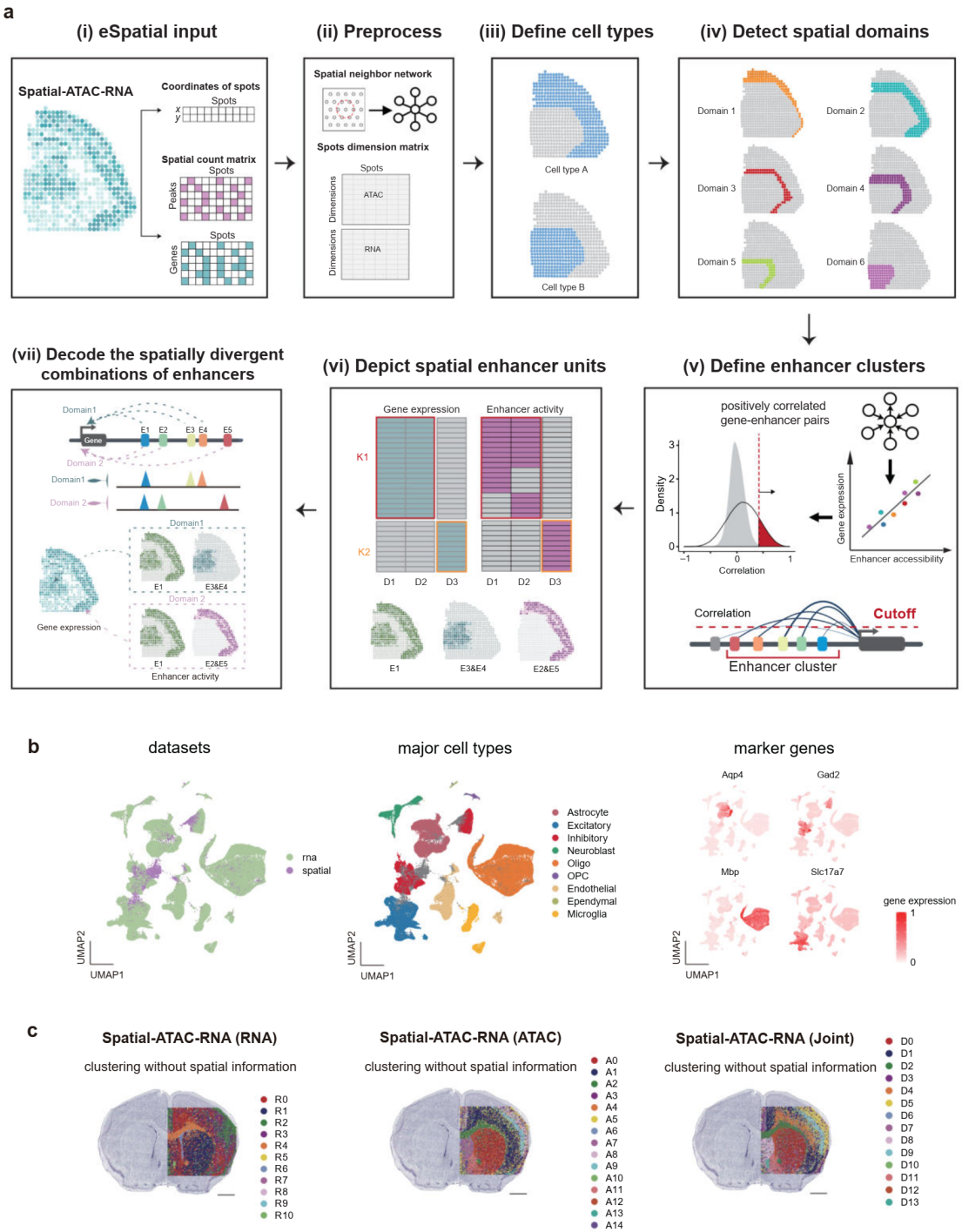
DIVERGENT COMBINATIONS OF ENHANCERS ENCODE SPATIAL GENE EXPRESSION

SUPPLEMENTARY INFOMATION

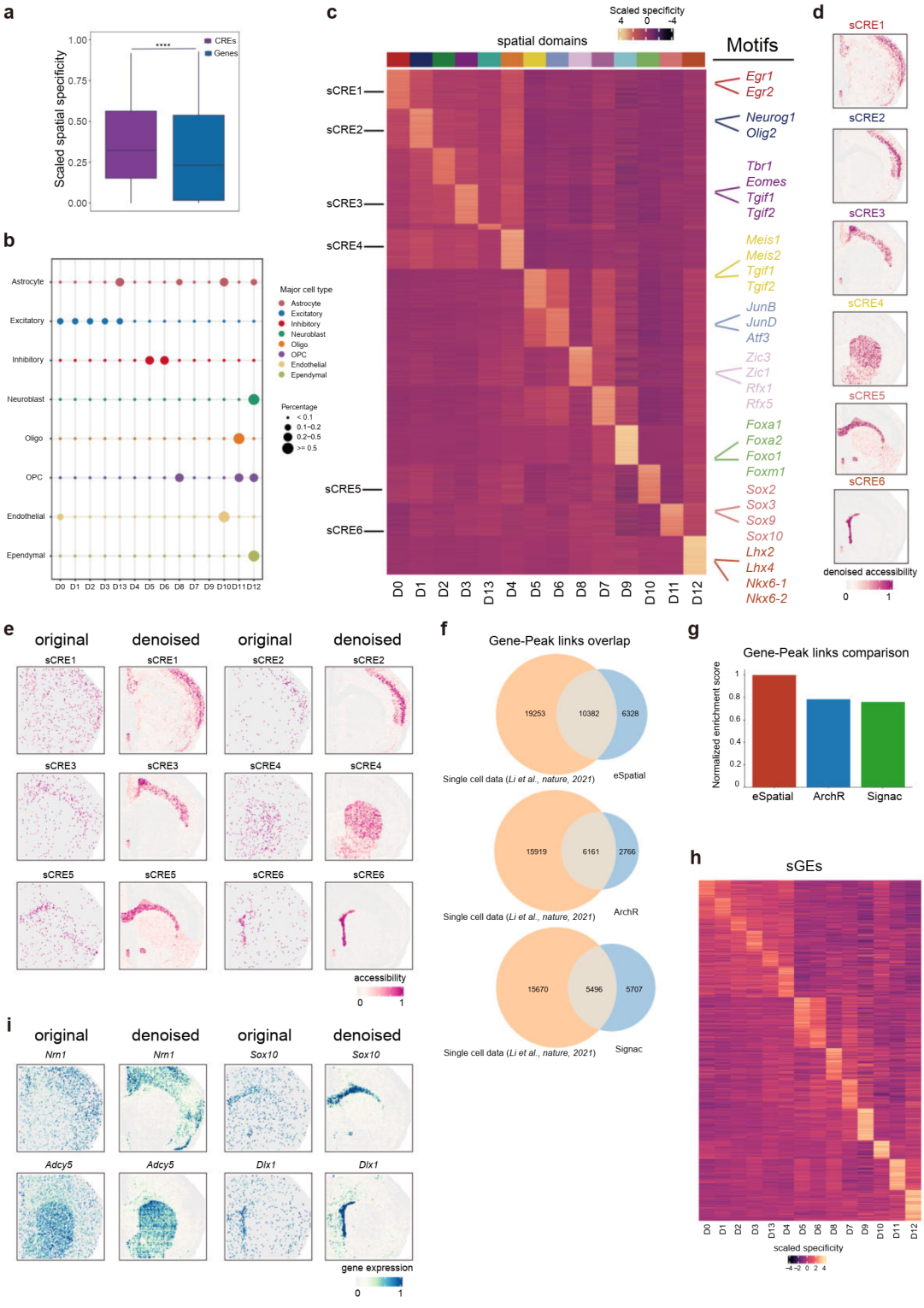
Supplementary Figures



Supplementary Fig. 1 Overview of eSpatial framework and cell clustering analysis of the P22 mouse brain.

(**a**) Overview of eSpatial framework: (**i**) eSpatial input: spatial count matrices (peak-by-spot matrix and gene-by-spot matrix) and spatial coordinates of spots. (**ii**) Preprocess: eSpatial reduces the dimensions of the gene-by-spot matrix

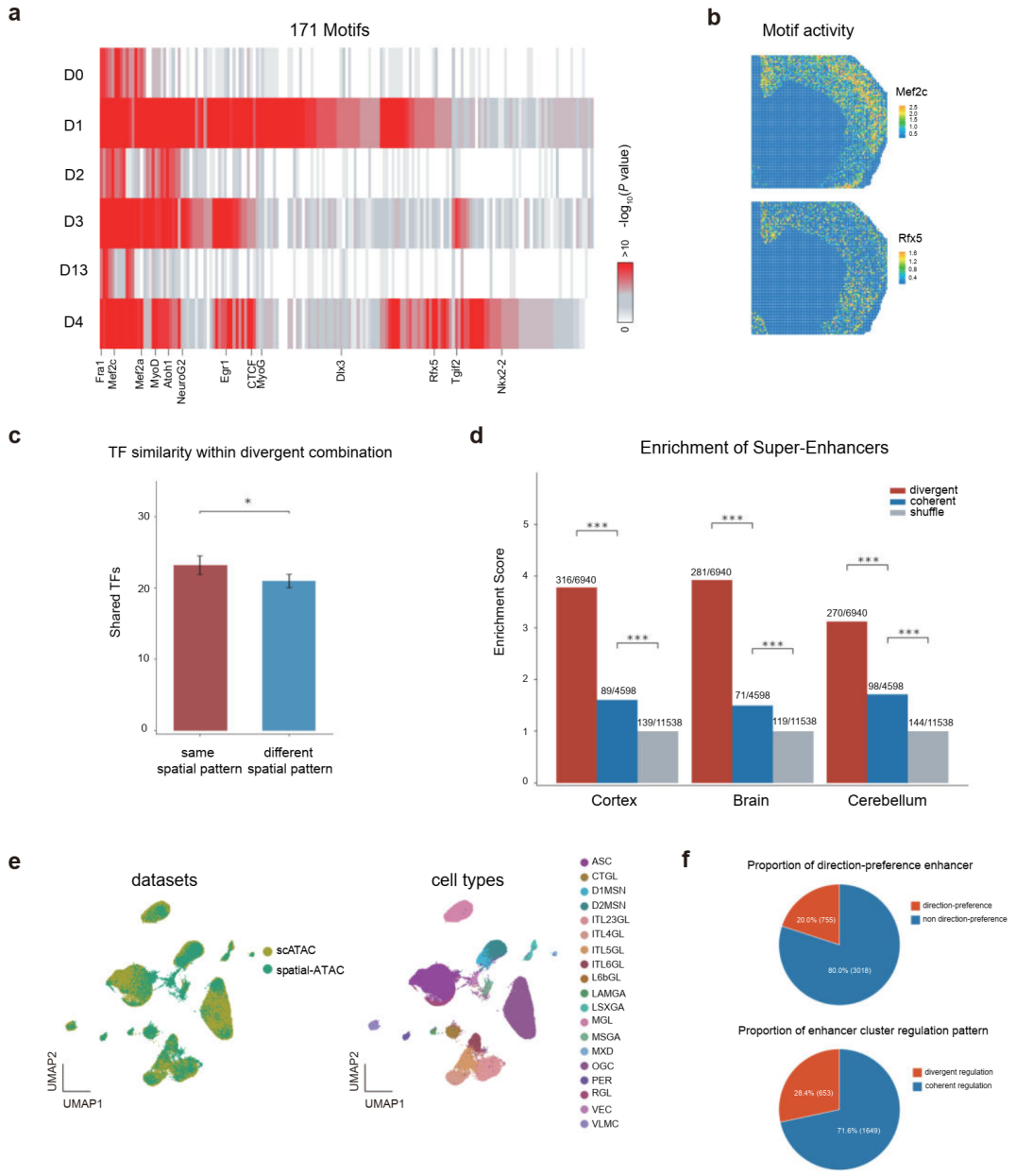
and peak-by-spot matrix separately. It converts the spatial coordinates into an undirected neighbor network based on Euclidean distance between two spots. (**iii**) Cell types: eSpatial integrates spatial RNA data with scRNA-seq data using Seurat to identify cell types across spatial locations. (**iv**) Spatial domains: eSpatial utilizes STAGATE for spatial clustering of the dimension matrix and spatial neighbor network identification. (**v**) Enhancer clusters: eSpatial links enhancers to their target genes by correlating enhancer chromatin accessibility with gene expression across spatial neighbor networks. (**vi**) Spatial enhancer units: eSpatial identifies spatial gene modules with similar expression patterns and clusters linked enhancers for genes in each module based on binary spatial specificity, defining spatial enhancer units based on the spatial patterns of enhancers. (**vii**) Spatial enhancer code: genes expressed across multiple spatial domains are controlled by the combinations of divergent enhancer units. (**b**) UMAP of P22 mouse brain for spatial-RNA after integration with reference single-cell RNA-seq mouse brain atlas data [1], colored by datasets (left) and major cell types (right). (**c**) Spatial distribution of RNA (left), ATAC (middle), and joint (right) clusters using spatial ATAC-RNA-seq data of mouse brain. The clustering process was without spatial information, distinguished from those in **Fig. 1**. Pixel size, 20 μm; scale bars, 1 mm.



Supplementary Fig. 2 Spatial-specific genes and *cis*-regulatory elements in the P22 mouse brain. Related to Fig. 1.

(**a**) Boxplot showing the spatial specificity of gene expression and chromatin accessibility of CREs. *p*-values were calculated using one-tailed Student's *t*-

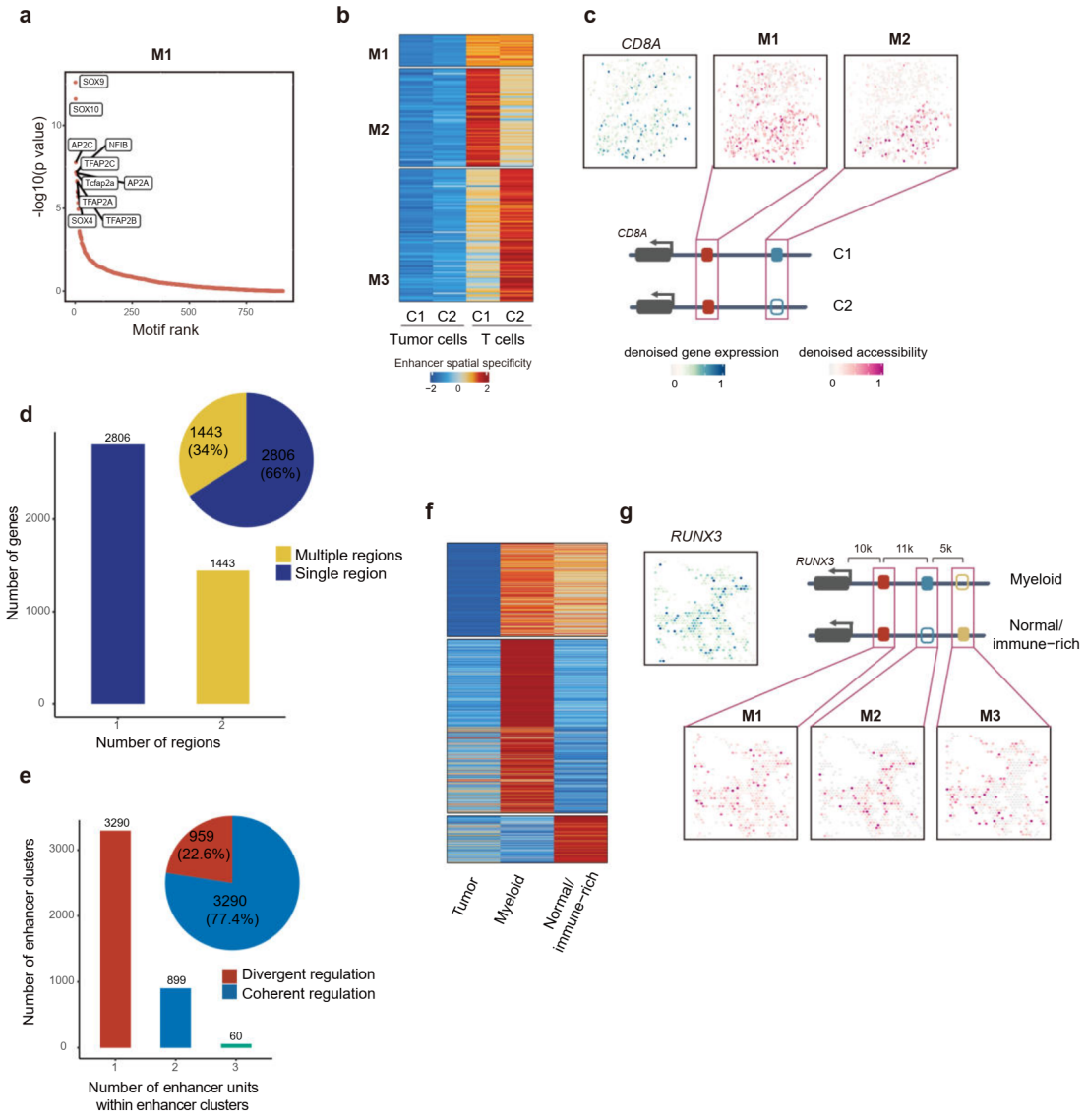
test. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; *n.s.*, not significant. (**b**) The proportions of spatial domains in each major cell type. Spot size indicates the percentage of each spatial domain in major cell types. (**c**) Heatmap showing scaled spatial specificity for sCREs (rows) across various spatial domains (columns). Representative enriched motifs are listed for selected spatial domains. (**d**) Spatial mapping of denoised chromatin accessibility for representative sCREs across different spatial domains. (**e**) Spatial mapping of the original (left) and denoised (right) chromatin accessibility for representative enhancers across different spatial domains. (**f**) Venn diagram illustrating the overlap between gene-enhancer links identified by eSpatial (as shown in **Fig. 1f**), ArchR [2], and Signac [3] on the same spatial epigenomic data, using the publicly available gene-enhancer links from mouse brain as the reference [4]. The numbers within each section of the Venn diagrams represent the count of unique and overlapping gene-peak links. (**g**) Bar plot showing the normalized enrichment scores of gene-enhancer links identified by eSpatial, ArchR, and Signac, relative to a mouse brain reference dataset. The enrichment scores are derived from Fisher's exact test, which have been normalized to allow for comparison. (**h**) Heatmap showing scaled spatial specificity for spatial-specific genes (sGEs) across different spatial domains. (**i**) Spatial mapping of original (left) and denoised (right) gene expression for representative genes across different spatial domains. Source data are provided as a Source Data file.



Supplementary Fig. 3 Spatial-specific *cis*-regulatory elements in the P22 mouse brain. Related to Fig. 2.

(**a**) Enrichment of known transcription factor (TF) motifs in distinct gene-enhancer subgroups regulating K1 genes in **Fig. 1f**. Known motifs calculate by HOMER [5]. (**b**) Spatial mapping of deviation scores for selected TF motifs. (**c**) Transcription factor (TF) similarity within divergent enhancer combinations across spatial domains. Statistical significance was tested (*$p < 0.05$, one-tailed paired *t* test). (**d**) Bar plot showing the enrichment of super-enhancers in divergent and coherent enhancer clusters from **Fig. 2d**, with shuffled enhancers

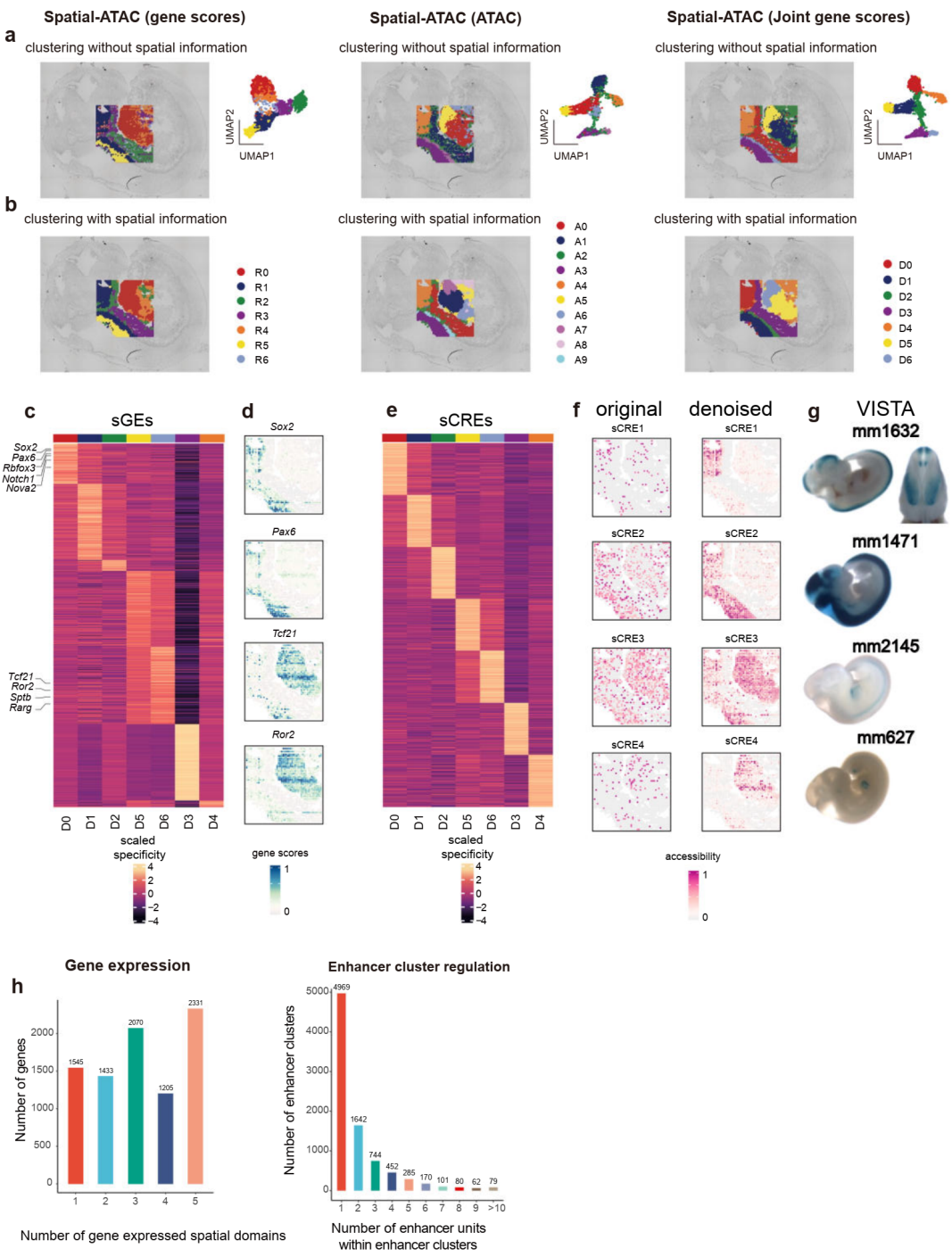
as a control background. The brain-related super-enhancers data was sourced from a comprehensive database [6]. Statistical significance was tested (***p < 0.001, binomial test). (**e**) UMAP of P22 mouse brain for spatial-ATAC after integration with reference single-cell ATAC-seq mouse brain atlas data [4], colored by datasets (left) and cell types (right). (**f**) The pie charts show the proportion of - enhancers with directionality biases of accessibility (top) and the corresponding enhancer cluster regulation patterns (bottom) classified by chromatin accessibility directionality in **Fig. 2g**. In the top panel, 20.0% (755) of putative enhancers show significant spatial biases ($|R| \geq 0.2$, $P < 0.05$; orange), while 80.0% (3,018) show non-directional preference (blue). In the bottom panel, 28.4% (653) of enhancer clusters are associated with directional enhancers, demonstrating divergent regulation (orange), compared with 71.6% (1,649) showing non directional preference, indicating coherent regulation (blue).



Supplementary Fig. 4 Spatial enhancer code of human tumors, Related to Fig. 3.

(**a**) Motif enrichment rank of enhancers with spatial patterns M1 from **Fig. 3f**. Top 10 motifs are highlighted. *P* values by a one-sided hypergeometric test. (**b**) Heatmap displaying the spatial patterns of enhancers regulating K4 genes in **Fig. 3c**. These enhancers were clustered into 3 patterns (M1-M3, rows) based on their binarized spatial specificity in two compartments. (**c**) Spatial mapping of denoised *CD8A* gene expression and denoised chromatin accessibility of three *CD8A* enhancers from two distinct enhancer units (M1 and M2). Schematic representation of the location of *CD8A* gene and *CD8A* enhancers.

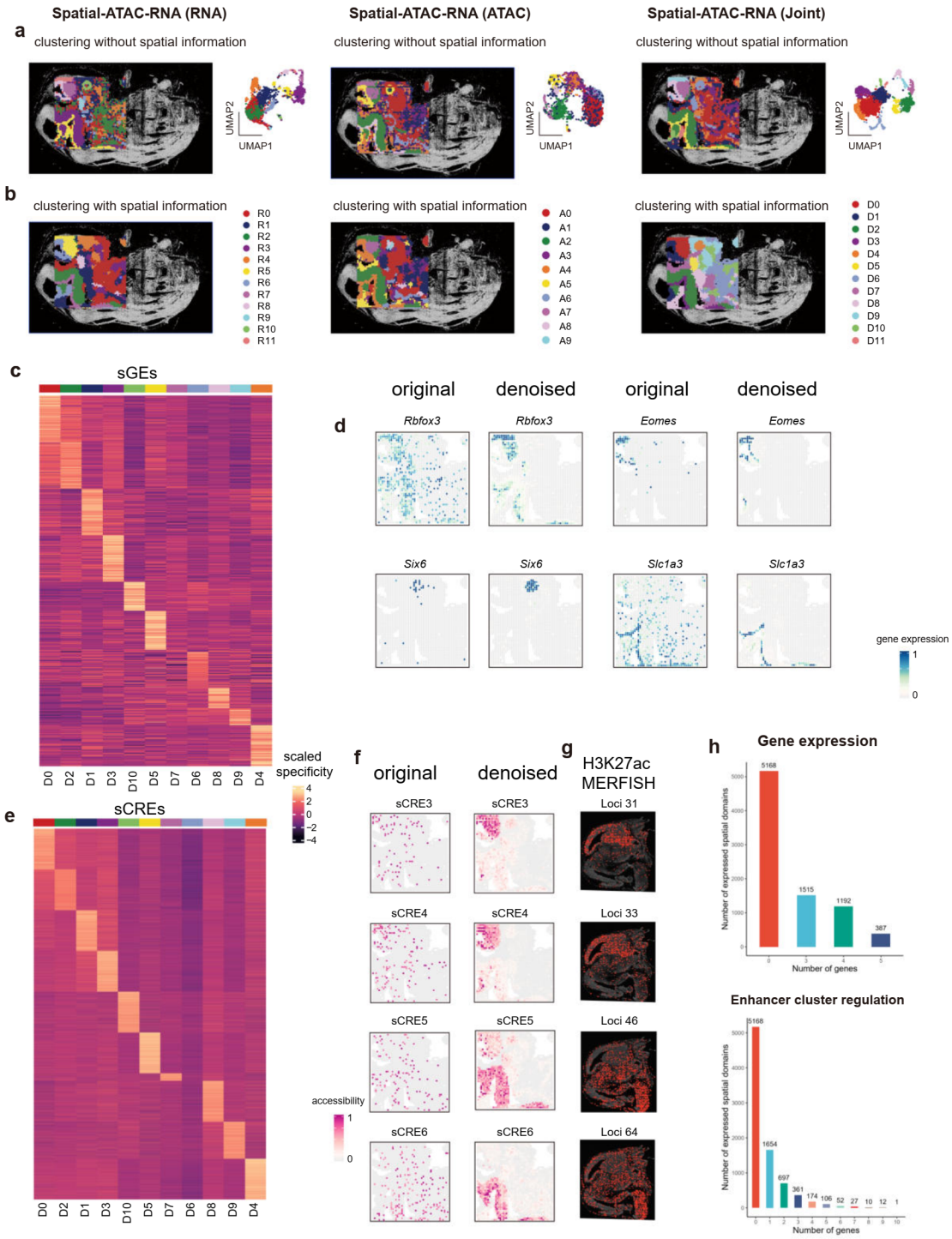
(**d**) Bar plot illustrating the number of regions in which genes are expressed. Pie chart showing the proportion of genes expressed across multiple regions (>=2) versus those specific to a single region. (**e**) Bar plot showing the number of enhancer clusters (*y*-axis) containing enhancers from various numbers of enhancer units (*x*-axis). Pie chart showing the proportion of divergent regulation (>=2 enhancer units within a cluster) versus coherent regulation (only 1 enhancer pattern). (**f**) Heatmap displaying the spatial patterns of enhancers regulating K7 genes in **Fig. 3k**. These enhancers were clustered into 3 patterns (M1-M3, rows) based on their binarized spatial specificity. (**g**) Spatial mapping of denoised *RUNX3* gene expression and denoised chromatin accessibility of three *RUNX3* enhancers from three distinct enhancer units (M1, M2, and M3) in the myeloid and normal/immune-rich regions. Schematic representation of the location of *CD8A* gene and *CD8A* enhancers.



Supplementary Fig. 5 Spatial-specific genes and *cis*-regulatory elements in e11 mouse embryos. Related to Fig. 4.

(**a-b**) Spatial distribution and UMAP of RNA, ATAC, and joint ATAC-RNA clusters in spatial ATAC-RNA-seq in e11 mouse embryos. The clustering process was without (**a**) or with (**b**) spatial information. (**c**) Heatmap showing scaled spatial specificity for spatial-specific genes (sGEs) across different

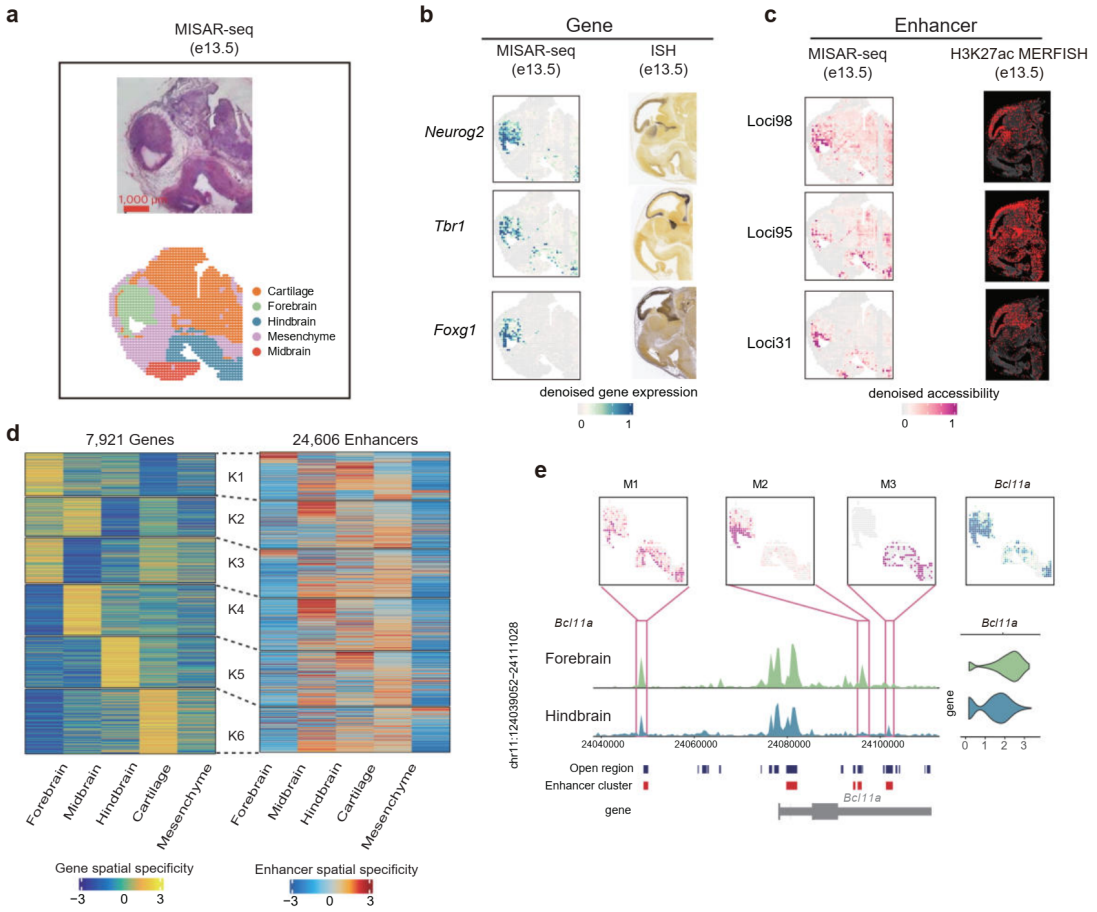
spatial domains. (**d**) Spatial mapping of gene activity scores for representative genes across different spatial domains. (**e**) Heatmap showing scaled spatial specificity for spatial-specific *cis*-regulatory elements (sCREs) across different spatial domains. (**f-g**) Spatial mapping of original (left) and denoised (right) chromatin accessibility (**f**), and VISTA enhancer reporter expression (**g**) for representative enhancers across different spatial domains. (**h**) Bar plot showing the number of spatial domains in which genes are expressed (left) and the number of enhancer clusters from enhancers within different numbers of enhancer units (right). Source data are provided as a Source Data file.



Supplementary Fig. 6 Spatial-specific genes and *cis*-regulatory elements in e13 mouse embryos. Related to Fig. 5.

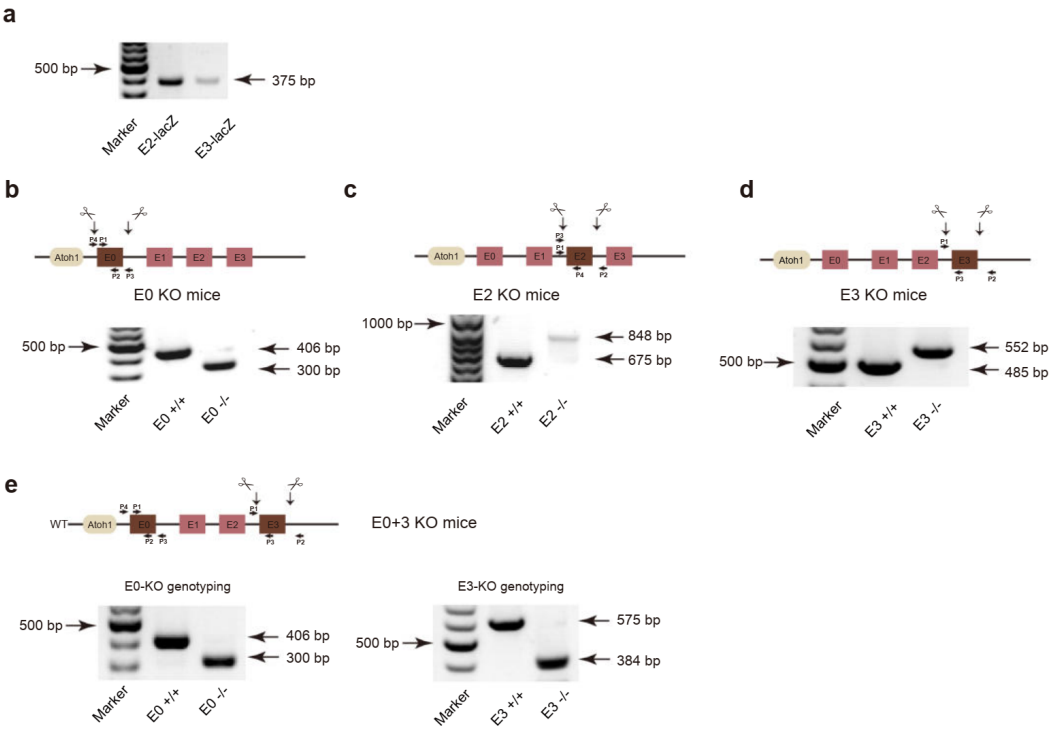
(**a-b**) Spatial distribution and UMAP of RNA, ATAC, and joint ATAC-RNA clusters in spatial ATAC-RNA-seq of e13 mouse embryos. The clustering process was without (**a**) or with (**b**) spatial information. (**c**) Heatmap showing scaled spatial specificity for spatial-specific genes (sGEs) across different

spatial domains. (**d**) Spatial mapping of original (left) and denoised (right) gene expression for representative genes across different spatial domains. (**e**) Heatmap showing scaled spatial specificity for spatial-specific *cis*-regulatory elements (sCREs) across different spatial domains. (**f-g**) Spatial mapping of original (left) and denoised (right) chromatin accessibility (**f**), and epigenomic MERFISH images of H3K27ac signals (**g**) for representative enhancers across different spatial domains. (**h**) Bar plot showing the number of spatial domains in which genes are expressed (left) and the number of enhancer clusters containing different numbers of enhancer units (right). Source data are provided as a Source Data file.



Supplementary Fig. 7 Spatial enhancer code in e13.5 mouse brain.

(**a**) H&E images of the e13.5 mouse brain (top). Spatial mapping of the spatial domains. Spatial data was from [7]. (**b**) Spatial mapping of denoised gene expression (left) and Allen RNA ISH images (right) for *Neurog2*, *Tbr1*, and *Foxg1*. (**c**) Spatial mapping of denoised chromatin accessibility (left) and epigenomic MERFISH images of H3K27ac signals (right) for the enhancers controlling *Neurog2*, *Tbr1*, and *Foxg1*. (**d**) Heatmap showing gene expression (left) and chromatin accessibility (right) for 24,606 gene-enhancer pairs. The gene expression was clustered using k-means clustering (k = 6). (**e**) Genome tracks showing the chromatin accessibility, peak site (labeled by open region, enhancer cluster), gene tracks (left), and gene expression (right) for *Bcl11a* in the forebrain and hindbrain. Spatial mapping of denoised chromatin accessibility of three representative enhancers from distinct enhancer units (M1, M2, and M3) and gene expression (top).



Supplementary Fig. 8 Validation of *Atoh1* spatial enhancer code, Related to Fig. 6.

(**a**) Agarose gel showing genotyping PCRs with E0-lacZ, E2-lacZ, and E3-lacZ genomic DNA. (**b-e**) Genotyping PCRs of E0-KO, E2-KO, E3-KO, and E0+3-DKO. Top: Strategy for creating enhancer KO mice using CRISPR/Cas9. Bottom: Agarose gel showing genotyping PCRs with E0-KO (**b**), E2-KO (**c**), E3-KO (**d**) and E0+3-DKO (**e**) from +/+ (WT), and -/- (KO) genomic DNA. E0 knockout location: Chr6:64,733,006-64,735,029. E2 knockout location: Chr6:64,773,213-64,774,102. E3 knockout location: Chr6:64,797,552-64,800,070. Double mutant E0+3 knockout location: Chr6:64,733,006-64,735,029 and Chr6:64,797,193-64,800,225.

Supplementary Tables

Supplementary Table 1. A list of all used datasets and accession numbers. (Related to Figs 1-6)

Dataset	Data Type	Details	Link
P22 mouse brain	spatial-ATAC-RNA	GSM6753043_MouseBrain_20um_100barcodes_ATAC_matrix.tsv.gz GSM6758285_MouseBrain_20um_100barcodes_ATAC_fragments.tsv.gz GSM6753043_MouseBrain_20um_100barcodes_ATAC_spatial.tar.gz	Gene Expression Omnibus with accession code GSE205055
e11 mouse embryo	spatial-ATAC	GSM6043255_ME11_20um_fragments.tsv.gz GSM6043255_ME11_20um_spatial.tar.gz	Gene Expression Omnibus (GEO) with accession code GSE171943
e13 mouse embryo	spatial-ATAC-RNA	GSM6799937_ME13_50um_matrix_merge.tsv.gz GSM6801813_ME13_50um_fragments.tsv.gz GSM6801813_ME13_50um_spatial.tar.gz	Gene Expression Omnibus with accession code GSE205055
e13.5 mouse fetal brain	MISAR-seq	E13_5-S1_raw_feature_bc_matrix.h5 E13_5-S1_filtered_fragments.tsv.bgz E13_5-S1_per_barcode_metrics.csv E13_5-S1-HE.jpg	National Genomics Data Center accession number (OEP003285, www.biosino.org/node/project/detail/OEP003285)
e13.5 mouse fetal brain	H3K27ac MERFISH	H3K27ac_embryo_loci_info.xlsx H3K27ac_embryo_plot	Zenodo (https://doi.org/10.5281/zenodo.7075964)
human melanoma	Slide-tags snRNA+snATAC	HumanMelanomaMultiome_features.tsv.gz HumanMelanomaMultiome_matrix.mtx.gz HumanMelanomaMultiome_barcodes.tsv.gz HumanMelanomaMultiome_atac.c	Broad Institute Single Cell Portal under the following accession numbers SCP2176

		sv HumanMelanomaMultiome_metadata.csv HumanMelanomaMultiome_spatial.csv	
human breast cancer	spatial-ATAC	fragments.tsv.gz 220327_C1_tissue_positions_list.csv metadata_combined.csv	Kindly shared by the authors of the original study
VISTA enhancers	Experimentally validated human and mouse noncoding fragments with gene enhancer activity as assessed in transgenic mice	Use gene, accession number, locus link, genomic position to find enhancer activity as assessed in transgenic mice	https://enhancer.lbl.gov/
Allen Brain Reference Atlases	ISH data	Use Gene Search to find ISH data for a specific gene of interest	https://atlas.brain-map.org/

Supplementary Table 2. List of primer and sgRNA sequences. (Related to Fig. 6)

	primer name	**sequence 5' to 3'**	**product and size**
lacZ	lacZ-F1	TTTCCATGTTGCCACTCGC	KI=375 bp
	lacZ-R1	AACGGCTTGCCGTTCAGCA	
E0-KO	E0-P1	CCCTCTCTCACACCCCATTAAC	WT=406 bp
	E0-P2	AATGTTAGCTGGCTGAGGGC	Het=406; 300 bp
	E0-P3	GGCGTGTGATTGACAAGTGG	KO= 300 bp
	E0-P4	CGTTTTATAGATAGTGAAATGGCAGC	
E2-KO	E2-P2	GCCCTGTTTTCAAGTCCCGA	wt=675bp
	E2-P3	CTGCAGGCTTCCCAGAAGAA	Het=675，848bp
	E2-P4	CAGGGACCTGACAACTCCAC	KO=848bp

E3-KO	E3-P1	GTCTGCAGCAGTGACACAGT	WT=485bp
	E3-P2	GGTCCTACCCACTTCTGGGA	Het=552; 485 bp
	E3-P3	ACCTCCTCAGATTCCCCTCC	KO=552bp
E0+E3 KO	E0-P1	CCCTCTCTCACACCCCATTAAC	WT=406 bp
	E0-P2	AATGTTAGCTGGCTGAGGGC	Het=406; 300 bp
	E0-P3	GGCGTGTGATTGACAAGTGG	KO= 300 bp
	E0-P4	CGTTTTATAGATAGTGAAATGGCAGC	
	E3-P1	cccacgtgttcttatcctct	WT=575 bp
	E3-P2	cagcttggagtcaagtgtct	Het=575; 384 bp
	E3-P3	ggtcctacccacttctggga	KO= 384 bp

Supplementary References

1. Zeisel A*, et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014 e1022 (2018).
2. Granja JM*, et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411 (2021).
3. Stuart T*, et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
4. Li YE*, et al.* An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129-136 (2021).
5. Heinz S*, et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
6. Khan A, Zhang X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res* **44**, D164-171 (2016).
7. Jiang F*, et al.* Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. *Nat Methods* **20**, 1048-1057 (2023).