger-ARC pipeline to (Satpathy et al. 2019) demultiplex, align to reference genome and transciptome, and generate gene-to-cell count matrices. The resulting filtered feature matrix was analyzed using the Seurat v5 pipeline. Rigorous quality control was applied to remove low-quality

cells based on standard metrics, including the number of total detected UMIs and mitochondrial gene fraction per cell. Potential doublets, identified by the co-expression of marker genes from multiple lineages, were also filtered out. To preserve biologically relevant populations, cells expressing FOXI1, characteristic of intercalated cells in the kidney, who were filtered out due to its low total gene counts, were recovered from filtered out cells.

The filtered gene expression matrix was normalized using scTransform to account for variations in sequencing depth and other technical factors. Principal Component Analysis (PCA) was performed to reduce dimensionality, followed by Harmony batch correction to integrate data from four samples and mitigate batch effects within the PCA embeddings. Clustering was conducted on the Harmony-corrected PCA space to delineate distinct cell populations. For visualization, Uniform Manifold Approximation and Projection (UMAP) was applied to the Harmony-corrected embeddings. Each cluster, representing a distinct cell type, was assigned to one of five main lineages based on the expression of top enriched marker genes. Subclustering was performed within each lineage using the same analytical pipeline to identify fine-grained cell types.

SeqFISH data processing

The seqFISH gene-by-cell matrix, generated by the pyFISH pipeline, was first filtered based on the probability that each segmented mask corresponds to a genuine cell. Segmentation was performed within the pyFISH pipeline using the Cellpose generalized model, utilizing the DAPI channel for nuclei and the 488 nm autofluorescence channel for cytoplasm (Stringer and Pachitariu 2025). Segmentations lacking a DAPI signal or corresponding to cells with an area smaller than 800 pixels were filtered out to exclude malformed cells. The filtered gene-by-cell matrix was then processed using the Seurat v5 pipeline, where cells with fewer than 10 total detected transcripts were removed. Subsequently, scTransform was applied to normalize differences in detection efficiency between barcoded and non-barcoded channels, as non-barcoded seqFISH typically exhibits higher detection efficiency. Standard principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) dimensionality reduction techniques were employed to cluster the cells. During

initial clustering, clusters with mixtured marker genes from different lineages were identified as spatial doublets, and were removed. Subclustering was then performed within each lineage following similar strategy, with lineage identification from the initial clustering.

Integration of single-nucleus RNA sequencing data and seqFISH data

Due to seqFISH+ profiling approximately 20 times more cells than the 10X multiome sequencing protocol (730,893 cells for seqFISH+ and 37784 cells for multiome sequencing), downsampling was necessary to enable computations to complete with reasonable time and resource for Tangram, a deep learning model that benefits from GPU acceleration for large datasets (Biancalani et al. 2021). To perform downsampling, the entire tissue section was manually divided into five anatomical lobes, each containing a varying number of cells. Within each lobe, cells were randomly downsampled to have the same number of cells in the input single-nucleus RNA sequencing (snRNA-seq) data from the multiome protocol.

Leave-One-Out Cross-Validation (LOOCV) was performed to assess the accuracy of gene imputation. For each gene, Tangram was used to predict its spatial expression pattern based on the remaining genes from the seqFISH data and the snRNA-seq data. The accuracy (spatial correlation metric) was measured by the cosine correlation between the predicted expression counts and the ground truth seqFISH+ counts for the left-out gene. To evaluate the robustness of the imputation, five rounds of Tangram imputations with identical settings were performed for the cells in lobe 4. For each gene, the Pearson correlation between every pair of the five imputation rounds was calculated, and the median of these correlations was used to quantify the robustness of imputation.

3.5 References

Abedini, Amin, Jonathan Levinsohn, Konstantin A. Klötzer, Bernhard Dumoulin, Ziyuan Ma, Julia Frederick, Poonam Dhillon, et al. 2024. “Single-Cell Multi-Omic and Spatial Profiling of Human Kidneys Implicates the Fibrotic Microenvironment in Kidney Disease Progression.” Nature Genetics 56 (8): 1712–24.

Biancalani, Tommaso, Gabriele Scalia, Lorenzo Buffoni, Raghav Avasthi, Ziqing Lu, Aman Sanger, Neriman Tokcan, et al. 2021. “Deep Learning and Alignment of Spatially Resolved Single-Cell Transcriptomes with Tangram.” Nature Methods 18 (11): 1352–62.

Bravo González-Blas, Carmen, Seppe De Winter, Gert Hulselmans, Nikolai Hecker, Irina Matetovici, Valerie Christiaens, Suresh Poovathingal, Jasper Wouters, Sara Aibar, and Stein Aerts. 2023. “SCENIC+: Single-Cell Multiomic Inference of Enhancers and Gene Regulatory Networks.” Nature Methods 20 (9): 1355–67.

Cheval, Lydie, Fabien Pierrat, Carole Dossat, Mathieu Genete, Martine Imbert-Teboul, Jean-Paul Duong Van Huyen, Julie Poulain, et al. 2011. “Atlas of Gene Expression in the Mouse Kidney: New Features of Glomerular Parietal Cells.” Physiological Genomics 43 (3): 161–73.

Chopp, Laura B., Xiaoliang Zhu, Yayi Gao, Jia Nie, Jatinder Singh, Parimal Kumar, Kelly Z. Young, et al. 2023. “Zfp281 and Zfp148 Control CD4+ T Cell Thymic Development and TH2 Functions.” Science Immunology 8 (89): eadi9066.

Combes, Alexander N., Luke Zappia, Pei Xuan Er, Alicia Oshlack, and Melissa H. Little. 2019. “Single-Cell Analysis Reveals Congruence between Kidney Organoids and Human Fetal Kidney.” Genome Medicine 11 (1): 3.

Cullen-McEwen, Luise, Megan R. Sutherland, and M. Jane Black. 2016. “The Human Kidney.” In Kidney Development, Disease, Repair and Regeneration, edited by Melissa H. Little, 27–40. Elsevier.

Eng, Chee-Huat Linus, Michael Lawson, Qian Zhu, Ruben Dries, Noushin Koulena, Yodai Takei, Jina Yun, et al. 2019. “Transcriptome-Scale Super-Resolved Imaging in Tissues by RNA seqFISH.” Nature 568 (7751): 235–39.

Fausto, Connor C., Fokion Glykofrydis, Navneet Kumar, Jack Schnell, Reka L. Csipan, Faith De Kuyper, Brendan Grubbs, et al. 2024. “Defining and Controlling Axial Nephron

Patterning in Human Kidney Organoids with Synthetic Wnt-Secreting Organizers." Synthetic Biology. bioRxiv. http://biorxiv.org/content/10.1101/2024.11.30.626171v1.

Guyer, Richard A., Rhian Stavely, Keiramarie Robertson, Sukhada Bhave, Jessica L. Mueller, Nicole M. Picard, Ryo Hotta, Julia A. Kaltschmidt, and Allan M. Goldstein. 2023. "Single-Cell Multiome Sequencing Clarifies Enteric Glial Diversity and Identifies an Intraganglionic Population Poised for Neurogenesis." Cell Reports 42 (3): 112194.

Hafemeister, Christoph, and Rahul Satija. 2019. "Normalization and Variance Stabilization of Single-Cell RNA-Seq Data Using Regularized Negative Binomial Regression." Genome Biology 20 (1): 296.

Hao, Yuhan, Tim Stuart, Madeline H. Kowalski, Saket Choudhary, Paul Hoffman, Austin Hartman, Avi Srivastava, et al. 2024. "Dictionary Learning for Integrative, Multimodal and Scalable Single-Cell Analysis." Nature Biotechnology 42 (2): 293–304.

Hughson, Michael, Alton B. Farris 3rd, Rebecca Douglas-Denton, Wendy E. Hoy, and John F. Bertram. 2003. "Glomerular Number and Size in Autopsy Kidneys: The Relationship to Birth Weight." Kidney International 63 (6): 2113–22.

Khoshdel-Rad, Niloofar, Amin Ahmadi, and Reza Moghadasali. 2022. "Kidney Organoids: Current Knowledge and Future Directions." Cell and Tissue Research 387 (2): 207–24.

Kim, Sunghyun, Kari Koppitch, Riana K. Parvez, Jinjin Guo, Maryanne Achieng, Jack Schnell, Nils O. Lindström, and Andrew P. McMahon. 2024. "Comparative Single-Cell Analyses Identify Shared and Divergent Features of Human and Mouse Kidney Development." Developmental Cell 59 (21): 2912–30.e7.

Kliewe, Felix, Andreas W. Kuss, Florian Siegerist, Sindy Schröder, Sandra Schordan, Nadine Artelt, Frances Kindt, et al. 2021. "Studies on the Role of the Transcription Factor Tcf21 in the Transdifferentiation of Parietal Epithelial Cells into Podocyte-Like Cells." June 19, 2021. https://www.cellphysiolbiochem.com/Articles/000378/.

Korsunsky, Ilya, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-Ru Loh, and Soumya Raychaudhuri. 2019. “Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony.” Nature Methods 16 (12): 1289–96.

Li, Nan, and Harvest F. Gu. 2022. “Genetic and Biological Effects of SLC12A3, a Sodium-Chloride Cotransporter, in Gitelman Syndrome and Diabetic Kidney Disease.” Frontiers in Genetics 13 (May):799224.

Lindström, Nils O., Guilherme De Sena Brandine, Tracy Tran, Andrew Ransick, Gio Suh, Jinjin Guo, Albert D. Kim, et al. 2018. “Progressive Recruitment of Mesenchymal Progenitors Reveals a Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis.” Developmental Cell 45 (5): 651–60.e4.

Lindström, Nils O., Jill A. McMahon, Jinjin Guo, Tracy Tran, Qiuyu Guo, Elisabeth Rutledge, Riana K. Parvez, et al. 2018. “Conserved and Divergent Features of Human and Mouse Kidney Organogenesis.” Journal of the American Society of Nephrology: JASN 29 (3): 785–805.

Lindström, Nils O., Rachel Sealfon, Xi Chen, Riana K. Parvez, Andrew Ransick, Guilherme De Sena Brandine, Jinjin Guo, et al. 2021. “Spatial Transcriptional Mapping of the Human Nephrogenic Program.” Developmental Cell 56 (16): 2381–98.e6.

Liu, Danli, Qiujia Li, Jing Luo, Qitong Huang, and Yubo Zhang. 2023. “An SPRI Beads-Based DNA Purification Strategy for Flexibility and Cost-Effectiveness.” BMC Genomics 24 (1): 125.

Ma, Sai, Bing Zhang, Lindsay M. LaFave, Andrew S. Earl, Zachary Chiang, Yan Hu, Jiarui Ding, et al. 2020. “Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin.” Cell 183 (4): 1103–16.e20.

McMahon, Andrew P. 2016. “Development of the Mammalian Kidney.” Edited by Paul M. Wassarman. Current Topics in Developmental Biology 117 (January):31–64.

Menon, Rajasree, Edgar A. Otto, Austin Kokoruda, Jian Zhou, Zidong Zhang, Euisik Yoon, Yu-Chih Chen, et al. 2018. “Single-Cell Analysis of Progenitor Cell Dynamics and Lineage Specification in the Human Fetal Kidney.” Development (Cambridge, England) 145 (16): dev164038.

Miao, Zhen, Michael S. Balzer, Ziyuan Ma, Hongbo Liu, Junnan Wu, Rojesh Shrestha, Tamas Aranyi, et al. 2021. “Single Cell Regulatory Landscape of the Mouse Kidney Highlights Cellular Differentiation Programs and Disease Targets.” Nature Communications 12 (1): 2277.

Nielsen, Søren, Tae-Hwan Kwon, Robert A. Fenton, and Jeppe Prætorious. 2012. “Anatomy of the Kidney.” In Brenner and Rector’s The Kidney, 31–93. Elsevier.

Oxburgh, Leif, Thomas J. Carroll, Ondine Cleaver, Daniel R. Gossett, Deborah K. Hoshizaki, Jeffrey A. Hubbell, Benjamin D. Humphreys, et al. 2017. “(Re)Building a Kidney.” Journal of the American Society of Nephrology: JASN 28 (5): 1370–78.

Polonsky, Michal, Louisa M. S. Gerhardt, Jina Yun, Kari Koppitch, Katsuya Lex Colón, Henry Amrhein, Barbara Wold, et al. 2024. “Spatial Transcriptomics Defines Injury Specific Microenvironments and Cellular Interactions in Kidney Regeneration and Disease.” Nature Communications 15 (1): 7010.

Satpathy, Ansuman T., Jeffrey M. Granja, Kathryn E. Yost, Yanyan Qi, Francesca Meschi, Geoffrey P. McDermott, Brett N. Olsen, et al. 2019. “Massively Parallel Single-Cell Chromatin Landscapes of Human Immune Cell Development and Intratumoral T Cell Exhaustion.” Nature Biotechnology 37 (8): 925–36.

Stringer, Carsen, and Marius Pachitariu. 2025. “Cellpose3: One-Click Image Restoration for Improved Cellular Segmentation.” Nature Methods 22 (3): 592–99.

Takei, Yodai, Yujing Yang, Jonathan White, Isabel N. Goronzy, Jina Yun, Meera Prasad, Lincoln J. Ombelets, et al. 2025. “Spatial Multi-Omics Reveals Cell-Type-Specific Nuclear Compartments.” Nature, April. https://doi.org/10.1038/s41586-025-08838-x.

Takei, Yodai, Shiwei Zheng, Jina Yun, Sheel Shah, Nico Pierson, Jonathan White, Simone Schindler, Carsten H. Tischbirek, Guo-Cheng Yuan, and Long Cai. 2021. “Single-Cell Nuclear Architecture across Cell Types in the Mouse Brain.” Science (New York, N.Y.) 374 (6567): 586–94.

Tran, Tracy, Cheng Jack Song, Trang Nguyen, Shun-Yang Cheng, Jill A. McMahon, Rui Yang, Qiuyu Guo, et al. 2022. “A Scalable Organoid Model of Human Autosomal Dominant Polycystic Kidney Disease for Disease Mechanism and Drug Discovery.” Cell Stem Cell 29 (7): 1083–1101.e7.

Wu, Hongwei, Fanna Liu, Yu Shangguan, Yane Yang, Wei Shi, Wenlong Hu, Zhipeng Zeng, et al. 2022. “Integrating Spatial Transcriptomics with Single-Cell Transcriptomics Reveals a Spatiotemporal Gene Landscape of the Human Developing Kidney.” Cell & Bioscience 12 (1): 80.

You, Yue, Luyi Tian, Shian Su, Xueyi Dong, Jafar S. Jabbari, Peter F. Hickey, and Matthew E. Ritchie. 2021. “Benchmarking UMI-Based Single-Cell RNA-Seq Preprocessing Workflows.” Genome Biology 22 (1): 339.

Chapter 4

CONCLUSION

4.1 FUTURE DIRECTIONS

Spatial technology has always fascinated me for its versatility in measuring biological systems at various scales. This thesis introduces two projects that demonstrate this versatility: In Chapter 2, we use spatial multi-omics to investigate the mouse cerebellum, capturing chromatin organization at single-cell resolution in a cell-type-specific manner. In Chapter 3, we profile a human fetal kidney section using seqFISH, integrating with paired single-cell sequencing data, emphasizing tissue-level cellular interactions. Both projects aim to reveal cellular diversity: one by focusing on subnuclear level resolution, and the other on cell-cell interactions across large sections, demonstrating the power of spatial technologies, such as seqFISH, to address varied research needs, from subnuclear to tissue-level scales. While spatial technologies like seqFISH enable the study of cell function in its native environment, they typically lack the whole-transcriptome or whole-genome coverage at single-base resolution provided by single-cell sequencing methods. In studies targeting over 10,000 genes or genomic loci, seqFISH's detection efficiency for individual genes diminishes due to massive multiplexing (Eng et al. 2019; Takei et al. 2025). To overcome this limitation, a key future direction is to enhance the number of transcripts or genomic loci detectable in a single experiment.

Signal amplification is critical for achieving brighter signals with a low signal-to-noise ratio, which shortens exposure times, thus allowing imaging of large areas and more rounds of hybridizations for greater multiplexing. Techniques like branched DNA (Xia et al. 2019), Rolling Circle Amplification (RCA) (Wang et al. 2018), Hybridization Chain Reaction (HCR) (Shah et al. 2016; Gandin et al. 2025), ClampFISH (Dardani et al. 2022), and in vitro transcription (Zombie) (Askary et al. 2020; Kudo et al. 2024) have been explored in various studies for multiplexing. However, each has limitations: branched DNA can be difficult to permeabilize into tissues, RCA and HCR can suffer from non-specific amplification, and

ClampFISH requires precise probe design. No single method is ideal, and further investigation is needed to optimize amplification strategies. Developing more efficient and error-tolerant barcoding schemes is another important aspect for in situ hybridization based methods like seqFISH and MERFISH. For in situ sequencing based methods, while they can theoretically offer a vast number of unique barcodes by extending the barcode sequence length, they face challenges due to signal crowding (Chang et al. 2023; Lee et al. 2022). In each sequencing cycle, transcripts are labeled with one of four fluorophores corresponding to the bases A, T, C, or G, and overlapping fluorescent signals from neighboring transcripts makes decoding impossible. Combining those methods with super-resolution or expansion microscopy could be the solution for resolving overlapping transcripts.

In our spatial multi-omics approach described in Chapter 2, imaging-based chromatin profiling measures the spatial proximity of DNA loci and subnuclear markers, rather than their direct molecular interactions, with a practical genomic resolution limited by the diffraction limit of conventional fluorescence microscopy to approximately 200–300 nm, corresponding to roughly 100–200 kb on the genomic scale. This constraint obscures finer-scale chromatin interactions, typically occurring at 10-20 kb (Greenwald et al. 2019). To address this, expansion microscopy and super-resolution microscopy can enhance immunofluorescence staining of chromatin marks, increasing the detection resolution to approximately 10-kb genomic resolution (Woodworth et al. 2021). It is also possible to directly detect the molecular interactions between histone modifications and specific genomic loci. For instance, Epigenomic MERFISH utilizes antibody-directed Tn5 tagmentation to spatially profile genomic loci associated with the H3K4me3 modification (Lu, Ang, and Zhuang 2022). Additionally, in situ proximity ligation assays (isPLA), traditionally used for detecting protein-protein interactions, could be adapted for DNA-protein interactions. In this method, a FISH probe targets a specific genomic locus, such as an enhancer or promoter, while an oligo-conjugated antibody binds to a histone modification. The oligonucleotides from both side can only be ligated and amplified when the genomic locus and targeted histone modification are in close proximity (Fredriksson et al. 2002; Sharanek et al. 2022). RIBOmap used similar strategy to detect ribosome-mRNA

interactions (Zeng et al. 2023). Alternatively, Fluorescence Resonance Energy Transfer (FRET) microscopy could detect those interactions by labeling DNA and histone modifications with donor and acceptor fluorophores that exhibit energy transfer when they are within 10nm proximity (Cremazy et al. 2005). Spatially resolving of finer scale of chromatin profiling could significantly deepen our understanding of cell-type-specific gene expression regulation within complex tissues.

4.2 REFERENCES

Askary, Amjad, Luis Sanchez-Guardado, James M. Linton, Duncan M. Chadly, Mark W. Budde, Long Cai, Carlos Lois, and Michael B. Elowitz. 2020. “In Situ Readout of DNA Barcodes and Single Base Edits Facilitated by in Vitro Transcription.” Nature Biotechnology 38 (1): 66–75.

Chang, Tianyi, Wuji Han, Mengcheng Jiang, Jizhou Li, Zhizhao Liao, Mingchuan Tang, Jianyun Zhang, et al. 2023. “Rapid and Signal Crowdedness-Robust in Situ Sequencing through Hybrid Block Coding.” Proceedings of the National Academy of Sciences of the United States of America 120 (47): e2309227120.

Cremazy, Frédéric G. E., Erik M. M. Manders, Philippe I. H. Bastiaens, Gertjan Kramer, Gordon L. Hager, Erik B. van Munster, Pernette J. Verschure, Theodorusw J. Gadella Jr, and Roel van Driel. 2005. “Imaging in Situ Protein-DNA Interactions in the Cell Nucleus Using FRET-FLIM.” Experimental Cell Research 309 (2): 390–96.

Dardani, Ian, Benjamin L. Emert, Yogesh Goyal, Connie L. Jiang, Amanpreet Kaur, Jasmine Lee, Sara H. Rouhanifard, et al. 2022. “ClampFISH 2.0 Enables Rapid, Scalable Amplified RNA Detection in Situ.” Nature Methods 19 (11): 1403–10.

Eng, Chee-Huat Linus, Michael Lawson, Qian Zhu, Ruben Dries, Noushin Koulena, Yodai Takei, Jina Yun, et al. 2019. “Transcriptome-Scale Super-Resolved Imaging in Tissues by RNA seqFISH.” Nature 568 (7751): 235–39.

Fredriksson, Simon, Mats Gullberg, Jonas Jarvius, Charlotta Olsson, Kristian Pietras, Sigrún Margrét Gústafsdóttir, Arne Ostman, and Ulf Landegren. 2002. “Protein Detection Using Proximity-Dependent DNA Ligation Assays.” Nature Biotechnology 20 (5): 473–77.

Gandin, Valentina, Jun Kim, Liang-Zhong Yang, Yumin Lian, Takashi Kawase, Amy Hu, Konrad Rokicki, et al. 2025. “Deep-Tissue Transcriptomics and Subcellular Imaging at High Spatial Resolution.” Science (New York, N.Y.) 388 (6744): eadq2084.

Greenwald, William W., He Li, Paola Benaglio, David Jakubosky, Hiroko Matsui, Anthony Schmitt, Siddarth Selvaraj, et al. 2019. “Subtle Changes in Chromatin Loop Contact Propensity Are Associated with Differential Gene Regulation and Expression.” Nature Communications 10 (1): 1054.

Kudo, Takamasa, Ana M. Meireles, Reuben Moncada, Yushu Chen, Ping Wu, Joshua Gould, Xiaoyu Hu, et al. 2024. “Multiplexed, Image-Based Pooled Screens in Primary Cells and Tissues with PerturbView.” Nature Biotechnology, October, 1–10.

Lee, Hower, Sergio Marco Salas, Daniel Gyllborg, and Mats Nilsson. 2022. “Direct RNA Targeted in Situ Sequencing for Transcriptomic Profiling in Tissue.” Scientific Reports 12 (1): 7976.

Lu, Tian, Cheen Euong Ang, and Xiaowei Zhuang. 2022. “Spatially Resolved Epigenomic Profiling of Single Cells in Complex Tissues.” Cell 185 (23): 4448–64.e17.

Shah, Sheel, Eric Lubeck, Maayan Schwarzkopf, Ting-Fang He, Alon Greenbaum, Chang Ho Sohn, Antti Lignell, et al. 2016. “Single-Molecule RNA Detection at Depth by Hybridization Chain Reaction and Tissue Hydrogel Embedding and Clearing.” Development (Cambridge, England) 143 (15): 2862–67.

Sharanek, Ahmad, Laura Raco, Vahab D. Soleimani, and Arezu Jahani-Asl. 2022. “In Situ Detection of Protein-Protein Interaction by Proximity Ligation Assay in Patient Derived Brain Tumor Stem Cells.” STAR Protocols 3 (3): 101554.

Takei, Yodai, Yujing Yang, Jonathan White, Isabel N. Goronzy, Jina Yun, Meera Prasad, Lincoln J. Ombelets, et al. 2025. “Spatial Multi-Omics Reveals Cell-Type-Specific Nuclear Compartments.” Nature, April. https://doi.org/10.1038/s41586-025-08838-x.

Wang, Xiao, William E. Allen, Matthew A. Wright, Emily L. Sylwestrak, Nikolay Samusik, Sam Vesuna, Kathryn Evans, et al. 2018. “Three-Dimensional Intact-Tissue Sequencing of Single-Cell Transcriptional States.” Science (New York, N.Y.) 361 (6400). https://doi.org/10.1126/science.aat5691.

Woodworth, Marcus A., Kenneth K. H. Ng, Aaron R. Halpern, Nicholas A. Pease, Phuc H. B. Nguyen, Hao Yuan Kueh, and Joshua C. Vaughan. 2021. “Multiplexed Single-Cell Profiling of Chromatin States at Genomic Loci by Expansion Microscopy.” Nucleic Acids Research 49 (14): e82.

Xia, Chenglong, Hazen P. Babcock, Jeffrey R. Moffitt, and Xiaowei Zhuang. 2019. “Multiplexed Detection of RNA Using MERFISH and Branched DNA Amplification.” Scientific Reports 9 (1): 7721.

Zeng, Hu, Jiahao Huang, Jingyi Ren, Connie Kangni Wang, Zefang Tang, Haowen Zhou, Yiming Zhou, et al. 2023. “Spatially Resolved Single-Cell Translatomics at Molecular Resolution.” Science (New York, N.Y.) 380 (6652): eadd3067.