METHODOLOGY | Article in Press

Joint-Cabernet enables scalable single-cell joint profiling of transcriptomes with DNA methylation and hydroxymethylation in the brain

Yali Bai (白雅丽) · **Tianjiao Yuan** (袁天骄) · **Liuhao Ren** (任刘豪) · **Yuhang Huan** (郇宇航) · **Fa Yang** (杨发) · **Yutong Li** (李雨桐) · **Aicui Zhang** (张爱翠) · **Yueyang Liu** (刘悦洋) · **Tian Tian** (田甜) · **Ningxia Kang** (康宁霞) · **Xuchao Xie** (谢戍超) · **Qiang Wang** (汪强) · **Danni Chen** (陈丹妮) · **Wei Chen** (陈炜) · **Yinghui Zheng** (郑英慧) · **Xiaoliang Sunney Xie** (谢晓亮) · **Yunlong Cao** (曹云龙)

Received: 28 June 2026 / Accepted: 11 November 2026

Abstract

DNA methylation, particularly CpG methylation, is vital to brain development and function, yet conventional bisulfite-based detection cannot distinguish 5-methylcytosine (5mC) from 5-hydroxymethylcytosine (5hmC), which is enriched in neurons. We develop Joint-Cabernet, a bisulfite-free single-cell platform simultaneously profiling 5mC, 5hmC, and transcriptomes at scale. Applying Joint-Cabernet to 47,712 adult mouse brain nuclei, we resolve cell-type-specific hydroxymethylation landscapes and identify elevated 5hmC in active genes and regulatory elements. We further uncover two modes of methylation-associated gene activation: progressive demethylation linked to neurodevelopmental genes and stable hydroxymethylation associated with cellular maintenance. Joint-Cabernet enables high-resolution dissection of gene regulation by 5mC and 5hmC.

Keywords Single-cell multi-omics · 5hmC · 5mC · Transcriptome · Epigenetic regulation · Brain

Background

The mammalian brain is a highly complex organ composed of diverse neuronal and glial cell types, each exhibiting specialized functions shaped by spatiotemporal gene expression programs [1–4]. These programs are tightly regulated by epigenetic mechanisms, including DNA methylation, chromatin accessibility, and three-dimensional genome architecture [5–10]. Among these regulatory layers, the methylation of cytosine at the fifth carbon position (5-methylcytosine, 5mC) has long been recognized as a stable repressive mark, essential for silencing retrotransposons, genomic imprinting, and X-chromosome inactivation. However, the discovery that the Ten-Eleven

Yali Bai, Tianjiao Yuan, Liuhao Ren, Yuhang Huan and Fa Yang contributed equally.

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

https://doi.org/10.1186/s13059-026-04351-0

Published online: 19 November 2026

Translocation (TET) family of dioxygenases can iteratively oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) has fundamentally expanded our understanding of DNA modification dynamics [11–15].

5mC is generally depleted from active regulatory regions, whereas 5hmC is enriched there. This enrichment is particularly evident in the mammalian brain, where 5hmC accumulates in the gene bodies of actively transcribed genes within postmitotic neurons [16–19]. This accumulation suggests that 5hmC is not merely a transient intermediate in the active DNA demethylation pathway but a stable epigenetic mark with distinct regulatory functions, potentially recruiting specific readers such as MeCP2 or MBD3 to modulate chromatin structure and gene expression [16, 20]. The balance between 5mC and 5hmC is critical for neurodevelopment, synaptic plasticity, and memory formation, and its dysregulation has been implicated in neurodegenerative disorders such as Alzheimer's disease and Huntington's disease [21–23]. Consequently, deconvoluting the individual contributions of 5mC and 5hmC is essential for decoding the epigenetic logic of the brain.

Despite the biological significance of these modifications, historical efforts to map the brain's epigenome have been severely hindered by the fundamental technical limitations of conventional bisulfite sequencing. Bisulfite treatment efficiently converts unmodified cytosines to uracil while leaving 5mC intact; however, it reacts identically with 5hmC, rendering 5mC and 5hmC utterly indistinguishable in standard bioinformatic pipelines. This base-level ambiguity has historically obscured the individual contributions of 5mC and 5hmC to epigenetic dynamics [24], forcing researchers to rely on composite signals that mask critical, locus-specific regulatory shifts. While recent innovations in single-cell methylome sequencing have revolutionized the construction of cell-type-specific atlases [5–7, 25], the vast majority of these platforms remain entirely blind to the 5hmC modality, collapsing two biologically opposing signals into a single deceptive metric.

To circumvent these challenges, the field has developed 5hmC sequencing strategies beyond conventional bisulfite readouts, such as TAB-seq [26] and oxBS-seq [27, 28], and more recently has witnessed the emergence of bisulfite-free chemical and enzymatic methods capable of discriminating 5mC and 5hmC at base resolution, such as TAPS [29], ACE-seq [30] and Six-letter seq [31]. However, translating these dual-profiling approaches to true single-cell resolution has exposed significant technical trade-offs regarding throughput, conversion efficiency, genomic coverage, and multi-omic compatibility. For instance, the SIMPLE-seq [32] methodology utilizes orthogonal chemical labeling to detect both modifications, but it suffers from suboptimal conversion rates of less than 90%, necessitating computational corrections and potentially reducing the sensitivity for detecting subtle epigenetic shifts in rare cell types. Alternatively, restriction enzyme-based single-cell platforms such as DARE-SOME [33] leverage methylation-sensitive endonucleases like HpaII and MspI. While highly accurate at the loci they capture, these approaches are intrinsically constrained by their absolute dependence on specific sequence motifs (e.g., CCGG contexts), resulting in the systematic undersampling of high-density CpG islands and vast intergenic regulatory domains. Another recent approach, Joint-snhmC-seq [34], successfully discriminates 5mC and 5hmC marks through bisulfite conversion coupled with APOBEC3A-mediated deamination, but lacks the capability to profile the transcriptome simultaneously. Most importantly, many existing technologies, such as TAB-seq and Six-letter seq, currently cannot be performed on single cells at scale, significantly limiting their usage in deconvoluting complex systems like the brain.

To address these challenges, we developed Joint-Cabernet, a high-throughput, bisulfite-free single-cell multi-omics platform that integrates transcriptomic profiling with the simultaneous detection of 5hmC and 5mC. Building upon our previously established Cabernet chemistry [35], the platform incorporates a highly optimized template-switching barcoding strategy to significantly enhance throughput and uniformity. Applying this method to 47,712 nuclei from adult mouse brains, we resolved long-standing limitations of bisulfite-based methods and generated a DNA hydroxymethylation and methylation atlas of the mouse brain, unveiling the distinct roles of 5hmC and 5mC in shaping brain function and organization.

Results

Technical development of Joint-Cabernet

The structural foundation of the platform relies on the previously established Cabernet methodology [35], which sequentially employs the Tn5 transposome for genomic fragmentation, followed by a non-destructive enzymatic cascade involving TET-mediated oxidation, β-glucosyltransferase (BGT)-mediated glycosylation, and APOBEC3A-mediated deamination. This bisulfite-free approach effectively preserves DNA integrity, leading to vastly superior genome recovery rates compared to highly degradative bisulfite methods. However, the original Cabernet protocol utilized a direct Tn5 dual-barcoding strategy that suffered from severe transposition uniformity issues, strictly precluding its application in high-throughput, plate-based single-cell experiments.

To engineer a highly scalable solution capable of profiling tens of thousands of cells, the platform implements a template-switching-based indexing strategy inspired by the symmetrical strand single-cell combinatorial indexing (s3) approach [36] (Fig. 1A and B). Standard Tn5 tagmentation workflows intrinsically discard up to 50% of the generated genomic fragments because viable sequencing libraries require the stochastic incorporation of two different adapters (forward and reverse) on opposite ends of a single DNA molecule; fragments receiving the same adapter on both ends cannot form viable clusters on the sequencing flow cell. The methodology bypasses this thermodynamic limitation by utilizing universal transposomes armed with a precisely engineered uracil-containing linker and a cytosine-depleted reverse adapter resistant to downstream deamination. During the initial gap-fill extension phase immediately following DNA fragmentation, these engineered uracil residues function as absolute, transient termination sites that forcefully arrest polymerase activity.

Subsequently, a customized secondary oligonucleotide is designed and introduced to facilitate a template-guided extension. This custom oligonucleotide contains three critical functional domains: an inline cellular barcode, a forward sequencing adapter, and a locked nucleic acid (LNA)-modified mosaic end (ME) reverse complement. The strategic incorporation of LNA critically elevates the annealing temperature of the oligonucleotide, decisively suppressing spurious intramolecular self-looping and guaranteeing high-fidelity extension even in complex genomic mixtures. Thermal cycling is iteratively applied to maximize processivity, ensuring near-complete conversion of the fragmented genome into viable, asymmetrically barcoded amplicons. Because

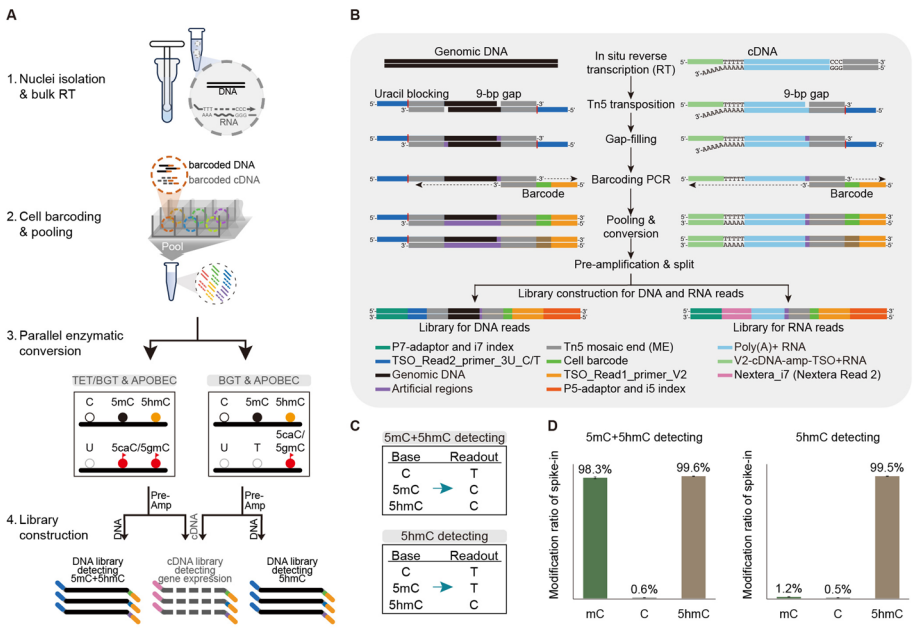


Fig. 1 Joint-Cabernet strategy and validation. **A** Workflow of brain dissection, nuclei preparation, and joint profiling of RNA, 5hmC, and 5mC. **B** Schematic of joint profiling of RNA, 5hmC and 5mC. **C** Decoding single cell 5mC and 5hmC. **D** Conversion rate for 5mC + 5hmC detecting and 5hmC detecting. Error bars represent SD

the uracil-induced termination strategically blocks the forward adapter from fully synthesizing the complementary reverse strand, the protocol eliminates the creation of artifactual fragments that could spuriously incorporate exogenous, synthetically methylated cytosines during amplification. This strict safeguard preserves the fidelity of the endogenous epigenetic map.

To simultaneously profile 5mC and 5hmC in single cells, the uniquely barcoded contents of 96 wells are evenly distributed and pooled for two parallel enzymatic processing channels. The first channel undergoes combined TET2 and BGT treatment to protect both 5mC and 5hmC. The second channel undergoes BGT treatment alone, protecting only endogenous 5hmC while leaving 5mC unprotected. Both channels are then subjected to the potent APOBEC3A deaminase, which selectively converts all unprotected cytosines into uracil. As the template-switching barcode extension incorporates modified nucleotides (containing methylated or hydroxymethylated dCTP), the appended cell barcodes are protected from APOBEC3A-mediated deamination. This bifurcation allows for the highly accurate quantification of 5hmC directly (with a demonstrated 99.5% detection accuracy, the proportion of CpG cytosines called as 5hmC in 5hmC spike-in), and of 5mC + 5hmC (with an accuracy of 98.3%, the proportion of CpG cytosines called as modified cytosines in the pUC19 spike-in), yielding the true 5mC levels via computational subtraction of the 5hmC signal from the combined 5mC + 5hmC channel at single-cell resolution (Fig. 1C and D, Additional file 1: Fig. S1A). It should be noted that the two-channel design does not resolve 5hmC and 5mC + 5hmC on the same DNA molecule, as each molecule is analyzed in only one channel.

To enable the synchronized capture of the transcriptome alongside these epigenetic modalities without physically splitting the cellular lysate, the methodology incorporates a highly specialized reverse transcription (RT) step prior to genomic tagmentation [37]. Bulk reverse transcription is executed in situ utilizing a customized poly-T primer engineered with a distinct reverse adapter that exhibits zero sequence homology with the genomic DNA adapters (Additional file 2: Table S1). This strict sequence divergence ensures that the resulting cDNA fragments and genomic DNA amplicons can be independently targeted and selectively amplified during downstream library preparation. To further insulate the transcriptomic data from genomic contamination, the RT reaction omits standard dCTP and instead uses 5-hydroxymethyl-dCTP (5hm-dCTP) exclusively. This biochemical substitution forces the reverse transcriptase to synthesize fully hydroxymethylated cDNA strands. Given that endogenous non-CpG methylation (mCH) exists at exceptionally low frequencies in the native mammalian genome, this pervasive artificial methylation serves as a robust, highly specific molecular barcode [38]. During bioinformatic processing, any sequencing read exhibiting a hyper-elevated mCH ratio will be instantly classified as an RNA-derived transcript, resolving cross-contamination artifacts between the transcriptomic and epigenomic libraries (Additional file 1: Fig. S1B).

Together, this integrative design generated matched 5hmC, 5mC, and transcriptome datasets from identical cells, resolving cell-type-specific methylome-transcriptome relationship analysis while maintaining single-cell resolution.

DNA methylation and hydroxymethylation landscapes of adult mouse brain

To demonstrate the power of Joint-Cabernet in resolving complex tissue heterogeneity, we applied the platform to profile 47,712 nuclei isolated from the cortex (superficial and deep layers), hippocampus, striatum, and olfactory bulb of adult mice (6–10 weeks old). Nuclei were labeled with anti-NeuN antibody and sorted via fluorescence-activated cell sorting (FACS), yielding 44,689 RNA-qualified nuclei (Additional file 3: Tables S4–S7). For methylation detection, reads with mCH levels < 50% were classified as DNA-derived. The following filtering criteria were applied: (i) 5hmC accuracy > 98%, (ii) 5mC conversion efficiency (< 2% for 5hmC-seq or > 96% for 5mC + 5hmC-seq), and (iii) unmodified cytosine background < 1% (Additional file 1: Fig. S1A). Ultimately, 43,075 neuronal and non-neuronal cells with matched transcriptomes, 5hmC profiles, and 5mC + 5hmC profiles passed quality control for downstream analysis. Sequencing generated an average 8.2 million 5hmC raw reads per cell with 5.9 million aligned reads covering 1.4 million CpG sites. Correspondingly, 8.3 million 5mC + 5hmC raw

reads were generated per cell, with 5.1 million aligned reads covering 1.4 million CpG sites (Additional file 1: Fig. S2A–S2E). RNA profiling detected on average transcripts from 2,741 genes per neuron and 1,352 genes per non-neuronal cell (Additional file 1: Fig. S2F and S2G).

To precisely determine cell-type-specific identities, the transcriptomic vectors generated by the platform were computationally projected onto a highly annotated reference single-cell RNA sequencing (scRNA-seq) atlas [2]. Utilizing robust principal component analysis and label transfer algorithms, hierarchical cellular clustering successfully recovered all canonical neuronal and non-neuronal subclasses (Additional file 1: Fig. S3A–S3F). Using marker genes from integrated clusters (Additional file 4: Table S8), we annotated 13 major types (e.g., excitatory: 'L2/3 IT CTX Glut', 'DG Glut'; inhibitory: 'STR D1 GABA') and 31 subtypes (Fig. 2A, Additional file 1: Fig. S4A and S4B). Subsequent spatial validation strongly corroborated these bioinformatic annotations, verifying that specific excitatory neuron subtypes mapped precisely to their known anatomical origin tissues, while diverse GABAergic populations tightly corresponded to their expected regional microenvironments (Additional file 1: Fig. S5A).

Interestingly, while the FACS-based NeuN gating strategy exhibited strong concordance with transcriptomic identities across the cortex and hippocampus, the analysis uncovered marked ambiguity within the olfactory bulb (Additional file 1: Fig. S5B–S5D). In this specific region, the fluorescent intensity of the NeuN marker presented a blurred, continuous distribution, leading to ambiguous gating boundaries. Consequently, the initial FACS-predicted identities in the olfactory bulb displayed a notable divergence from the final high-resolution transcriptional profiles, highlighting the limitations of relying solely on single-protein physical sorting techniques and underscoring the necessity of using the transcriptome as the unbiased indicator of cellular identity.

Importantly, global metrics of true 5mCG and 5hmCG demonstrated pronounced variability across different cellular lineages. In sharp contrast, the composite 5mCG + 5hmCG signal—the exact signal generated by bisulfite protocols—remained static across divergent neuronal and glial populations (Fig. 2B and C). This striking divergence highlights a fundamental limitation of bisulfite-based measurements in capturing cellular epigenetic heterogeneity in the brain and demonstrates that dynamic 5mC-to-5hmC conversion represents a major source of epigenetic plasticity in the brain. Furthermore, the data showed that non-CpG modifications (5mCH + 5hmCH) mirrored the true 5mCH levels, confirming that TET-mediated oxidative activity is strongly enriched at classical CG motifs.

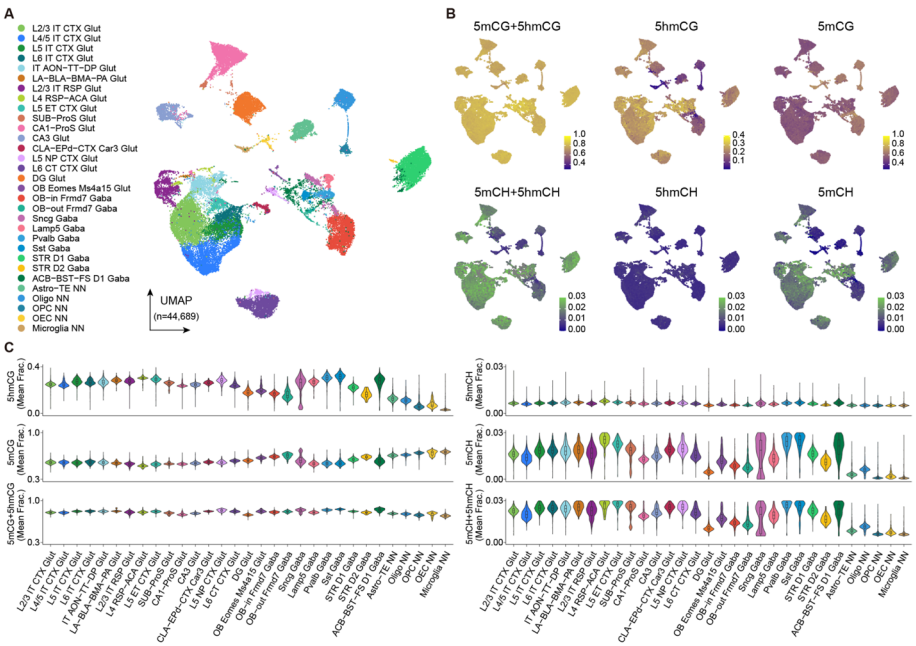


Fig. 2 Multi-modal profiling of mouse brain cells with Joint-Cabernet. **A** Uniform manifold approximation and projection (UMAP) plot of transcriptome-based clustering of joint profiling data ($n = 44{,}689$), colored by subclasses annotated with marker genes. **B** Cellular global 5mC + 5hmC, 5hmC, and 5mC levels at CG and CH sites, visualized with UMAP. **C** Violin plots showing the abundance of 5mC + 5hmC, 5hmC, and 5mC levels at CG and CH sites across different cell types

Consistent with theoretical models of neurobiology, fully terminally differentiated, postmitotic neurons displayed superior global 5hmC levels compared to actively dividing non-neuronal cells [11] (Fig. 2C). This substantial accumulation is a direct consequence of the postmitotic state: because mature neurons do not undergo DNA replication, the replication-dependent dilution of 5hmC marks is completely abolished, allowing TET enzymes to chronically deposit and maintain the modification over the lifespan of the cell. Notably, the degree of 5hmCG enrichment was highly subtype-specific. Somatostatin-expressing (Sst) GABAergic interneurons—a population that originates early in neurodevelopment from the medial ganglionic eminence (MGE) and maintains prolonged integration within mature cortical circuits—exhibited the highest concentrations of 5hmCG globally. Conversely, neuronal populations residing in the dentate gyrus (DG) and the olfactory bulb (OB) displayed markedly depleted 5hmC levels. This relative depletion is directly linked to the ongoing, active adult neurogenesis occurring within the subgranular zone (SGZ) and subventricular zone (SVZ), which continuously repopulate these specific brain regions with highly proliferative, replication-competent immature cells that constantly dilute their 5hmC profiles through cell divisions [39, 40]. Surprisingly, even within transcriptionally homogeneous cellular clusters, single-cell resolution revealed substantial stochastic variation in global 5hmC abundance, suggesting that TET-mediated oxidation operates with a certain degree of intrinsic molecular stochasticity.

DNA hydroxymethylation marks actively transcribed DNA

The integration of the matched transcriptomic and epigenomic matrices allowed for a comprehensive and systematic dissection of how these chemical marks directly influence transcriptional output. By stratifying genes into non-expressed, low-expressed, and high-expressed categories across neuronal subtypes, we observed pronounced differences in 5mCG and 5hmCG distribution in transcriptionally active loci (Fig. 3A). At transcription start sites (TSS) and core promoter elements, both true 5mCG and 5hmCG were depleted in actively transcribed loci, supporting the classical model that the absence of CG methylation is required for the assembly of the transcriptional initiation machinery. However, despite the absolute reduction in total modifications at the TSS, the relative ratio of 5hmCG to total methylation spiked significantly, suggesting that localized, promoter-proximal hydroxymethylation may actively facilitate the transition from a poised to an active transcriptional state, potentially by actively repelling repressive methylation-binding proteins that otherwise lock the promoter in a closed state.

The most pronounced regulatory divergence occurred within the gene bodies. Across all 31 characterized neuronal and non-neuronal subtypes, the activation of transcription was universally hallmarked by a severe depletion of gene body 5mCG, coupled with a massive accumulation of 5hmCG. Crucially, the aggregate 5mCG + 5hmCG composite signal across these active gene bodies remained stable, creating a deceptive equilibrium, masking the active, frequent conversion of 5mC to 5hmC. This dynamic shift establishes 5hmCG as a definitive epigenetic signature of active transcription in the mammalian brain. In contrast, modifications at non-CpG (CH) sites exhibited minimal 5hmC conversion. This scarcity of 5hmCH perfectly explains why historical bisulfite measurements of CH methylation correlate tightly with gene expression [41, 42]—because bisulfite-measured CH methylation is almost entirely composed of true 5mCH, devoid of the confounding 5hmC overlap that ruins CG measurements. Notably, while 5hmCH displayed a positive correlation with gene expression, its low absolute abundance and poor signal-to-noise ratio fundamentally restrict its utility as a primary predictive biomarker (Additional file 1: Fig. S6A–S6C).

The superior resolving power of this methodology was highlighted in the analysis of cell-type-specific biomarker genes. For instance, the *Cdh18* gene—a definitive marker for layer 6 corticothalamic excitatory neurons—displayed a massive surge of 5hmCG and a reduction of 5mCG within its gene body, accurately tracking its selective transcription in that specific cellular layer. Yet, the composite bisulfite signal across the *Cdh18* gene body remained entirely flat across all cortical layers, rendering the activation event invisible to bisulfite assays (Fig. 3B). Global Pearson correlation matrices solidified this trend across the transcriptome: with gene body 5hmCG associating positively with expression and 5mCG showing stronger anti-correlation than combined signals

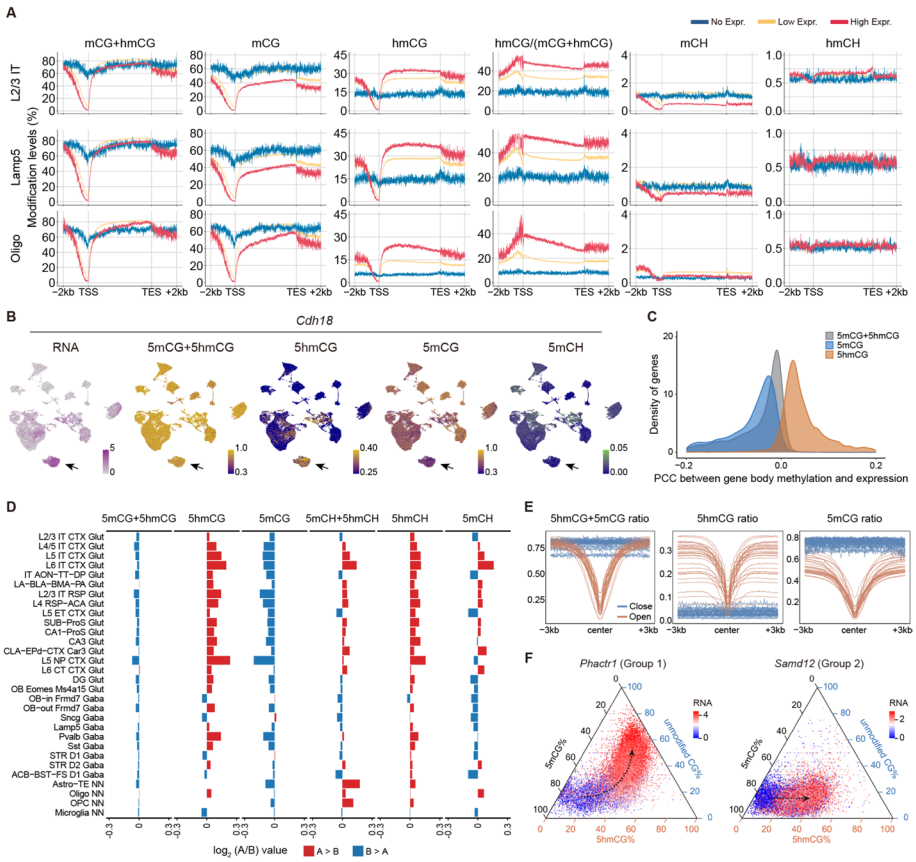


Fig. 3 Accurate characterization of the regulatory functions of the DNA methylation and hydroxymethylation. **A** DNA modification levels around genebodies of genes categorized by expression levels: no expression (blue), low expression (yellow), and high expression (red). **B** UMAP plots colored by gene expression and genebody modification rates of *Cdh18* (a marker for 'L6 CT CTX Glut', which is indicated by a black arrow). **C** Distribution of Pearson correlation coefficients (PCCs) between gene expression and genebody CG modifications across single cells. **D** Differences in DNA modification levels between A and B compartments at the subclass level. **E** DNA modification ratios of 5mCG + 5hmCG, 5hmCG, and 5mCG are shown in open or closed chromatin regions, with each subclass represented by one line. **F** Ternary plots showing genebody 5mCG, 5hmCG and unmodified CG levels (%), with RNA expression coded in color. Examples from two distinct gene groups are presented to illustrate the differential regulatory mechanisms in gene activation

(Fig. 3C). Notably, we observed intra-subclass gene activation coincident with 5mCG-to-5hmCG conversion on the gene body, which is undetectable by conventional combined methylation signals (Additional file 1: Fig. S6D). Together, these findings establish gene body 5hmCG as a dynamic marker of neuronal gene transcription and emphasize the distinct roles of 5hmC and 5mC CpGs in relation to neuronal gene expression.

Hydroxymethylation is prevalent in active chromatin

Beyond linear gene models, the dataset was subsequently mapped onto high-resolution three-dimensional chromatin conformation maps. The mammalian genome is topologically partitioned into actively transcribed, structurally open A compartments and densely packed, repressed B compartments [43]. Using single-cell chromatin conformation data [6], we observed that active A compartments are enriched with 5hmC and depleted of 5mC (Fig. 3D, Additional file 1: Fig. S7A). Analysis of 100-kb-bin compartment score changes (where positive scores reflect a structural shift toward open A compartments) across fundamentally different cell types revealed a tight positive correlation with 5hmCG levels and, in contrast, a negative correlation with 5mCG levels. However, the

combined 5mC + 5hmC signal entirely obscures this intricate opposing regulation (Additional file 1: Fig. S7B), once again highlighting the absolute necessity of distinguishing these distinct chemical modifications.

Building on previous reports of cell-type-specific CREs (cis-regulatory DNA elements) in the adult mouse brain [9], we analyzed the methylation pattern in open chromatin regions. Interestingly, while conventional bisulfite sequencing signals show similar methylation levels at the periphery of open regions compared to closed chromatin regions, the underlying molecular composition differs fundamentally. The flanks of active, open chromatin are in fact sites of intense epigenetic editing, characterized by a massive enrichment of 5hmCG and a total reduction of 5mCG (Fig. 3E, Additional file 1: Fig. S7C and S7D). The highly accessible centers of these open chromatins likely physically accommodate the bulky TET enzyme complexes, allowing them to rapidly catalyze the full oxidative cascade from 5mC to unmodified cytosine, resulting in a localized cratering of all cytosine modifications at the center of the CRE, bordered by a dense distribution of 5hmC.

5hmCG marks distinct modes of gene activation in neuronal cells

Certain crucial regulatory proteins, such as MeCP2 and MBD3, have been definitively shown to bind directly to 5hmC [16, 20], strongly suggesting that 5hmC may have independent functional roles far beyond simply acting as a transient intermediate in active DNA demethylation. A key, long-standing question in complex epigenetic regulation is whether 5hmC itself acts as a truly distinct, standalone regulatory mark. Using the immense resolution of our Joint-Cabernet data, we can simultaneously examine multiple intersecting DNA modifications directly alongside gene expression from the exact same cells, allowing us to systematically and precisely assess their interactions and regulatory influences. To thoroughly investigate this, we classified thousands of genes into distinct regulatory groups strictly based on the correlative relationship between their gene body modifications and their actual transcriptional activity (Additional file 1: Fig. S8A). Among 3,337 specific genes that were highly positively associated with 5hmCG, 3,254 (an overwhelming 97.5 percent) exhibited a strict negative correlation with 5mCG. This strong inverse relationship is consistent with the biochemical conversion of 5mC to 5hmC during standard active demethylation. However, deeper analysis suggests that simple demethylation is far from the only functional outcome. Within these actively demethylated and heavily transcriptionally upregulated genes, two major regulatory patterns distinctly emerge.

Group 1 consists of 2,376 highly dynamic genes that undergo a progressive, multi-step transition from 5mCG directly to 5hmCG and subsequently all the way to entirely unmodified CG, accompanied by steadily increasing gene expression across various neuronal subtypes (Fig. 3F, Additional file 1: Fig. S8B). This complex process occurs in two distinct stages: an initial, rapid conversion of 5mCG to 5hmCG, heavily associated with moderate transcriptional activation, followed by a much further, prolonged transition to fully unmodified CG, directly linked to a far more pronounced activation state. Group 2, in contrast, consists of 840 distinct genes whose intense transcriptional activation is primarily associated solely with the conversion of 5mCG to 5hmCG, entirely without a substantial terminal shift toward unmodified CG (Fig. 3F, Additional file 1: Fig. S8C). The prolonged persistence of 5hmC in this group suggests that 5hmC may possess deep functions entirely independent of its role as a mere demethylation intermediate, potentially acting directly through recognition by highly specific 5hmC-binding proteins or unique transcription factors.

Group 1 and Group 2 exhibit divergent genomic and functional characteristics beyond their DNA modification dynamics (Additional file 1: Fig. S8D). Genes in Group 1 are significantly longer (median length: 109.7 kb vs. 61.5 kb in Group 2) and contain more CpG sites (median CpG number: 1126 vs. 597). A more complex genomic architecture may favor progressive demethylation, requiring multi-step epigenetic remodeling to achieve transcriptional activation. In contrast, Group 2 genes are shorter with fewer CpGs, suggesting their activation may rely on localized rather than large-scale chromatin restructuring. These gene structural differences align with the functional specialization and expression dynamics of each group. Group 1 is enriched for neurodevelopmental processes (GO: nervous system development, synapse organization, Additional file 1: Fig. S8A, Additional file 5:

Tables S9–S10) and contains most of cell-type-specific neuronal markers. Notably, Group 1 genes exhibit significantly higher variability in RNA expression across neuronal subtypes (Additional file 1: Fig. S8E), reflecting their roles in establishing cell identity through context-dependent regulation. Group 2, however, shows distinct enrichment for basic cellular maintenance functions (GO: primary metabolic process, cellular component organization). The lower expression variability of genes in Group 2 across cell types aligns with their housekeeping roles. Together, the advanced Joint-Cabernet multi-omics approach definitively resolves distinct, complex DNA methylation patterns that would be entirely indistinguishable and lost to conventional bisulfite analyses. The distinct and stable distribution of 5hmC in Group 2 strongly suggests that 5hmC may have functional relevance beyond its role as a transient intermediate in active DNA demethylation.

Differentially hydroxymethylated regions across neurons

Differential 5hmC distribution may play a significant role in neural development and cell differentiation [44, 45]. To explore its potential regulatory functions in different neuronal subtypes, we identified 200,438 cell-type-specific differentially hydroxymethylated regions (DHMRs) (Fig. 4A–C). Among these regions, which predominantly overlap with intronic and intergenic elements, 153,580 exhibited hyper-5hmCG in at least one neuronal subtype. Given that over 40% of the DHMRs are located within gene bodies, we investigated their correlation

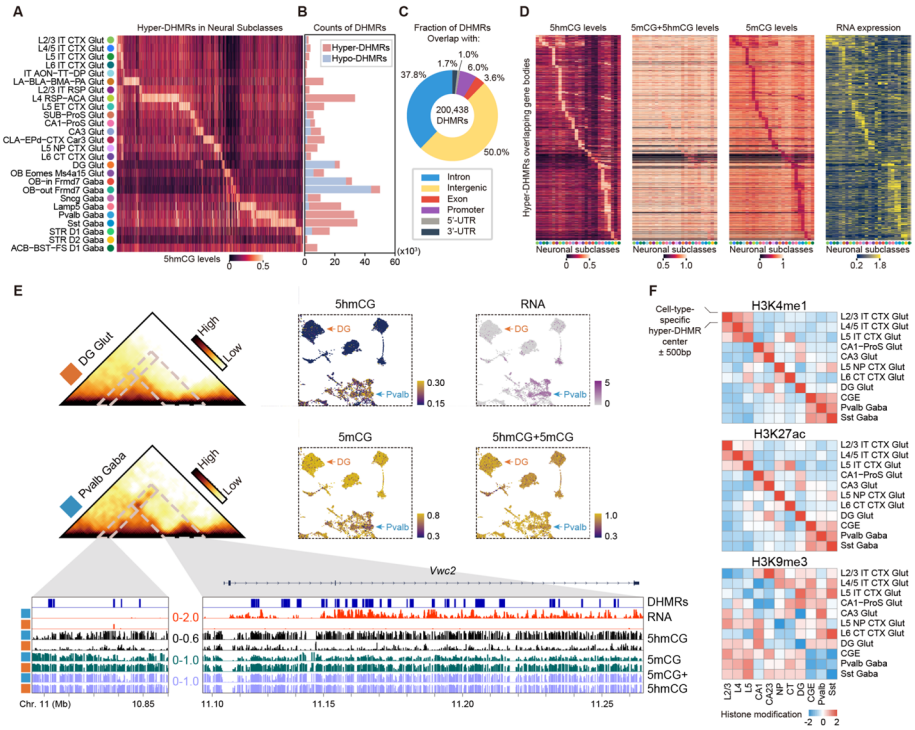


Fig. 4 Regulatory roles of differentially hydroxymethylated regions in neurons. **A** The 5hmCG ratios in hyper-differentially hydroxymethylated regions (DHMRs) across neuronal subclasses. **B** The distribution of DHMRs across neuronal subclasses. **C** The annotation of hyper-DHMRs by genomic elements. **D** The 5hmCG, 5mCG + 5hmCG, and 5mCG levels of genebody hyper-DHMRs, along with the RNA expression levels of the corresponding genes. **E** The *Vwc2* distal regulatory region (chr11: 10,805,000–10,860,000), identified by 3C data from Liu et al.[2], shows up-regulation of 5hmCG and down-regulation of 5mCG, along with concomitant gene activation in the 'Pvalb Gaba' subclass. The contact maps show the interaction intensity. UMAP and IGV visualizations show the *Vwc2* RNA expression levels and the 5hmCG, 5mCG, and 5mCG + 5hmCG levels in the upstream distal region. **F** Heatmaps showing the average histone modification signals of cell types identified in the Paired-Tag dataset[42] on hyper-DHMR centers ± 500 bp. Rows represent the cell-type-specific hyper-DHMRs detected in this study, and columns represent the matched cell types from the Paired-Tag dataset

with gene expression patterns across neuronal cell types. By mapping hyper-DHMRs to their corresponding gene loci (Additional file 6: Table S11), we identified 963 genes whose expression levels showed a significant positive correlation with 5hmCG signals within these regions. These DHMRs displayed a marked depletion of 5mCG signals and were also highly cell-type-specific, with minimal variation in the combined 5mCG + 5hmCG signals across different cell types (Fig. 4D).

In addition to the DHMRs located within gene bodies, we identified a substantial number of DHMRs in intergenic regions, which are likely to play critical roles in gene regulation. To investigate this, we used 3 C data to determine the target genes of intergenic regions containing DHMRs (Additional file 1: Fig. S9A). By focusing on the correlation of interaction strength and gene expression levels (Additional file 1: Fig. S9B), we identified 4,105 expression-associated interaction anchors (Additional file 7: Table S12). 5hmCG levels at these distal anchors showed a significant positive correlation with the expression levels of their corresponding target genes (Additional file 1: Fig. S9C). For instance, the *Vwc2* gene, specifically expressed in 'Pvalb GABA' neurons and known for its role in neuronal development in both mouse and human [46], showed increased interaction strength with an anchor located 255 kb upstream. This enhanced interaction was associated not only with gene activation but also with a marked accumulation of 5hmCG and significant depletion of 5mCG in the distal anchor (Fig. 4E). Similar results can also be observed for the *Etl4* gene and 160 kb upstream distal anchor in 'L5 NP CTX Glut' subclass (Additional file 1: Fig. S9D–S9F).

5hmC has been shown to influence chromatin states by interacting with other epigenetic marks [47]. Histone modifications, such as H3K27ac and H3K4me1, are known markers of active enhancers, while H3K9me3 is associated with heterochromatin and repressed regions. To investigate the relationship between 5hmC and these modifications, we retrieved histone mark signal data for various cell types from the Paired-Tag dataset [48] and analyzed the differential distribution patterns of these marks on cell-type-specific hyper-DHMRs. We found that hyper-DHMRs in each cell type were enriched with active marks, including H3K4me1 and H3K27ac, and showed a depletion of repressive marks, such as H3K9me3 (Fig. 4F). This multi-omic intersection suggests that TET-mediated 5hmC deposition is not merely a passive consequence of transcription, but may instead play an active role in shaping the open chromatin enhancer landscape associated with neuronal lineage specification.

Spatially resolved epigenetic landscapes based on location imputation

Spatially resolved patterns of DNA methylation and hydroxymethylation are essential for understanding how epigenetic regulation interfaces with cytoarchitecture, neuronal identity, and circuit organization in the brain. In particular, 5mC and 5hmC exhibit pronounced cell-type- and region-specific distributions that are tightly coupled to transcriptional programs and neuronal function. However, despite their central roles in neurodevelopment and brain plasticity, spatially resolved profiling of 5mC and 5hmC remains largely unexplored. While standard single-cell transcriptomics provides excellent cellular resolution, it inherently obliterates the spatial organization of the complex brain architecture during tissue dissociation. To resurrect this critical spatial context, the methylation matrices were computationally imputed onto high-spatial-resolution Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH) reference map [49–51]. The analysis successfully projected the 5mC and 5hmC profiles back onto their exact spatial coordinates within the mouse brain sections (Fig. 5A, Additional file 8: Table S13). To ensure the accuracy of spatial location imputation, we performed integration and spatial prediction separately for excitatory neurons, inhibitory neurons, and non-neuronal cells (Fig. 5B and C). Among these categories, excitatory neurons exhibited pronounced spatial specificity, with subregion boundaries faithfully preserved in the spatial embeddings. For instance, projection neurons corresponding to different cortical layers were accurately localized within their respective layers ('L2/3 IT CTX Glut', 'L5 ET CTX Glut', 'L6 CT CTX Glut').

This spatial projection allowed for the discovery of previously uncharacterized, geographically driven epigenetic gradients. For instance, the *Grik3* gene, which encodes glutamate receptors involved in various neurophysiological processes [52], exhibited a dorsal–ventral gradient of increasing expression in the intratelencephalic (IT)

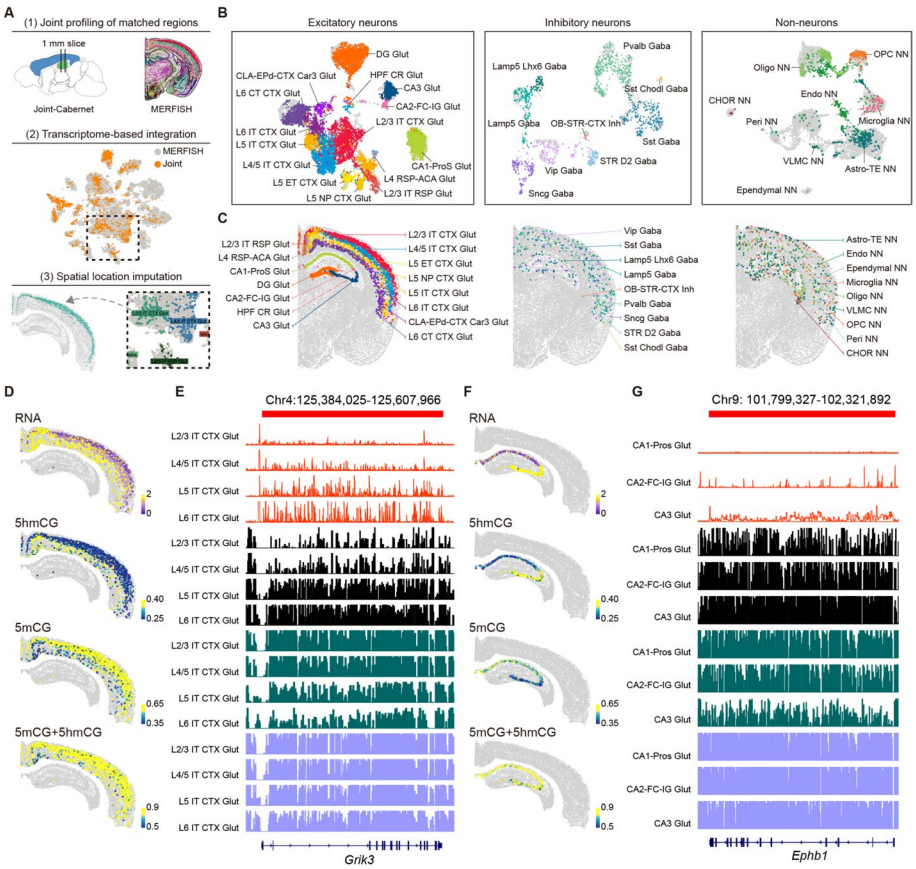


Fig. 5 Spatially resolved epigenetic landscapes based on location imputation with MERFISH data. **A** Schematic of the spatial projection of 5hmCG and 5mCG profiles through transcriptome-based integration with MERFISH data. **B** UMAP plot showing excitatory neurons (left), inhibitory neurons (middle), and non-neurons (right), with gray dots representing cells from the MERFISH data. **C** Projected spatial locations of excitatory neurons (left), inhibitory neurons (middle), and non-neurons (right). **D** Spatially projected RNA expression, 5hmCG, 5mCG, and 5mCG + 5hmCG levels of *Grik3* in related subclasses in the cortex. **E** IGV tracks corresponding to (**D**). **F** Spatially projected RNA expression, 5hmCG, 5mCG, and 5mCG + 5hmCG levels of *Ephb1* in related subclasses in the hippocampus. **G** IGV tracks corresponding to (**F**)

neurons of subcortical layers. This expression pattern was accompanied by a rise in gene body 5hmCG levels and a decrease in 5mCG, with minimal change observed in the combined 5mCG + 5hmCG signal (Fig. 5D and E). Similarly, *Ephb1*, a gene crucial for axon guidance and dendritic spine morphogenesis [53], also showed a gradient expression pattern in the cornu ammonis (CA) region of the hippocampus (Fig. 5F and G). Our analysis revealed that the association of 5hmC with gene activation and 5mC with repression occur in spatially organized patterns, suggesting a contribution to spatial epigenetic regulation. These findings highlight the importance of spatially resolved epigenomic analyses and provide a conceptual framework for future studies investigating how epigenetic modifications encode spatial information in complex tissues such as the brain.

Discussion

Joint-Cabernet represents an advanced, highly scalable single-cell multi-omics platform that resolves the long-standing challenges in distinguishing 5hmC and 5mC while synchronously capturing transcriptomes from the exact same cells. By applying this method to complex mouse brains, our study discovered highly dynamic hydroxymethylation landscapes that tightly correlate with transcriptional plasticity, chromatin organization, and intricate spatial epigenetic patterning. These findings, which would have remained entirely invisible by conventional bisulfite sequencing, substantially advance the understanding of complex epigenetic regulation in the highly dynamic mammalian brain and provide a framework for dissecting the functional role of 5hmC in neurodegenerative diseases.

While the platform enables high-throughput profiling, it is important to acknowledge that single-cell genomic coverage remains inherently limited and this plate-based method still require more extensive labors than

droplet-based methods common for RNA-seq and ATAC-seq. Moreover, the physical separation of material from each cell into two channels for profiling 5hmC and 5mC + 5hmC precludes the co-detection of both modifications on the same DNA molecules. Consequently, large genomic bins are required to mitigate the sparsity of single-cell measurements and obtain robust estimates of genomic modification levels. Additionally, the platform's integrated transcriptomic profiling exhibits lower gene detection sensitivity compared to dedicated, standalone single-cell RNA sequencing methods, requiring integration with external reference transcriptomic datasets for the most robust analysis. Also, the reliance on mCH levels for contamination removal restricts the application in species with extensive mCH modifications. These limitations pose challenges in resolving single-cell level correlations and stochastic epigenetic fluctuations, highlighting the need for technical optimizations to enhance sequencing quality.

Conclusions

Joint-Cabernet serves to aggressively enable future, massive-scale studies that prioritize completely defining 5hmC and 5mC dynamics specifically during brain maturation, aging, and neurodegeneration. With rapidly emerging evidence strongly suggesting that the entire mammalian brain undergoes substantial age-related epigenetic remodeling, the ability to seamlessly capture these concurrent modalities is increasingly critical. In the context of neurodegenerative diseases, the massive epigenetic drift observed during aging is anticipated to actively drive large-scale transcriptional dysregulation. Our technology provides a powerful pivot to enable future therapeutic exploration.

Methods

Mouse brain tissues

All mouse experiments were approved by the Institutional Animal Care and Use Committee for Rodents and Rabbits of Changping Laboratory (approval number: CPNL-AP-2025-0006) and conducted following the ethical guidelines. Adult male C57BL/6 J mice (7–10 weeks old) were obtained from Beijing HFK Bioscience Co., Ltd. (Beijing, China) and housed on a 12-h light/dark cycle with food and water ad libitum in a dedicated SPF facility room before sacrifice and rapid brain extraction. The superficial cortex, deep cortex, hippocampus, olfactory bulb, and striatum were freshly dissected and immediately placed in ice-cold phosphate-buffered saline (PBS) for subsequent processing. The superficial cortex and deep cortex were further dissected from the cortex by separating them roughly at the midline. For Joint-Cabernet samples, the tissues were dissected and immersed in 1 mL of RNAlater Solution (Invitrogen, AM7020) at 4 °C overnight before proceeding to nuclei isolation, fixation, in situ reverse transcription, antibody labeling, and sorting. For samples integrated with the MERFISH data, several P49 mice (three in the first batch and five in the second) were sacrificed to obtain fresh 1000-μm-thick coronal brain sections centered at 7300 μm from the anterior end, as referenced by the Allen Brain Atlas CCFv3 (http://atlas.brain-map.org). These sections were further dissected into left cortex, right cortex, left hippocampus, and right hippocampus before proceeding to nuclei isolation, fixation, in situ reverse transcription, antibody labeling, and sorting.

Nuclei isolation

Nuclei were isolated as previously described in the Dip-C protocol [54], with minor modifications. All buffers used in sample preprocessing were freshly supplemented with RNaseOUT (Thermo Fisher Scientific, 10,777,019) at a 1:800 ratio. Briefly, brain tissues were first washed in 1 mL of ice-cold PBS and transferred into 2 mL of

nuclei isolation medium with Triton (0.25 M sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM HEPES pH 8.0, 1 µM DTT, 0.1% Triton X-100).

Tissues were homogenized using a 2 mL Dounce homogenizer (Sigma D8938) with 5 strokes of the loose (A) pestle and 15 strokes of the tight (B) pestle, followed by centrifugation at 100 g, 4 °C for 8 min. The pellet was resuspended in 1 mL of nuclei isolation medium without Triton (0.25 M sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM HEPES pH 8.0, 1 µM DTT) and centrifuged again under the same conditions. The resulting pellet was resuspended in 0.4 mL of nuclei isolation medium without Triton and filtered through a 70 µm strainer (Stemcell, 27,216), with two additional washes of the strainer using 0.2 mL of nuclei isolation medium without Triton. Nuclei isolation was completed by another centrifugation at 100 g, 4 °C for 8 min, followed by resuspension in 1 mL of PBS-RI.

Fixation and in situ reverse transcription

Nuclei were fixed and subjected to in situ reverse transcription as previously described in the SPLiT-seq protocol [55], with some modifications. Briefly, nuclei pellets were resuspended in 1 mL of ice-cold PBS-RI (PBS supplemented with RNaseOUT at a 1:800 ratio) before adding 3 mL of cold 1.33% formaldehyde solution (16% formaldehyde, Thermo Fisher Scientific, 28,906, diluted in 1 × PBS) and gently mixed.

The nuclei were fixed for 10 min on ice with occasional inversion for mixing, followed by the addition of 160 µL of 5% Triton X-100 (Sigma-Aldrich, 93,443–100 ML) for a 3-min permeabilization on ice. The nuclei were then centrifuged at 500 g, 4 °C for 5 min, and resuspended in 500 µL PBS-RI, and mixed with 500 µL of cold 100 mM Tris–HCl pH 8.0 (Invitrogen, AM9855G) and 20 µL 5% Triton X-100. After another centrifugation, the nuclei were resuspended in 200 µL of 0.5 × PBS-RI and 4 µL of 5% Triton X-100, then centrifuged and resuspended in 0.5 × PBS-RI before counting.

8,000 to 16,000 fixed nuclei per reaction were then in situ reverse transcribed in 20 µL RT reactions (4 µL of 5 M Betaine (Sigma-Aldrich, B0300-5VL), 4 µL of 5 × RT buffer (Thermo Fisher Scientific), 0.125 µL of RNaseOUT, 0.25 µL of SUPERase-In (Thermo Fisher Scientific, AM2696), 1.43 µL of 7 mM-each dNTP with hydroxymethyl-dCTP instead of dCTP (prepared by diluting dNTPs set (NEB, N0356S, excluding dCTP), and 5-hydroxymethyl-dCTP (Jena Bioscience, NU-932 L) in 0.1 × TE), 2 µL of 100 µM V2-Poly-dT_TSO + RNA primer (Additional file 2: Table S1), 2 µL of Maxima H Minus RTase (Thermo Fisher Scientific, EP0753), 6.195 µL of 0.5 × PBS-RI containing the nuclei) with the following program: 50 °C 10 min, 3 * [8 °C 12 s, 15 °C 45 s, 20 °C 45 s, 30 °C 30 s, 42 °C 2 min, 50 °C 3 min], 50 °C 5 min, and held at 4 °C. 30-µL RT mix was then added to the 20-µL reaction for template switching (4.8125 µL of 5 M Betaine, 10 µL of 20% Ficoll PM-400 (Sigma-Aldrich, F5415-25 ML), 6 µL of 5 × RT buffer, 0.3125 µL of RNaseOUT, 0.625 µL of SUPERase-In, 4 µL of 7 mM-each dNTP with 5-hydroxymethyl-dCTP instead of dCTP, 1.25 µL of 100 µM ME19bp + 2 G-TSO primer (Additional file 2: Table S1), 3 µL of Maxima H Minus RTase), and incubated at 25 °C for 30 min and 42 °C for 90 min, before being held at 4 °C. Every 8 or 16 reactions from the same samples were pooled and 0.45 µL of 10% Triton X-100 was added per reaction, before centrifugation at 1,500 g, 4 °C for 15 min. The pellets were resuspended in 1 mL blocking buffer-RI (0.5% BSA in PBS, supplemented with 1:800 volume of RNaseOUT), and again centrifuged at 1,500 g, 4 °C for 15 min. All the pellets from each sample were resuspended and pooled in 1 mL blocking buffer-RI before being stored overnight in a 4 °C refrigerator.

Antibody labeling and sorting

Nuclei were antibody labelled as previously described, with optimizations. To nuclei in 1 mL blocking buffer-RI, 200 µL of antibody mix (1.2 µL (fresh nuclei) or 0.5 µL (fixed nuclei) Alexa Fluor 488-conjugated anti-NeuN antibody (Millipore, MAB377X) diluted in 150 µL PBS-RI and 50 µL blocking buffer-RI and pre-incubated at room temperature for 5 min with protection from light) was added, and incubated for 25 min at 4 °C with rotation. DAPI (Sigma, MBD0015) was then added at a final concentration of 2 µg/mL before a 5-min incubation at

4 °C with rotation. The nuclei were pelleted through centrifugation at 1,500 g, 4 °C for 15 min, and resuspended in 1 mL blocking buffer-RI for a 20-min incubation at 4 °C with rotation as a wash step. Finally, the nuclei were centrifuged at 1,500 g, 4 °C for 15 min, and resuspended in PBS-RI for single-nucleus sorting into 96-well plates with a BD FACSAria SORP sorter (BD Biosciences) equipped with an 85-μm nozzle. A standard gating strategy based on DAPI staining, SSC/FSC levels, and NeuN antibody staining, was established for all samples of each brain region. Single nuclei were sorted into wells containing 2 μL lysis buffer (20 mM Tris–HCl pH 8.0, 20 mM NaCl, 0.15% Triton X-100, 15 mM DTT, 1 mM EDTA, 3 μg Qiagen Protease (Qiagen, 19,157), 0.5 μM ssCarrierDNA (Additional file 2: Table S1), and 3 spike-in control (0.1 pg of non-methylated Lambda DNA, 0.03 pg of CpG methylated pUC19, 0.3 pg of fully hydroxymethylated ClaI, as described in Cabernet), with columns 1–11 sorted with NeuN + nuclei and column 12 with NeuN- nuclei. After sorting, plates were briefly centrifuged before lysing with the following program: 50 °C 1 h, 65 °C 1 h, 70 °C 15 min and held at 4 °C. Lysed plates were stored at − 80 °C for later use.

Transposome preparation

The transposon design was inspired by the 's3' design. The TSO_3U_Tn5 transposome was designed in a unified fashion. The Tn5_3U_oligo_V2 oligo was dissolved in 0.1 × TE, mixed at a ratio of 1:1.1 with the Tn5_comp oligo, and annealed through heating to 95 °C for 1 min and gradual cooling down at a speed of − 0.1 °C/3 s till 25 °C, in 1 × annealing buffer (10 mM Tris–HCl pH 8.0, 50 mM NaCl, 1 mM EDTA). The transposase was purified after expression from the pTXB1-Tn5 plasmid (Addgene). Comparable results can be generated with commercial Tn5 transposase (Vazyme Biotech Co., Ltd.). The transposase (6.25 μM monomer) and the annealed transposons (6.25 μM) were then assembled by mixing at a volume ratio of 1:1.1 and incubated at 25 °C for 30 min. The assembled transposome was then diluted with Tn5 storage buffer (20% glycerol, 1 mM DTT, 0.1 mM EDTA, 50 mM NaCl, 10 mM Tris–HCl pH 8.0) to the concentration of 300 nM monomer and stored at − 80 °C. Another kind of Tn5, used for RNA library preparation in the Joint-Cabernet experiments, Nextera_i7_Tn5, was also prepared according to the protocol above, except the uracil-containing primer was substituted by the Nextera_i7 oligo (Additional file 2: Table S1).

Joint-Cabernet library preparation and sequencing

Single-nucleus lysate was transposed in a 10-μL reaction of 10 mM Tris–HCl pH 8.0, 5 mM MgCl2, 8% PEG-8000 (Sigma-Aldrich, P1458), 1.045 nM homemade TSO_3U_Tn5 transposome by incubation at 55 °C for 10 min. Transposases were removed by addition of 1 μL 2 mg/mL Qiagen protease, combined with 1 μL 12 × Tn5 stop buffer of 0.6 M NaCl, 90 mM EDTA, and 0.02% Triton X-100, and incubation at 50 °C for 40 min and 70 °C for 15 min.

Then, 5.5 μL of contents was moved from each well to a corresponding well in a new plate for the protection and detection of 5mC + 5hmC, and the remaining contents in each well was for the protection and detection of 5hmC only.

Gaps before the/U/U/U/blockage in the transposed DNA were filled, and cell barcodes were added, by the addition of 4.42 μL TSO-gap-PCR mix (2 μL Q5 reaction buffer (NEB), 2 μL Q5 High GC Enhancer (NEB), 0.035 μL 1 M MgCl2, 0.286 μL 5hm-dNTP (7 mM each dNTP, with 5-hydroxymethyl-dCTP instead of dCTP for the plates intended for 5hmC detection) or 5 m-dNTP (7 mM each dNTP, with 5-methyl-dCTP instead of dCTP for the plates intended for 5mC + 5hmC detection), 0.1 μL Q5 high-fidelity DNA polymerase (NEB, M0491L)), and 0.2 μL of 50 μM well-unique LNA-modified cell barcode primer (Additional file 2: Table S2). The following program was used for PCR amplification: 4 °C (pre-cooling of the block), 50 °C 3 min (gap-filling), 98 °C 30 s, 10 × [98 °C 10 s, 61 °C 20 s, 72 °C 1 min], 72 °C 5 min, held at 4 °C.

After cell barcoding, contents from each 96-well plate were pooled together into a single 5-mL tube for purification with the addition of 40 ng sonicated carrier lambda DNA in water and 1270 μL 30% diluted SPRI beads

(Beckman Coulter, the dilution buffer was composed of 19% PEG-8000, 2.5 M NaCl, 10 mM Tris–HCl pH 8.0, 1 mM EDTA, 0.05% Tween 20 (Millipore,655,204-100ML)). The tubes were thoroughly vortexed, incubated at room temperature for 10 min, briefly centrifuged and placed onto magnet racks (Stemcell, 18,103) until clear. Four rounds of washing with ~ 6.5 mL of 80% ethanol was performed before elution with 48 µL 1 mM Tris–HCl pH 8.0 on a magnet rack for PCR tubes.

The DNA eluted for 5hmC detection was added to a 40-µL BGT reaction mix (1.76 µL 2 mM UDP-glucose (NEB), 8.8 µL NEBuffer 4 (NEB), 4.4 µL T4 Phage β-glucosyltransferase (NEB, M0357L), and 25.04 µL water) before incubation at 37 °C for 2 h. The BGT reaction was terminated by addition of 4.4 µL Proteinase K (NEB, P8107S) and incubation at 37 °C for 30 min. Purification was performed with 1.8 × SPRI beads and the DNA was eluted using 23 µL 1 mM Tris–HCl pH 8.0. DNA denaturation was performed through the addition of 2 µL 0.25 M freshly diluted NaOH, incubation at 50 °C for 10 min, and immediate cooling by quickly moving the samples onto an aluminum chill block at 0 °C. 25 µL of APOBEC reaction mix (1 µL APOBEC (NEB, E7125L), 5 µL APOBEC reaction buffer (NEB), 0.5 µL BSA (NEB), 1 µL 0.12 M freshly diluted HCl, and 16 µL water) was added while keeping the denatured product on the cold aluminum chill block. Deamination by APOBEC reaction was then performed through incubation at 37 °C for 16 h.

The DNA eluted for 5mC + 5hmC detection, underwent TET/BGT reaction instead of BGT reaction, which was performed as follows: 33-µL TET reaction mix (2.4 µL water, 18 µL reconstituted TET2 Reaction buffer (NEB), 1.8 µL Oxidation Supplement (NEB), 1.8 µL 100 mM DTT (NEB), 1.8 µL Oxidation Enhancer (NEB), 7.2 µL TET2 (NEB)) was added and mixed before the addition of 9 µL of 1:1250 diluted Fe(II) Solution (NEB) and incubation at 37 °C for 1 h, after which the reaction was stopped by the addition of 1.8 µL Stop Reagent (NEB) and incubation at 37 °C for 30 min. The TET/BGT reaction products were purified and APOBEC deaminated as described above for the detection of 5hmC, except that the APOBEC reaction time was reduced to 3 h.

Pre-amplification was then performed for both kinds of enzymatic conversion products as follows. After the APOBEC reaction, the first round of pre-amplification was performed by adding 50 µL 2 × KAPA HiFi HotStart Uracil + ReadyMix (Roche, KK2802), 1 µL 100 µM TSO_Read1_primer_V2 and 1 µL 100 µM V2-cDNA-amp-TSO + RNA primer, and running the following PCR program: 98 °C 45 s, 14 × [98 °C 15 s, 60 °C 30 s, 72 °C 3 min), 72 °C 3 min, and held at 4 °C. The second round of pre-amplification was performed by further adding 1 µL 100 µM TSO_Read2_primer_3U_C/T, and running the following PCR program: 98 °C 45 s, 4 × [98 °C 15 s, 60 °C 30 s, 72 °C 3 min), 72 °C 3 min, and held at 4 °C. The pre-amplification products were purified using 1.6 × SPRI beads and eluted in 52 µL of EB, with 24 µL of which headed for library preparation of 5-(hydroxy) methyl-cytosines detection, and another 24 µL of which headed for library preparation of RNA detection.

Libraries detecting 5-(hydroxy)methyl-cytosines were both prepared by adding 25 µL 2 × KAPA HiFi HotStart ReadyMix (Roche, KK2602), 0.5 µL 100 µM TSO custom P5 index primer and 0.5 µL 100 µM TSO custom P7 index primer (Additional file 2: Table S3) to the 24 µL pre-amplified product, and running the following PCR program: 98 °C 45 s, 5 × [98 °C 15 s, 60 °C 30 s, 72 °C 90 s), 72 °C 3 min, and held at 4 °C. Purification and size-selection were performed respectively with DNA Clean & Concentrator-5 (ZYMO, D4013) and Select-A-size DNA Clean & Concentrator (ZYMO, D4080, cutoff chosen at 150 bp) according to the manufacturer's guidelines.

For RNA library preparation, 24 µL of pre-amplified product from each enzymatic conversion path was pooled with its counterpart originating from the same source plate. For example, plate A was split into plate A1 (for 5mC + 5hmC detection) and plate A2 (for 5hmC detection), which were cell barcoded, pooled, purified, enzymatically converted in different ways, and pre-amplified as 2 conjugated samples, according to the methods described above; the pre-amplified products from plate A1 and A2, which both represent gene expression profiles of cells in plate A, were pooled to generate a combined 48-µL (24-µL each) product for subsequent RNA library preparation.

The combined 48-µL pre-amplified products were added 50 µL 2 × KAPA HiFi HotStart ReadyMix, 1 µL 100 µM TSO_Read1_primer_V2 and 1 µL 100 µM V2-cDNA-amp-TSO + RNA primer (Additional file 2: Table S1) before running the following PCR program: 98 °C 45 s, 5 × [98 °C 15 s, 60 °C 30 s, 72 °C 3 min),

72 °C 3 min, and held at 4 °C. This PCR product was purified using 1.2 × SPRI beads, and eluted in 100 µL EB. The elution was termed as cDNA Amp product, as it had been further enriched for RNA reads.

4 ng of the cDNA Amp product was transposed in a 10-µL reaction of 10 mM Tris–HCl pH 8.0, 5 mM MgCl2, 8% PEG-8000, 10.8 nM home-made Nextera_i7_Tn5 transposome by incubation at 55 °C for 10 min, and stopped in a 12-µL reaction as described above. 37 µL of Q5 mix (15.43 µL water, 10 µL Q5 reaction buffer (NEB), 10 µL Q5 High GC Enhancer (NEB), 0.07 µL 1 M MgCl2, 1 µL dNTP mix (NEB, N0447S, 10 mM each), 0.5 µL Q5 high-fidelity DNA polymerase (NEB)), 0.5 µL 100 µM TSO custom P5 index primer and 0.5 µL 100 µM Nextera P7 index primer (Additional file 2: Table S3), were added before performing the following PCR procedures: 65 °C 3 min, 98 °C 30 s, 8 × [98 °C 10 s, 60 °C 30 s, 72 °C 90 s], 72 °C 2 min, and held at 4 °C. The PCR products were added 54 µL EB and 56 µL SPRI beads for removal of the over-long fragments by retaining 150 µL of the supernatants, which were further added 20 µL SPRI beads, purified, and eluted by adding 20.5 µL EB. 20 µL of the elution was then collected as crude RNA library, which was pooled and further size-selected using Select-A-size DNA Clean & Concentrator (cutoff chosen at 300 bp) to get the final pooled RNA libraries.

We designed a custom balance library, which is generated by PCR using Balance_V3 oligo as the template, with Balance_s3P5 and Balance_s3P7 as primers, for improving the low nucleotide diversity caused by the fixed Tn5 ME 19 bp sequence following the cell barcode in the Read1 and the low-CG nature of (hydroxy)methyl-libraries that convert unmodified Cs to Ts (Additional file 2: Table S1). It must be noted that custom sequencing primers (Additional file 2: Table S1) were used for Illumina sequencing as required by the library structures. For sequencing that only involved the (hydroxy)methyl-libraries, all sequencing primers were replaced by the custom sequencing primers, while in the cases involving sequencing the RNA libraries, custom sequencing primers related to the P7 end were spiked in, as the RNA libraries had standard Nextera P7 ends. Libraries were first sequenced in a 2 × 150 bp mode on an Illumina MiSeq system for QC, with the custom balance library added at a ratio of 30%. The libraries passing QC were sequenced on an Illumina NovaSeq 6000 system in a 2 × 150 bp mode, using one S4 flow cell for around 24 (hydroxy)methyl-libraries or 288 RNA libraries, with the custom balance library added at a ratio of 20%.

Quantification and statistical analysis

Quality control

Raw sequencing reads of DNA and RNA for each single cell were processed following the Joint-Cabernet pipeline (Code is available on GitHub at: https://github.com/yunlongcaolab/MouseBrain_epiAtlas.git). Individual cell libraries were filtered as follows: DNA 5mC + 5hmC libraries require align rate > 20%, coverage > 0.001, pUC19 mCG ratio > 0.96, lambda mCG ratio < 0.01; DNA 5hmC libraries require align rate > 20%, coverage > 0.001, pUC19 mCG ratio < 0.02, lambda mCG ratio < 0.01, ClaI mCG ratio > 0.98; RNA libraries require align rate ≥ 20%, separated RNA reads ratio ≥ 70%, RNA reads ≥ 20,000, gene number ≥ 300, Qualimap Intronic ≥ 40%, Qualimap Intergenic ≤ 30%.

Transcriptome-based integration

We applied the Robust Principal Component Analysis (RPCA) method in Seurat (v5.1.0, https://github.com/satijalab/seurat) [56] to integrate snRNA-seq data from multiple brain regions, including the cortex, hippocampus, olfactory bulbs, and striatum, with the reference dataset from Yao et al.. The 'RunPCA' function was utilized to compute the top 20 principal components (PCs) based on variably expressed genes. Clustering was conducted on the integrated expression data using a shared-nearest-neighbor (SNN) graph-based approach with Louvain community detection, implemented through the 'FindClusters' function with a resolution of 1.8, before visualization

with Uniform Manifold Approximation and Projection (UMAP) [57]. After integration, each cell in a Seurat cluster was assigned the most prevalent original label of the cells from the reference dataset in that cluster.

DNA methylation around gene bodies with different expression levels

The median value of non-zero expression levels in the single-cell expression matrix from the reference dataset from Yao et al. was calculated, and genes were then divided into 3 groups according to their average expression levels in the reference dataset for each subclass: genes with zero expression, genes with expression no larger than the median, and genes with expression over the median. The average methylation levels around the gene bodies of the genes in the 3 groups were then calculated by applying the 'computeMatrix' function from deeptools [58] onto the subclass-level methylation BigWig files generated from ALLC files, with the following parameters: regionBodyLength = 5000, binsize = 5, upstream = 2000, and downstream = 2000, before plotting using the 'geom_line' function from the ggplot2 package (https://ggplot2.tidyverse.org) in R. We note that true 5mC levels, whether averaged across cells or within genomic regions, were deduced by subtracting the corresponding 5hmC levels from the corresponding 5mC + 5hmC levels before clipping negative values to 0.

Correlation between RNA expression and gene body DNA methylation

We calculated the Pearson correlation coefficients (PCCs) between RNA expression levels and gene body mean DNA methylation levels across cells, for all genes, with the *p*-values adjusted using the BH method. We performed 5000 random permutations of the sample labels to generate a null distribution of PCCs, from which permutation-derived *p*-values were calculated. Correlations with adjusted *p*-values < 0.05 were considered significant.

The 'geom_density_line' function from the R package ggplot2 was used to plot the density distribution of PCCs. The 'geom_pointdensity' function from ggplot2 was utilized to create scatter plots depicting the relationships between RNA expression and DNA methylation levels, with fitting lines generated by the 'stat_cor' function.

3D-compartment-level analysis of DNA modifications

We performed a comprehensive analysis of Hi-C data for different subclasses by applying the cooler package (v0.10.3, https://github.com/open2c/cooler) [59] onto the 100-kb-resolution contact matrix for each subclass from Liu et al.. The Hi-C matrices were normalized using the cooler balance tool. We then converted the processed data from COOL format to HDF5 format using hicConvertFormat (v3.0) [60], enabling efficient downstream analysis. To examine the structural variation of the contact matrix, we performed principal component analysis (PCA) using the hicPCA [60] tool. The first three principal components (PC1, PC2, and PC3) were extracted and saved in BEDGraph format. For further analysis, we classified the genome into two compartments (A and B) based on the sign of the PC1 values, and we filtered out any entries where PC1 was zero. Finally, we calculated the 5hmC, 5mC, and 5mC + 5hmC levels in CG and CH contexts across A/B compartments for each subclass.

Chromatin accessibility and DNA methylation dynamics

We obtained publicly available subclass-specific ATAC-seq chromatin regulatory elements (CREs). Open chromatin regions were defined as genomic intervals centered at CRE midpoints and extending 3 kb upstream and downstream. Closed chromatin regions were designated as consecutive 6 kb genomic intervals devoid of any CRE overlap. We then quantified the levels of 5hmCG, 5mCG, and the combined 5mCG + 5hmCG ratio in both open and closed chromatin regions across different subclasses.

D(H)MR analysis

First, we merged single-cell ALLC files into pseudo-bulk ALLC files according to the labels from clustering analysis, using the 'merge-allc' command from ALLCools (v1.0.8, https://github.com/lhqing/ALLCools). Next, we divided the genome into homogeneous methylated candidate regions for DMR-calling ('segments'), using the 'segment –min_cpg 3 –max_bp 2000' command from wgbstools (v0.2.0, https://github.com/nloyfer/wgbs_tools) with the subclass-level merged ALLC files. The mean CG methylation levels for the segments in individual cells were then generated with the 'generate-dataset' command from ALLCools. We then conducted DMR-calling in 3 stratified levels: (1) among excitatory neurons, inhibitory neurons, and non-neurons; (2) among all neuronal subclasses; (3) among all non-neuron subclasses. Segments with lengths of at least 200 bp then underwent DMR-calling using two-sided Wilcoxon rank-sum test by comparing cells from 1 group and cells from all other groups. Segments were regarded as DMRs if the BH-adjusted *p*-values were not above 0.05, the numbers of cells with non-NA values in the group were at least 10, and the mean methylation levels of the group differed from the robust means of all groups (the mean methylation levels of groups with methylation levels between 25 and 75th percentiles) by over + 0.3 (hyper) or below − 0.3 (hypo). The same process was carried out for DHMR analysis, except that the threshold of the differences of mean methylation levels from robust means were changed from ± 0.3 to ± 0.2 as the dynamic range of 5hmCG levels is smaller than 5mCG + 5hmCG levels. Annotation of the DMRs and DHMRs was carried out using the 'annotate_by_beds' method of the 'RegionDS' class from ALLCools with the parameter 'slop = 250'. For overlapping of DMRs and DHMRs with gene bodies and promoters, only genes with expression levels over 0.2 in at least 1 subclass and the coefficient of variation (CV) of expression levels among subclasses over 0.5, were used, and the overlapping thresholds were set to max(0.6*min(gene_length, D(H)MR_length), 500 bp (for overlapping gene bodies)/100 bp (for overlapping promoters)).

Linking distal regions with hyper-DHMR to genes through 3 C data

First, loops (dots) were called using matching neuronal subclass pseudo-bulk imputed 10-kb-resolution 3 C interaction matrixes generated through snm3C-seq with the command 'chromosight detect –min-dist 20,000 –max-dist 500,000' from chromosight [61]. Next, only loops with one anchor overlapping genes with expression levels over 0.2 in at least 1 subclass and the coefficient of variation (CV) of expression levels among subclasses over 0.4 (the genic anchor), and the other anchor not overlapping with any genes but having at least 1 hyper-DHMR within the 30-kb regions centering at this anchor (the intergenic anchor with hyper-DHMR), were preserved, linking distal regions with hyper-DHMR to genes. The 5hmCG levels of the 30-kb regions centering at the intergenic anchor with hyper-DHMR were then calculated through the 'generate-dataset' command from ALLCools. The 3 C interaction between the intergenic anchors and the linked genes were calculated by summing up the balanced 3 C interaction between the intergenic anchor and all the 10-kb bins overlapping with the gene. The Pearson correlation coefficients (PCCs) between the RNA expression levels of the linked genes and the 5hmCG levels in the distal regions, as well as the Pearson correlation coefficients (PCCs) between the RNA expression levels of the linked genes and the 3 C interaction between the intergenic anchors and the linked genes, were then calculated.

Histone modification analysis in hyper-DHMRs

We analyzed the enrichment of histone modifications in the hyper-DHMRs across different subclasses to explore their potential regulatory roles. The histone modifications examined included H3K4me1, H3K27ac, and H3K9me3. Using publicly available subclass-specific histone modification data (GSE152020), we quantified the signal intensity of these modifications in hyper-DHMRs.

Spatial projection of Joint-Cabernet cells using MERFISH data

We selected cells with z-values ranging from 7.33 to 7.34 and located in the brain regions of isocortex and hippocampus in the downloaded MERFISH data as reference data for spatial projection, and categorized the cells into excitatory neurons, inhibitory neurons, and non-neurons according to the 'subclass_transfer' information in the metadata. Accordingly, we categorized our Joint-Cabernet data of brain coronal sections into excitatory neurons, inhibitory neurons, and non-neurons by integrating with 10×RNA-seq data from Yao et al. The excitatory neurons, inhibitory neurons, and non-neuronal cells from those two sources were then integrated separately before performing label transfer. Spatial projection was then carried out by assigning the coordinates of the nearest MERFISH cell in UMAP coordinates with matching brain region and subclass, to Joint-Cabernet cells.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1186/s13059-026-04351-0.

Acknowledgements We thank Y. Chen and C. Geng at Peking University High-throughput Sequencing Center operated by BIOPIC for sorting of nuclei and assisting with Illumina Miseq sequencing. We used AI-assisted tools, including ChatGPT and DeepSeek, to help edit and polish the manuscript for language clarity.

Peer review information Wenfei Jin and Wenjing She were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions YC and XSX supervised the study. TY, YB, LR, YH and YC developed the Joint-Cabernet. YB, TY, LR, YH, FY, YL, AZ, YL, TT, NK, DC, XX, QW, WC and YZ generated and preprocessed the single-cell data. YB, TY, LR, YH, FY, AZ and YL performed bioinformatic analyses. YB, TY, LR and YC wrote the manuscript. All authors read and approved the final manuscript.

Funding This work is financially supported by National Natural Science Foundation of China (2023011477, YC) and Changping Laboratory (2025 B-06-01, YC).

Data availability Raw and processed data of Joint-Cabernet were deposited in GEO under the accession number GSE314268 [62]. The analysis code is available both via Zenodo at: https://doi.org/10.5281/zenodo.22141637 [63], and via GitHub at https://github.com/yunlongcaolab/MouseBrain_epiAtlas.git, under the MIT license [64]. The 10X scRNA-seq data is available at: https://allen-brain-cell-atlas.s3.us-west-2.amazonaws.com/index.html#expression_matrices/WMB-10Xv3/20230630/ [65]. The MERFISH data is available at: https://datasets.cellxgene.cziscience.com/f9f6cc29-7d06-47be-a798-e4e3b36a86b2.h5ad [66]. The scATAC-seq data is available in GEO under the accession number GSE246791 [67] and at: https://catlas.org/catlasv1/wholemousebrain/data/cCREs/ [68]. The histone modification data is available in GEO under the accession number GSE152020 [69]. The 3C data is available at: https://data.nemoarchive.org/biccn/grant/u19_cemba/ecker/epigenome/cellgroup/m3C-seq/mouse/processed/counts/ [70].

Declarations

Ethics approval and consent to participate This study was conducted in accordance with the ethical guidelines, with approval from the Institutional Animal Care and Use Committee for Rodents and Rabbits of Changping Laboratory (approval number: CPNL-AP-2025-0006).

Consent for publication Not applicable.

Competing interests YC, XSX, TY and YB are inventors on a patent filed by Changping Laboratory that covers template-switching-based single-cell DNA methylation and hydroxymethylation sequencing, which constitutes part of the Joint-Cabernet method described in this work. This patent is directed to the scale-up of the method and does not impact the reproducibility or use of the method by other researchers.

References

1. Yao Z, Liu H, Xie F, Fischer S, Adkins RS, Aldridge AI, et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. Nature. 2021;598:103–10.
2. Yao Z, van Velthoven CTJ, Kunst M, Zhang M, McMillen D, Lee C, et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. Nature. 2023;624:317–32.
3. Siletti K, Hodge R, Mossi Albiach A, Lee KW, Ding SL, Hu L, et al. Transcriptomic diversity of cell types across the adult human brain. Science. 2023;382:eadd7046.
4. Jorstad NL, Song JHT, Exposito-Alonso D, Suresh H, Castro-Pacheco N, Krienen FM, et al. Comparative transcriptomics reveals human-specific cortical features. Science. 2023;382:eade9516.
5. Liu H, Zhou J, Tian W, Luo C, Bartlett A, Aldridge A, et al. DNA methylation atlas of the mouse brain at single-cell resolution. Nature. 2021;598:120–8.
6. Liu H, Zeng Q, Zhou J, Bartlett A, Wang BA, Berube P, et al. Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. Nature. 2023;624:366–77.
7. Tian W, Zhou J, Bartlett A, Zeng Q, Liu H, Castanon RG, et al. Single-cell DNA methylation and 3D genome architecture in the human brain. Science. 2023;382:eadf5357.
8. Li YE, Preissl S, Hou X, Zhang Z, Zhang K, Qiu Y, et al. An atlas of gene regulatory elements in adult mouse cerebrum. Nature. 2021;598:129–36.
9. Zu S, Li YE, Wang K, Armand EJ, Mamde S, Amaral ML, et al. Single-cell analysis of chromatin accessibility in the adult mouse brain. Nature. 2023;624:378–89.
10. Li YE, Preissl S, Miller M, Johnson ND, Wang Z, Jiao H, et al. A comparative atlas of single-cell chromatin accessibility in the human brain. Science. 2023;382:eadf7044.
11. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324:929–30.
12. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324:930–5.
13. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333:1300–3.
14. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333:1303–7.
15. Wu X, Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet. 2017;18:517–34.
16. Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell. 2012;151:1417–30.
17. Szulwach KE, Li X, Li Y, Song CX, Wu H, Dai Q, et al. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat Neurosci. 2011;14:1607–16.
18. Wen L, Li X, Yan L, Tan Y, Li R, Zhao Y, et al. Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. Genome Biol. 2014;15:R49.
19. Mellen M, Ayata P, Heintz N. 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. Proc Natl Acad Sci U S A. 2017;114:E7812-21.
20. Yildirim O, Li R, Hung JH, Chen PB, Dong X, Ee LS, et al. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell. 2011;147:1498–510.
21. Zhao J, Gu T, Gao C, Miao G, Palma-Gudiel H, Yu L, et al. Brain 5-hydroxymethylcytosine alterations are associated with Alzheimer's disease neuropathology. Nat Commun. 2025;16:2842.
22. Villar-Menendez I, Blanch M, Tyebji S, Pereira-Veiga T, Albasanz JL, Martin M, et al. Increased 5-methylcytosine and decreased 5-hydroxymethylcytosine levels are associated with reduced striatal A2AR levels in Huntington's disease. Neuromolecular Med. 2013;15:295–309.
23. Wang F, Yang Y, Lin X, Wang JQ, Wu YS, Xie W, et al. Genome-wide loss of 5-hmC is a novel epigenetic feature of Huntington's disease. Hum Mol Genet. 2013;22:3641–53.
24. Kozlenkov A, Li J, Apontes P, Hurd YL, Byne WM, Koonin EV, et al. A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. Sci Adv. 2018;4:eaau6190.
25. Heffel MG, Zhou J, Zhang Y, Lee DS, Hou K, Pastor-Alonso O, et al. Temporally distinct 3D multi-omic dynamics in the developing human brain. Nature. 2024;635:481–9.
26. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell. 2012;149:1368–80.
27. Booth MJ, Ost TW, Beraldi D, Bell NM, Branco MR, Reik W, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc. 2013;8:1841–51.
28. Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science. 2012;336:934–7.

29. Liu Y, Siejka-Zielinska P, Velikova G, Bi Y, Yuan F, Tomkova M, et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. Nat Biotechnol. 2019;37:424–9.
30. Schutsky EK, DeNizio JE, Hu P, Liu MY, Nabel CS, Fabyanic EB, et al. Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. Nat Biotechnol. 2018. https://doi.org/10.1038/nbt.4204.
31. Fullgrabe J, Gosal WS, Creed P, Liu S, Lumby CK, Morley DJ, et al. Simultaneous sequencing of genetic and epigenetic bases in DNA. Nat Biotechnol. 2023;41:1457–64.
32. Bai D, Zhang X, Xiang H, Guo Z, Zhu C, Yi C. Simultaneous single-cell analysis of 5mC and 5hmC with SIMPLE-seq. Nat Biotechnol. 2025;43:85–96.
33. Viswanathan R, Cheruba E, Wong PM, Yi Y, Ngang S, Chong DQ, et al. DARESOME enables concurrent profiling of multiple DNA modifications with restriction enzymes in single cells and cell-free DNA. Sci Adv. 2023;9:eadi0197.
34. Fabyanic EB, Hu P, Qiu Q, Berrios KN, Connolly DR, Wang T, et al. Joint single-cell profiling resolves 5mC and 5hmC and reveals their distinct gene regulatory effects. Nat Biotechnol. 2024;42:960–74.
35. Cao Y, Bai Y, Yuan T, Song L, Fan Y, Ren L, et al. Single-cell bisulfite-free 5mC and 5hmC sequencing with high sensitivity and scalability. Proc Natl Acad Sci U S A. 2023;120:e2310367120.
36. Mulqueen RM, Pokholok D, O'Connell BL, Thornton CA, Zhang F, O'Roak BJ, et al. High-content single-cell combinatorial indexing. Nat Biotechnol. 2021;39:1574–80.
37. Di L, Fu Y, Sun Y, Li J, Liu L, Yao J, et al. RNA sequencing by direct tagmentation of RNA/DNA hybrids. Proc Natl Acad Sci U S A. 2020;117:2886–93.
38. Luo C, Liu H, Xie F, Armand EJ, Siletti K, Bakken TE, et al. Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. Cell Genom. 2022. https://doi.org/10.1016/j.xgen.2022.100107.
39. Aimone JB, Li Y, Lee SW, Clemenson GD, Deng W, Gage FH. Regulation and function of adult neurogenesis: from genes to cognition. Physiol Rev. 2014;94:991–1026.
40. Lim DA, Alvarez-Buylla A. The adult ventricular-subventricular zone (V-SVZ) and olfactory bulb (OB) neurogenesis. Cold Spring Harb Perspect Biol. 2016. https://doi.org/10.1101/cshperspect.a018820.
41. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. Global epigenomic reconfiguration during mammalian brain development. Science. 2013;341:1237905.
42. Luo C, Keown CL, Kurihara L, Zhou J, He Y, Li J, et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science. 2017;357:600–4.
43. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326:289–93.
44. Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol. 2011;29:68–72.
45. Han D, Lu X, Shih AH, Nie J, You Q, Xu MM, et al. A highly sensitive and robust method for genome-wide 5hmC profiling of rare cell populations. Mol Cell. 2016;63:711–9.
46. Koike N, Kassai Y, Kouta Y, Miwa H, Konishi M, Itoh N. Brorin, a novel secreted bone morphogenetic protein antagonist, promotes neurogenesis in mouse neural precursor cells. J Biol Chem. 2007;282:15843–50.
47. He B, Zhang C, Zhang X, Fan Y, Zeng H, Liu J, et al. Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. Nat Commun. 2021;12:4249.
48. Zhu C, Zhang Y, Li YE, Lucero J, Behrens MM, Ren B. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nat Methods. 2021;18:283–92.
49. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015;348:aaa6090.
50. Zhang M, Eichhorn SW, Zingg B, Yao Z, Cotter K, Zeng H, et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. Nature. 2021;598:137–43.
51. Zhang M, Pan X, Jung W, Halpern AR, Eichhorn SW, Lei Z, et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. Nature. 2023;624:343–54.
52. Pinheiro PS, Perrais D, Coussen F, Barhanin J, Bettler B, Mann JR, et al. GluR7 is an essential subunit of presynaptic kainate autoreceptors at hippocampal mossy fiber synapses. Proc Natl Acad Sci U S A. 2007;104:12181–6.
53. Henkemeyer M, Itkis OS, Ngo M, Hickmott PW, Ethell IM. Multiple EphB receptor tyrosine kinases shape dendritic spines in the hippocampus. J Cell Biol. 2003;163:1313–26.
54. Tan L, Ma W, Wu H, Zheng Y, Xing D, Chen R, et al. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. Cell. 2021;184:741-758 e717.
55. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science. 2018;360:176–82.
56. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42:293–304.
57. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018. https://doi.org/10.1038/nbt.4314.

58. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. DeepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44:W160-165.
59. Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics. 2020;36:311–6.
60. Wolff J, Backofen R, Gruning B. Loop detection using Hi-C data with HiCExplorer. Gigascience. 2022. https://doi.org/10.1093/gigascience/giac061.
61. Matthey-Doret C, Baudry L, Breuer A, Montagne R, Guiglielmoni N, Scolari V, et al. Computer vision for pattern detection in chromosome contact maps. Nat Commun. 2020;11:5795.
62. Bai Y, Yuan T, Ren L, Huan Y, Yang F, Li Y, et al. Single-cell joint profiles of DNA methylation, hydroxymethylation and RNA expression in the adult mouse brain. Gene Expression Omnibus. 2026. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE314268.
63. Bai Y, Yuan T, Ren L, Huan Y, Yang F, Li Y, et al. Analysis code related to Joint-Cabernet profiles in the mouse brain. Zenodo. 2026. https://doi.org/10.5281/zenodo.22141637.
64. Bai Y, Yuan T, Ren L, Huan Y, Yang F, Li Y, et al. Analysis code related to Joint-Cabernet profiles in the mouse brain. GitHub. 2025. https://github.com/yunlongcaolab/MouseBrain_epiAtlas.
65. Yao Z, van Velthoven CTJ, Kunst M, Zhang M, McMillen D, Lee C, et al. RNA atlas of the mouse brain. Allen Brain Cell Atlas. 2023. https://allen-brain-cell-atlas.s3.us-west-2.amazonaws.com/index.html#expression_matrices/WMB-10Xv3/20230630/.
66. Zhang M, Pan X, Jung W, Halpern AR, Eichhorn SW, Lei Z, et al. Spatial MERFISH atlas of the mouse brain. CELLxGENE. 2023. https://datasets.cellxgene.cziscience.com/f9f6cc29-7d06-47be-a798-e4e3b36a86b2.h5ad.
67. Zu S, Li YE, Wang K, Armand EJ, Mamde S, Amaral ML, et al. Subclass-specific chromatin accessibility bigwig profiles in the adult mouse brain. Gene Expression Omnibus. 2023. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE246791.
68. Zu S, Li YE, Wang K, Armand EJ, Mamde S, Amaral ML, et al. Subclass-specific cCREs identified by chromatin accessibility in the adult mouse brain. CATlas. 2023. https://catlas.org/catlasv1/wholemousebrain/data/cCREs/.
69. Zhu C, Zhang Y, Li YE, Lucero J, Behrens MM, Ren B. Subclass-specific histone modifications bigwig profiles in the mouse brain. Gene Expression Omnibus. 2020. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152020.
70. Liu H, Zeng Q, Zhou J, Bartlett A, Wang BA, Berube P, et al. Subclass-specific mcool files describing the 3C profiles in the mouse brain. NEMO Archive. 2023. https://data.nemoarchive.org/biccn/grant/u19_cemba/ecker/epigenome/cellgroup/m3C-seq/mouse/processed/counts/.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Yali Bai (白雅丽) [1] · Tianjiao Yuan (袁天骄) [1] · Liuhao Ren (任刘豪) [2,3] · Yuhang Huan (郇宇航) [1,2] · Fa Yang (杨发) [1,2] · Yutong Li (李雨桐) [1,2] · Aicui Zhang (张爱翠) [1] · Yueyang Liu (刘悦洋) [1] · Tian Tian (田甜) [1] · Ningxia Kang (康宁霞) [1] · Xuchao Xie (谢戌超) [1] · Qiang Wang (汪强) [1] · Danni Chen (陈丹妮) [1] · Wei Chen (陈炜) [1] · Yinghui Zheng (郑英慧) [2] · Xiaoliang Sunney Xie (谢晓亮) [1,2] · Yunlong Cao (曹云龙) [1,2,3]

✉ Xiaoliang Sunney Xie (谢晓亮)
sunneyxie@biopic.pku.edu.cn

✉ Yunlong Cao (曹云龙)
yunlongcao@pku.edu.cn

[1] Changping Laboratory, Beijing 102206, China

[2] Present Address: Biomedical Pioneering Innovation Center (BIOPIC), School of Life Sciences, Peking University, Beijing 100871, China

[3] Peking-Tsinghua Center for Life Sciences, Beijing, China

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/ .