RESEARCH

A ternary-code DNA methylome atlas of mouse tissues

Sol Moe Lee[1], David C. Goldberg[1], Cameron Cloud[1], Jared B. Parker[2], Christopher Krapp[3], Christian E. Loo[4], Elliot Kim[1], Ivan Zhao[1], Chengcheng Jin[5], Rishi Porecha[6], Marisa S. Bartolomei[3], Rahul M. Kohli[2*] and Wanding Zhou[1,7*]

*Correspondence:
rkohli@pennmedicine.
upenn.edu; wanding.
zhou@pennmedicine.upenn.edu

[1] Center for Computational
and Genomic Medicine,
The Children's Hospital
of Philadelphia, Philadelphia, PA
19104, USA
[2] Department of Medicine,
University of Pennsylvania,
Philadelphia, PA 19104, USA
[3] Department of Cell
and Developmental Biology,
Epigenetics Institute, University
of Pennsylvania Perelman School
of Medicine, Philadelphia, PA
19104, USA
[4] Graduate Group
in Biochemistry and Biophysics,
University of Pennsylvania,
Philadelphia, PA 19104, USA
[5] Department of Cancer Biology,
University of Pennsylvania
Perelman School of Medicine,
Philadelphia, PA 19104, USA
[6] Illumina, Inc., San Diego, CA
92122, USA
[7] Department of Pathology
and Laboratory Medicine,
University of Pennsylvania,
Philadelphia, PA 19104, USA

Abstract

Background: DNA cytosine modifications, including 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), are key epigenetic regulators with distinct functions. Dissecting the ternary code (C, 5mC, 5hmC) across tissues and cell types remains a critical priority due to the limitations of traditional profiling methods based on bisulfite conversion.

Results: Here, we leverage the combined bisulfite and enzymatic (bACE) conversion with the Mouse Methylation BeadChip to generate 265 base-resolution ternary-code modification maps of 5mC and 5hmC across 29 mouse tissue types spanning 8–76 weeks of age and both sexes. Our atlas reveals a complex grammar of 5hmC distribution, jointly shaped by cell mitotic activity, chromatin states, and interplay with 5mC at the same and neighboring CpG sites. Of note, we demonstrate that 5hmC significantly complements 5mC-based biomarkers in delineating cell identity in both brain and non-brain tissues. Each modification state, including 5hmC alone, accurately discriminates tissue types, enabling high-precision machine learning classification of epigenetic identity. Furthermore, the ternary methylome variations extensively implicate gene transcriptional variation, with age-related changes correlated with gene expression in a tissue-dependent manner.

Conclusions: Our work reveals how tissue, sex, and age jointly govern the dynamics of the two cytosine modifications, augments the scope of DNA modification biomarker discovery, and provides a reference atlas to explore epigenetic dynamics in development and disease.

Keywords: DNA methylation, Hydroxymethylation, Epigenetics, Mouse, Cell identity, Aging, Transcription regulation

Background

Modifications of cytosine bases in mammalian genomic DNA are crucial for regulating cellular and organismal functions [1] and are increasingly utilized as diagnostic biomarkers [2]. Among these modifications, 5-methylcytosine (5mC) in the CpG dinucleotide

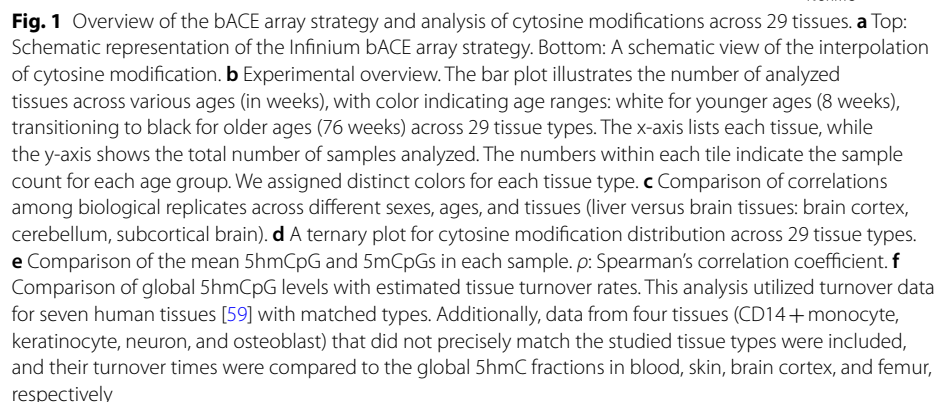Lee *et al. Genome Biology*   (2025) 26:343

Page 2 of 30

context is the most common, with 5-hydroxymethylcytosine (5hmC), an oxidative derivative, following in prevalence [3]. 5mC marks are introduced and maintained by DNA methyltransferases (DNMTs), enabling transmission of the epigenetic profiles across cell divisions and over the course of tissue development [4]. TET enzymes can oxidize 5mC to 5-hydroxymethylcytosine (5hmC), which must be reestablished after cell division. 5hmC can be further oxidized to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), although the levels of 5fC and 5caC are orders of magnitude below those of 5hmC [5–7]. 5hmC has been shown to play a critical role in diverse processes, including maintaining pluripotency [8, 9], neuronal and somatic tissue development [10–12], aging [13, 14], brain disorders [15, 16], and human cancers [2, 17, 18].

While 5mC and 5hmC are increasingly recognized as antagonistic cytosine modifications with distinct biological roles [19, 20], a comprehensive analysis of their dynamics across various tissue types and age groups is lacking. Prior studies based on bisulfite sequencing (BS-seq), enzymatic methyl sequencing (EM-seq), and APOBEC-coupled epigenetic sequencing (ACE-seq) have generally assessed their combined signal as 5-modified cytosine (5modC) or focused on either one of the two modifications in isolation. Notably, the extensive use of bisulfite conversion in DNA methylome profiling, which does not discriminate between 5hmC and 5mC [21], has led to the largest body of knowledge on 5modC in human and mouse tissues [22–24] and cell types [25–27]. Several base resolution methods have also been developed to look at 5mC alone [28–30] and 5hmC alone [31–33]. These methods have contributed to understanding the 5hmC landscape in neuronal tissues and related disorders [20]—given the high prevalence of 5hmC in neurons [3, 34, 35]—as well as in embryonic stem cells [36, 37] and various human tissues [11, 12]. However, these investigations often studied 5hmC in isolation and are typically based on a limited number of subjects. Recently, bulk and single-cell co-assays of 5mC and 5hmC have emerged [38–40]. However, due to the high profiling cost, the application of these methods has been restricted to specific tissue types, age groups, or sexes, limiting the comprehensive exploration of their profiles across biologically variable samples.

Given the critical biological roles of cytosine modifications, there has been increasing interest in developing a reference atlas across diverse cell types and tissues. Recent work with bisulfite-based approaches has demonstrated that 5modC can be used to define cell identity with high precision [25–27]. This capability has enabled successful applications such as cell type deconvolution from heterogeneous tissues [41], cancer classification [42], liquid biopsy, and lineage tracking [43, 44]. Besides approaches that combine 5mC and 5hmC into the 5modC signal, 5hmC alone has been proposed and applied in translational contexts, including liquid biopsy, due to its potential to delineate cell states [45–49]. Given the distinct yet complementary roles of 5mC and 5hmC, these established efforts suggest that a ternary-code modification atlas (C, 5mC, 5hmC) would likely provide a more precise and nuanced molecular footprint of cell identity than either modification alone.

Beyond defining cell identity, resolving 5hmC can help elucidate its role in gene regulation. At base resolution, 5hmC levels typically range from 0 to 20% across cells [3], with tissue- or age-related differences often below 5%, necessitating high quantitative precision in assays. Previous studies [38, 50] have also shown that 5hmC can be

Lee *et al. Genome Biology*   (2025) 26:343

Page 3 of 30

highly focal, potentially impacting transcription factor (TF) binding and cis-regulation of gene expression at enhancers [51]. Thus, high-depth analysis parsing of 5mC and 5hmC at information-rich CpG sites offers the potential promise of new insights into tissue-specific gene regulation.

Here, to newly dissect ternary-code cytosine modifications across diverse tissues, we utilized the Infinium Mouse Methylation BeadChip [23, 52, 53] in tandem with Bisulfite-assisted APOBEC-Coupled Epigenetic (bACE) conversion (Fig. 1a). We hypothesized that the Infinium BeadChips would be well-suited for scalable ternary methylome profiling, as they are highly quantitative (reflecting ~ 100 × sequencing depth [54]) and were designed to target enhancers and other 5hmC-enriched genomic

Fig. 1 Overview of the bACE array strategy and analysis of cytosine modifications across 29 tissues. **a** Top: Schematic representation of the Infinium bACE array strategy. Bottom: A schematic view of the interpolation of cytosine modification. **b** Experimental overview. The bar plot illustrates the number of analyzed tissues across various ages (in weeks), with color indicating age ranges: white for younger ages (8 weeks), transitioning to black for older ages (76 weeks) across 29 tissue types. The x-axis lists each tissue, while the y-axis shows the total number of samples analyzed. The numbers within each tile indicate the sample count for each age group. We assigned distinct colors for each tissue type. **c** Comparison of correlations among biological replicates across different sexes, ages, and tissues (liver versus brain tissues: brain cortex, cerebellum, subcortical brain). **d** A ternary plot for cytosine modification distribution across 29 tissue types. **e** Comparison of the mean 5hmCpG and 5mCpGs in each sample. ρ: Spearman's correlation coefficient. **f** Comparison of global 5hmCpG levels with estimated tissue turnover rates. This analysis utilized turnover data for seven human tissues [59] with matched types. Additionally, data from four tissues (CD14 + monocyte, keratinocyte, neuron, and osteoblast) that did not precisely match the studied tissue types were included, and their turnover times were compared to the global 5hmC fractions in blood, skin, brain cortex, and femur, respectively

Lee *et al. Genome Biology*   (2025) 26:343

Page 4 of 30

regions. Furthermore, the bACE array strategy permits a parallel workflow to resolve 5hmC from 5modC [55, 56]. In this approach, samples first undergo bisulfite conversion and are then subsequently split, with one half directly prepared for assay to jointly profile 5mC and 5hmC (5modC) and the other half receiving additional APOBEC deamination treatment to distinguish 5hmC alone. This discrimination is possible because, in bisulfite-converted DNA, 5mC undergoes enzymatic deamination, while 5hmC, which bisulfite converts to cytosine-5-methylenesulfonate (CMS), resists deamination. The resulting profiles of 5modC and 5hmC then allow for the fraction of 5mC alone to be derived by subtraction. Unlike previous array adaptations that depend on computational correction of background signals [57, 58], the bACE array thus provides base-resolution, direct measurements of 5hmC, adding this critical dimension to help reveal the ternary code.

Using this approach, we generated a comprehensive ternary-code methylation atlas from C57BL/6 J mice, spanning 29 tissue types, both sexes, and an age range of 8–76 weeks, allowing population-scale analysis of these epigenetic marks. This mouse model is well-suited for atlas construction, given its well-characterized genetic background, rich phenotype characterization, and extensive use in prior epigenetic studies [3, 60]. Our atlas offers foundational insights into cytosine modifications in mice and establishes principles likely translatable to humans, providing a robust resource for exploring cell identity and epigenetic regulation across species.

Results

A ternary-code DNA methylome atlas of C57BL/6 J mouse tissues

We collected 265 mouse tissue samples, encompassing 29 tissue types from 32 C57BL/6 J mice (Fig. 1b). DNA was extracted from each tissue sample and underwent bisulfite and bACE conversion before Infinium array profiling (Fig. 1a). As 5hmC fractions are frequently low and subject to signal background influence on the Infinium array readout, we leveraged a reference DNA set titrating methylation across the full range [23] to derive a standard interpolation curve for each probe. Preprocessed array data were then calibrated using these standards to obtain direct measurements of %5modC (5mC + 5hmC) and %5hmC (Fig. 1a, Methods). To benchmark accuracy at single-CpG resolution, we compared our bACE array dataset with published TAB-seq and ACE-seq profiles from mouse cortex, a tissue with relatively high 5hmC levels. We observed strong concordance with Spearman correlation coefficients reaching 0.78, higher than the 0.65 correlation observed between the ACE-seq and TAB-seq datasets themselves (Additional file 1: Fig. S1a). To further test whether this agreement extends across a broader dynamic range of 5hmC abundance, we next compared global %5hmC levels in four tissue types using the bACE array and orthogonal measurements from low-pass ACE-seq (Methods). In these comparisons, interpolated (corrected) array values aligned more closely with ACE-seq measurements (Additional file 1: Fig. S1b). Pairwise Wilcoxon tests confirmed that uncorrected data differed significantly from ACE-seq ($p < 1.5e - 10$), whereas corrected values were not statistically distinguishable ($p = 0.057$; Additional file 1: Fig. S1c). Although this does not establish strict equivalence, it indicates that interpolation markedly improves concordance with the reference sequencing method. Furthermore, our datasets across the entire cohort showed stronger correlations among biological

replicates than among samples with different covariates (sex, age), confirming the reproducibility of our approach (Fig. 1c).

With our comprehensive profiles across the 265 samples in hand, we first focused on the global differences in C, 5mC, and 5hmC across tissues. Given the specific importance of modifications in CpG contexts, we plotted the average total %CpG modification levels for each sample in a ternary plot, where the three values sum to unity (Fig. 1d). Across tissues, the levels of cytosine modifications range from 40.8 to 58.2% for unmodified C, 36.3 to 58.2% for 5mC, and 0.5 to 14.6% for 5hmC. Similar tissues were observed to cluster together, with 5hmC level adding a dimension that distinguished tissues in these ternary plots. Examination of specific tissues revealed that neuron-rich tissues, such as the subcortical brain and cortex, showed the highest 5hmC levels, consistent with prior measurements using liquid chromatography-mass spectrometry (LC–MS/MS) to quantify DNA modifications [61]. In contrast, immune-related tissues like blood and thymus exhibited the lowest levels of 5hmC (Fig. 1d, Additional file 1: Fig. S1d). Notably, our focus on CpG modifications was substantiated by examining probes in the non-CpG (CpH) context. Overall, the 5hmCpH fraction was generally low in most tissues, with the exception of neuronal tissues, where some CpA sites exhibited relatively higher levels of 5hmC compared to other tissue types (Additional file 1: Fig. S1e).

Given the ability to parse global 5modC into its component 5mC and 5hmC levels, we next excluded unmodified C from consideration and examined correlations between global 5hmC and 5mC levels. Across tissues, we noted that 5hmC levels are negatively correlated with 5mC (Fig. 1e), likely reflecting the fact that 5hmC is generated from 5mC. At the same time, tissues with similar levels of 5mC can differ significantly in their 5hmC content. For example, colonic tissue shows an average of 1.8% 5hmC while eye samples show 8% 5hmC, although both tissues have 44% 5mC on average (Additional file 1: Fig. S1f, red). Conversely, similar global 5modC levels can be associated with widely divergent compositions of 5mC and 5hmC. For example, both the subcortical brain and blood have global 5modC levels around 52–58%, yet 23.5% of the 5modC is accounted for by 5hmC in the brain, while only 2% of 5modC is 5hmC in the blood (Additional file 1: Fig. S1f, gray). In considering factors contributing to the global 5hmC prevalence, we noted that global tissue levels of 5hmC correlated positively with tissue turnover time (Spearman's rho: 0.82, Fig. 1f), potentially aligning with the non-heritable nature of 5hmC over cell division and its accumulation with longer-lived cells. Collectively, global levels of unmodified C, 5mC, and 5hmC varied across tissues, with 5hmC levels highest in neuron-rich tissues and lowest in immune tissues, consistent with a positive association with tissue turnover time.

Local neighborhoods around 5mC and 5hmC sites

We next moved from examining global 5mC and 5hmC to understanding the local neighborhood context around sites where these modifications are prevalent. In this analysis, the two modifications show distinctive profiles. For 5mC, we examined CpGs neighboring other highly modified CpGs (> 94%). These neighboring CpGs also show correlated high methylation levels, with 5mC levels that average > 80% within 200 bp, likely reflecting DNMT processivity [62] (Fig. 2a). By contrast, when focusing on highly 5hmC-modified sites (> 20%), the scale of spatial 5hmC correlation was shorter, potentially indicating
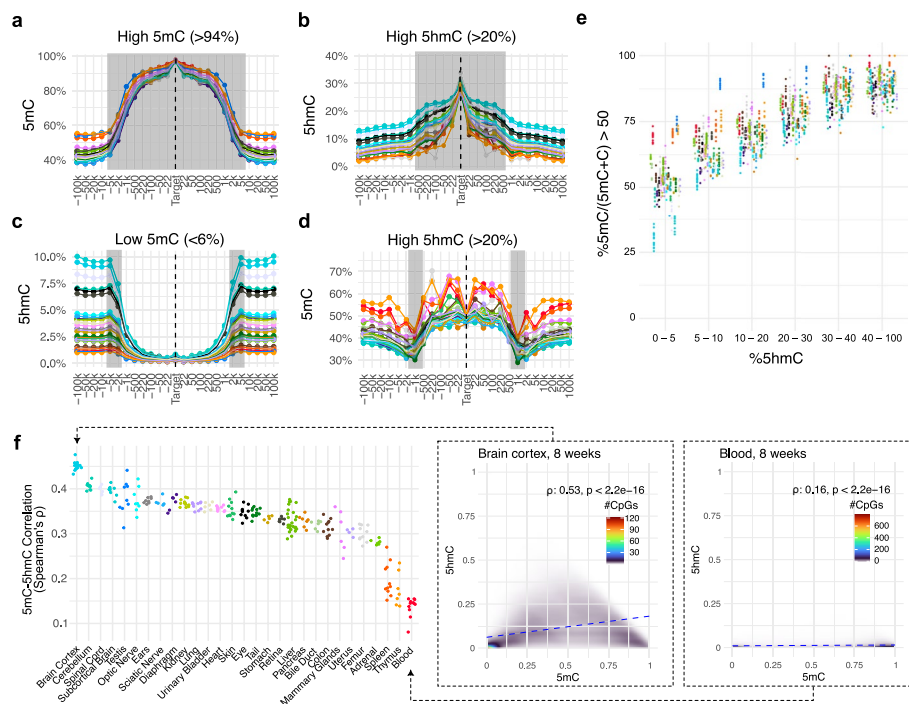
Lee *et al. Genome Biology*   (2025) 26:343

Page 6 of 30

Fig. 2 Local neighborhoods around 5mC and 5hmC sites. Distribution of **a** 5mC levels with high 5modC content (over 94%) and **b** 5hmC levels with high 5hmC content (over 20%). The color of points and lines represents tissues. Distribution of **c** 5hmC level with low 5mC content (less than 6%) and **d** 5mC level with high 5hmC (over 20%). **e** 5hmCpG fraction correlation with the 5mCpG. Each dot represents a sample. The x-axis shows the mean proportion of 5hmCpG, while the y-axis represents the percentage of methylated sites, defined by the ratio, of 5mC over the sum of 5mC and C, greater than 0.5. **f** Spearman correlation between 5mC and 5hmC across CpG probes for each sample, with two example scatter density plots illustrating the correlation between 5mC and 5hmC showing a representative brain cortex (206402260033_R05C01) and a blood (206402260033_R05C01) sample. The dashed blue line represents the fitted linear regression. Spearman's correlation coefficient (ρ) and corresponding *p* value are displayed at the top

weaker TET processivity [63] (Fig. 2b). Furthermore, while the neighborhood around prevalent 5mC tends to show similar profiles across tissues, CpGs neighboring existing 5hmCs exhibited greater variability across tissues, suggesting that 5hmC marks are more localized, particularly in tissues with lower global levels (Fig. 2b). More generally, we found that both 5mCs and 5hmCs were inversely correlated with local CpG density (Additional file 1: Fig. S2a, b). We also examined the cross-talk between the two modifications in local neighborhoods. CpGs with low 5mC were associated with an absence of 5hmC in the neighborhoods extending up to 2–5 kbps away (Fig. 2c), likely reflecting the substrate constraint that 5hmC must be generated from 5mC. Conversely, CpGs with high 5hmCs were associated with high 5mCs at immediate neighbors but were often proximal to regions with low 5mC (~ 1 kbps), suggesting that TETs act at sites that are locally enriched for 5mC but adjacent to regulatory chromatin (Fig. 2d). In contrast, high 5modC alone did not predict 5hmC in surrounding regions, indicating TET activity is not solely directed by 5mC abundance in a neighborhood (Additional file 1: Fig. S2c).

With a sense of the local neighborhoods for 5mC and 5hmC, we also looked to understand tradeoffs in the distribution of modifications at given individual CpGs. Across the data sets, we observe that sites with 5hmC tended to have more 5mC than unmodified cytosine (Fig. 2e). Sites with low 5hmC carry different levels of 5mC depending on the

tissue (Fig. 2e). They tend to be unmodified in neuronal tissues and are more associated with 5mC in immune-related tissues such as blood, thymus, and spleen (Fig. 2e). Consistent with these findings, Spearman correlation analysis between 5mC and 5hmC profiles within in each sample revealed positive relationships between the two modifications, most prominent in neuronal tissues exhibiting high global 5hmC levels (Fig. 2f). Additionally, while studies using bisulfite conversion have often suggested that 5modC tends to have a bimodal distribution at individual CpGs, we find that the 5mC pattern is contingent on 5hmC. When 5hmC is low, the fraction of 5mC within the combined pool of 5mC and unmodified C also tends to have a bimodal distribution. This fraction trends toward a unimodal distribution as the 5hmC fraction increases (Additional file 1: Fig. S2d, e). Together, these patterns suggest that 5hmC dynamically tracks 5mC, particularly in neuronal tissues where this coupling is more pronounced. In contrast, in proliferative tissues, the distribution of 5hmC appears more independent of 5mC, indicating distinct regulatory functions.

In summary, 5mC and 5hmC show distinct spatial correlation patterns: 5mC extends more broadly, while 5hmCs are locally confined. 5hmC distribution depends on 5mC as the substrate for oxidative generation of 5hmC, a relationship more pronounced in long-lived tissues.

5mC-5hmC interplay: chromatin and gene context

Given that tissues exhibit distinct ternary-code profiles, we next focused on understanding how the distribution of 5mC and 5hmC differs as a function of genomic elements in the different tissues. To explore global ternary-code profiles by chromatin states, we intersected each CpG with consensus chromatin states from an 18-state mouse model (Methods), which was derived from 66 ENCODE chromHMM calls [64]. Overall, the genomic 5hmC distribution suggests enrichment at actively transcribed gene bodies and enhancers (Fig. 3a, Additional file 1: Fig. S3a), consistent with prior reports [65]. While this pattern held across tissue types, the relative distributions varied by tissue. Brain cortex, cerebellum, and liver exhibited higher levels of 5hmC in gene bodies than enhancers, while most other tissue types skewed toward higher levels in enhancers specifically (Fig. 3b). As another approach to assess the relative prevalence of 5hmC as a function of genomic elements, we examined sites with high 5modC (> 90%) across chromatin states and parsed the relative contribution of 5hmC (Fig. 3c). In neuronal tissues, 5hmC contributed up to 14% of cytosine modifications at gene bodies, dropping to ∼ 1% in proliferating tissues like blood, spleen, and thymus. Furthermore, while 5modC is rarely found at promoter and enhancer CpG sites, 5hmC can contribute significantly (up to 23% and 26%, respectively) to the total modification signal (Fig. 3c).

Next, we focused on more precisely defining the distribution of 5hmC and 5modC around gene bodies. Both cytosine modifications are more prevalent in gene body DNA than in intergenic regions (Additional file 1: Fig. S3b, c). 5modC tends to be slightly biased toward the 3′-end of the gene bodies (Additional file 1: Fig. S3b), while 5hmC is more frequently found at the 5′-end of gene bodies closer to promoters, suggesting regulatory roles (Fig. 3d, Additional file 1: S3c). To link expression levels to the distribution of modifications, we examined the most and least expressed genes from brain and liver tissue. In both tissues, the 5modC patterns tend to be similar for the most expressed
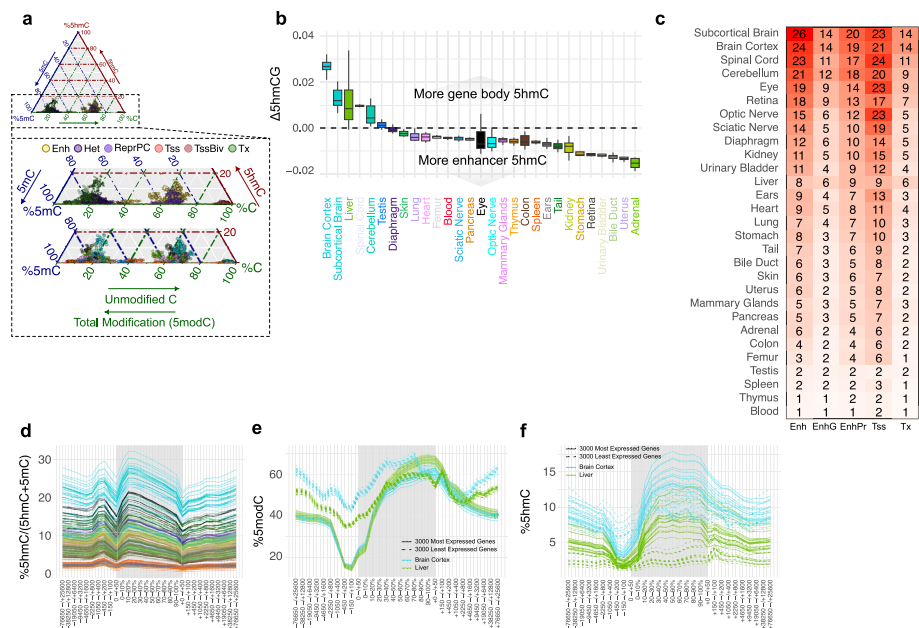
Lee *et al. Genome Biology*   (2025) 26:343

Page 8 of 30

Fig. 3 5mC-5hmC interplay: chromatin and gene context. **a** Distribution of cytosine modifications across 29 tissue types categorized by chromatin states, with colors representing chromatin states (top) and tissues (bottom). **b** Distribution of global 5hmCpG levels between gene body and enhancer regions. The x-axis is arranged based on the difference in 5hmCpG fractions between these regions. **c** The numbers in each tile represent the fraction of the mean beta value of 5hmCpGs. The mean value was calculated for CpG probes with high 5modC levels (greater than 0.9). **d** The mean beta value distribution of covered CpGs in each bin for % of 5hmC/5modC around transcription start sites (TSS) and transcription termination sites (TTS). Distribution of mean beta values for covered CpGs in each bin for **e** 5modC and **f** 5hmC around the TSS and TTS in brain cortex and liver tissues. Transcriptome data were obtained from [66] (Methods)

genes, with more promoter modifications in the least expressed genes, consistent with the canonical epigenetic silencing model of promoter methylation (Fig. 3e). However, with 5hmC, genes with higher expression in both tissues are associated with fewer promoter modifications but more gene body modifications than genes with lower expression (Fig. 3f). In sum, 5hmC and 5mC show distinct genomic and tissue distributions, with 5hmC contributing substantially to total modification in a gene expression-dependent manner.

The global governors of the ternary DNA methylomes

Having examined the global, regional, and genome element-specific cytosine modification levels, we next aimed to integrate across the entire ternary-code methylome profile to understand the contribution of tissue, age, and sex in defining methylome profile similarity. To this end, we visualized the ternary code as 2D t-SNE embeddings encompassing each of the three modification forms separately (Fig. 4a). With each modification, tissue type was the predominant influence on sample clustering, followed by age and sex, as quantified by entropy-based uncertainty coefficients (Fig. 4b). Hierarchical clustering across 29 tissues suggested that each of the cytosine modifications is also tracked by developmental lineage, indicating that tissues are separated according to their embryonic germ layer origins, cytosine modification levels, and turnover rates. For example, using any single modification (unmodified C, 5mC, or 5hmC), neuronal and
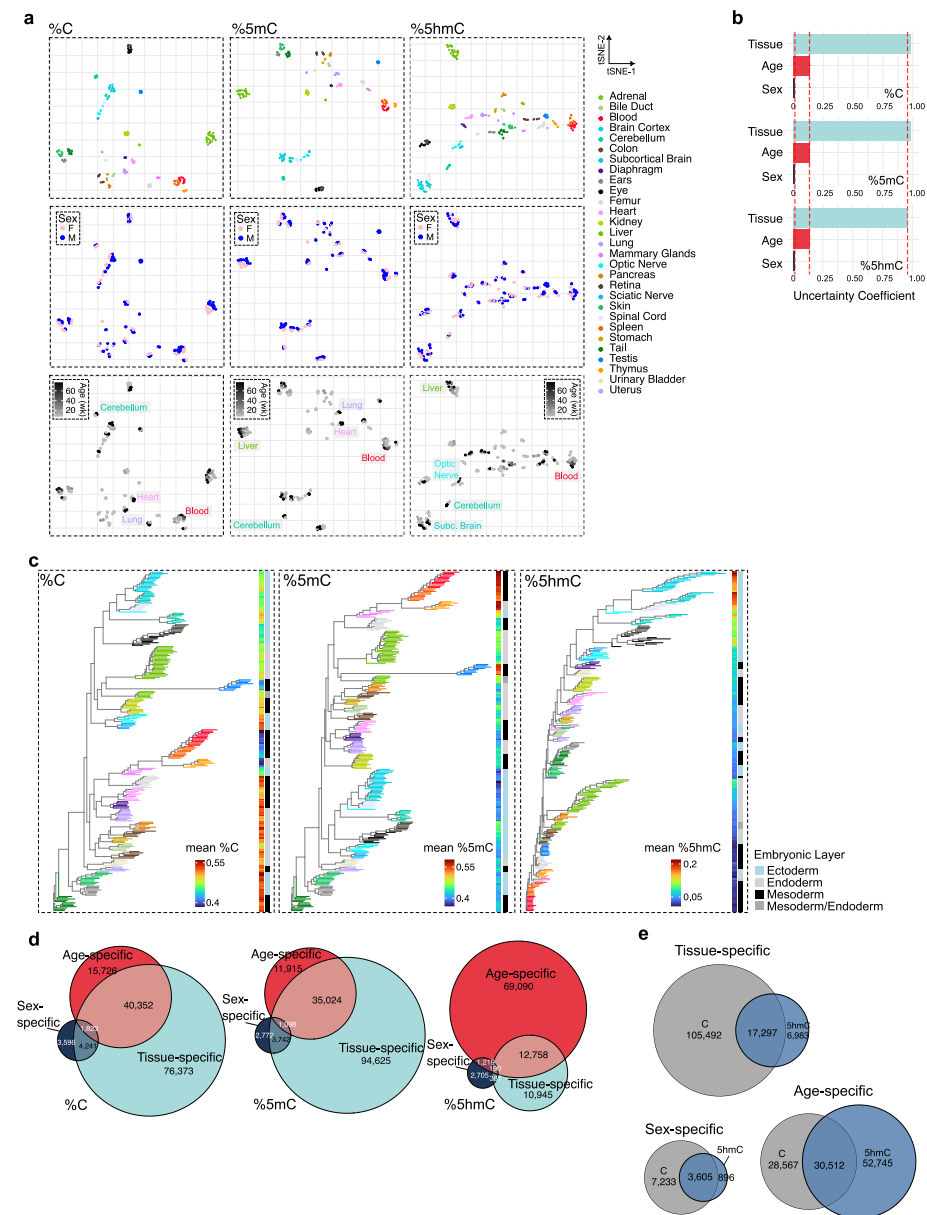
Fig. 4 The global governors of the ternary DNA methylomes. **a** t-SNE maps of unmodified cytosine (left), 5mC (middle), and 5hmC (right) across 29 tissue types, where each point represents a sample. The order from top to bottom reflects tissue type, sex, and age. **b** Uncertainty coefficients for predicting sample clustering membership based on tissue type, age, and sex for unmodified cytosine (top), 5mC (middle), and 5hmC (bottom). The uncertainty coefficient represents the proportion of total information in clustering explained by random discrete variables. **c** Unsupervised hierarchical clustering of mouse tissues based on levels of unmodified cytosine (left), 5mC (middle), and 5hmC (right). **d** Comparison of age, sex, and tissue-specific modifications in 5modC (left), 5mC (middle), and 5hmC (right). **e** Venn diagrams for the overlap of tissue-specific (top), sex-associated (bottom left), and age-associated (bottom right) modifications between 5modC and 5hmC. Analyses in **d** and **e** were performed on 12 tissue types (liver, brain cortex, blood, skin, eye, femur, heart, lung, testis, optic nerve, cerebellum, and subcortical brain)

immune-related tissues formed distinct clusters consistent with t-SNE patterns (Fig. 4c). Focusing on 5modC and 5hmC, we also examined which genomic elements might be contributing to clustering. As with 5modC, tissue specificity was most defined by 5hmC

at gene bodies and enhancers and less so by promoter status (Additional file 1: Fig. S4a–c).

As noted, although tissue identity is the dominant factor separating samples, age and sex are also contributing factors. With each of the three modifications, within each tissue cluster, samples of different ages separate from one another (Fig. 4a). While sex-based separation was also evident across samples, greater separation is associated with unmodified C and 5mC than with 5hmC (Fig. 4a, Additional file 1: Fig. S4d).

Given that tissue, sex, and age can all contribute to shaping the ternary-code methylome, we sought to parse CpGs that make a sample distinctive based on each modification state. To this end, we performed a CpG locus-based multivariate regression on each contributing factor. The number of significant CpGs specific to each predictor is consistent with the order of tissue, age, and then sex (Fig. 4d). Notably, a substantial number of CpGs are subject to influence by two or even all three factors (Fig. 4d). Focusing on CpGs defined by unmodified C (equivalent to those defined by 5modC) and 5hmC, 13.3% ($N = 17{,}297$) of tissue-specific modifications are common to both unmodified C and 5hmC, while 5.4% ($N = 6983$) of tissue-specific modifications are unique to 5hmC (Fig. 4e). In contrast, 5hmC features a greater proportion of age-specific modifications, with 47.2% ($N = 52{,}745$) of the total age-associated CpGs unique to 5hmC, likely reflecting age-dependent accumulation of 5hmC in tissues with longer-lived cells (Fig. 4e).

In summary, tissue type, sex, and age are global determinants of cytosine modification states in the ternary-code methylome, with each factor influencing CpG sets that are substantially overlapping yet also distinct.

5hmC augments cytosine modification-based cell identity definition

Bisulfite-based 5modC (and linked unmodified C) signals have been used as molecular barcodes for cell identity [27]. To evaluate whether 5hmC alone can also delineate tissue identity, we developed predictive models of tissue types based strictly on 5hmC profiles. We selected the 2000 most variable 5hmC features to define the model using a training subset of the data ($N = 73$) and achieved 100% accuracy in cross-validation test cohorts ($N = 32$) (Fig. 5a). The most important feature subset, according to SHapley Additive exPlanations (SHAP) values, consists of CpGs exhibiting 5hmC marks unique to specific tissues or limited tissue subsets (Additional file 1: Fig. S5a). Overall, this analysis demonstrates that 5hmC alone can readily reveal cell identities as effectively as total cytosine modifications.

Modifications uniquely reflecting tissue identity were characterized by either hyper- or hypo-modifications that distinguish the target tissue from all other tissue types, which we collectively refer to as the *one-vs-rest* or OvR sites. Using a nonparametric linear discriminant analysis, we identified many such sites from our diverse tissue atlas (Fig. 5b, c, Additional file 1: Fig. S5b). In contrast to 5modC, where there are more hypomodified tissue-defining sites (defined by unmodified C), 5hmC signatures predominantly appear as hyper-hydroxymethylation (Fig. 5c). Notably, distinct OvR 5hmC signatures are still observed in tissues such as blood and thymus (Fig. 5c), suggesting a role for 5hmC even with tissues with globally low 5hmC levels. When parsing 5mC from 5modC, the 5mC-only signatures were similar to those from 5modC (Additional file 1: Fig. S5b), confirming that unmodified C is the primary contributor to the OvR 5modC signatures.
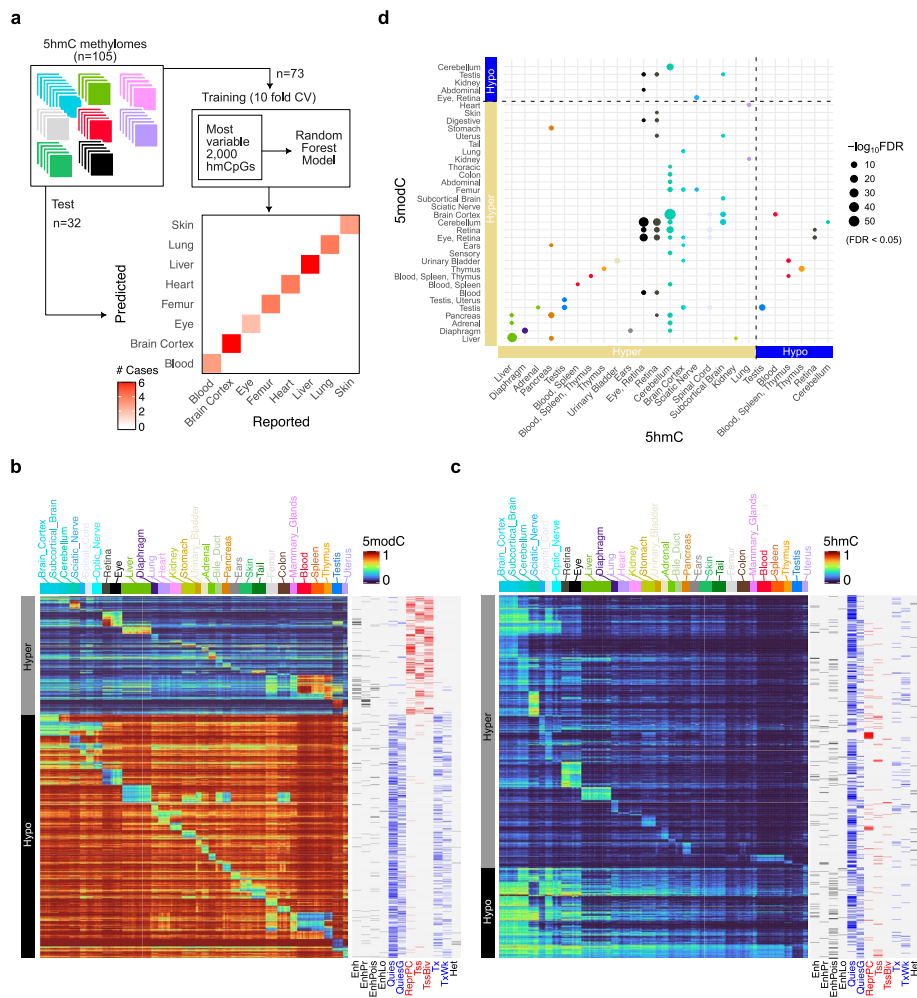
Fig. 5 5hmC augments cytosine modification-based cell identity definition. **a** Schematic illustration of the random forest model workflow for tissue prediction using the 5hmC dataset from 8 tissue types (Methods). The lower section of the figure shows a heatmap comparing predicted versus actual cases from the test set. Tissue-specific DNA methylation patterns displaying CpGs (rows) that are hypermethylated (red) or hypomethylated (blue) in each tissue (columns organized by tissue type) for **b** 5modC and **c** 5hmC. **d** Enrichment analysis of the tissue signatures for 5hmC and 5modC

Neuronal tissues with high 5hmC were an exception to this overall pattern, with the 5mC-only signal diminished relative to 5modC, indicating that hyper-5modC signatures in neuronal tissues, in fact, largely originate from the presence of 5hmC (Fig. 5c). We also observed that OvR sites segregate by chromatin states in a predictable fashion. For example, OvR hyper-5modC signatures were enriched in regions with low baseline methylation in most tissues, such as gene promoters (Tss) and H3K27me3-marked regions (TssBiv and ReprPC) that are marked by red bars in Fig. 5b. Conversely, OvR hypo-5modCs were found in highly methylated regions, including gene bodies (Tx and TxWk) and quiescent regions (Quies and QuiesG), all denoted by blue bars in Fig. 5b. In contrast, OvR hyper-5hmC signatures were often enriched in quiescent regions, suggesting a dependence on the baseline 5mC for 5hmC generation. A direct comparison of 5hmC and 5modC OvR signatures revealed an extensive overlap of the corresponding

hyper-modification signatures across most tissues, suggesting a previously overlooked role of 5hmC in hyper-5modC-based cell identity footprints (Fig. 5d).

In sum, we showed that 5hmC alone can reliably predict tissue identity. Even tissues with globally low 5hmC, such as blood, harbor distinct 5hmC signatures at CpGs not marked by 5mC. This demonstrates that 5hmC contributes independently and augments cytosine methylation-based definitions of cell identity.

5hmC delineates cell identity at TF binding vicinities of lineage-defining genes

We further queried the OvR signature distributions for mechanistic links to tissue biology. Specifically, we were interested in reconciling the relationship between 5hmC and the TF binding. Based on our earlier analysis, 5hmCs are enriched at enhancers. However, they are also distinctively localized from hypo-5mC, which is commonly associated with direct TF binding [67, 68]. By focusing on hyper-5hmCs and hypo-5modCs, each representing the more common tissue-specific changes for 5hmC and 5modC, respectively (Fig. 5b, c) [69], we probed for their association with TF binding and nearby associated genes.

Significant overlaps with TF binding sites (TFBS) and gene sets were more frequent in 5modC signatures than in 5hmC signatures (Fig. 6a, b, Additional file 1: Fig. S6a–c). For 5modCs, overlaps with specific TFBSs were more predominant (Fig. 6a, triangles), followed by other target gene set associations, likely reflecting the dependence of TF binding directly on the presence or absence of 5mC [68]. For 5hmC, tissue-specific hyper-5hmCs were less associated with the direct TF bindings (Fig. 6a). Nonetheless, their genomic locations are still proximal to genes marking tissue functions (Fig. 6a, b) and are enriched in tissue-specific chromatin states, as evident in the liver, cerebellum, and other tissues (Fig. 6c).

Next, we more closely examined the relationship between cytosine modifications and TFBSs, focusing on 5hmC distribution. Figure 6d highlights two example TFs: RXRA-B in the liver and NEUROD2 in the brain. As expected, 5mC is low at the binding sites for both examples. However, for RXRA, we also observe that 5hmC is enriched near RXRA-B binding sites, supporting the model that 5hmC can localize adjacent to, but not directly at, binding sites, which often prefer unmodified cytosines. Notably, NEUROD2 in the brain demonstrates how this enrichment is not universal. Here, we observe minimal 5hmC enrichment around NEUROD2 binding sites (Fig. 6d, bottom row), which likely reflects alternative mechanisms through which 5hmC accumulates in longer-lived tissues.

Integrating across modification states, the enrichment patterns can reveal complex transcriptional networks composed of genes with tissue-specific functions and expression profiles (Fig. 6e–h, Additional file 1: Fig. S6d–f). To highlight a few notable findings, liver-specific modification changes capture hepatocyte differentiation and function, as reflected by the loss of 5modC at binding sites of metabolic regulators such as RXRA [70] and the hyper-5hmC marking of key liver metabolism genes such as *Akr1c6* in prostaglandin metabolism [71, 72] (Fig. 6e). Cerebellum-specific modification changes are associated with the binding of neurogenic regulators like NEUROD2, NEUROG2, and MECP2 (Additional file 1: Fig. S6d). Eye-specific modification alteration implicates vision-related cell maintenance, as exemplified by modification loss at the binding sites

Lee *et al. Genome Biology* (2025) 26:343

Page 13 of 30

Fig. 6 5hmC delineates cell identity at TF binding vicinities of lineage-defining genes. **a** A scatter plot showing enrichment of tissue-specific 5modC (left) and 5hmC (right) by genes, TFBSs, and gene sets. **b** Gene Ontology (GO) term enrichment analysis for tissue-specific hyper-5hmCs for liver (top left), blood and spleen (top right), cerebellum (bottom left), and eye (bottom right) tissues. **c** ChromHMM enrichment analysis for hypermethylated tissue-specific 5hmC (top) and hypomethylated tissue-specific 5modC (bottom). **d** Genomic distribution of 5mC level (left) and 5hmC (right) centered on RXRA-B binding sites (top) and NEUROD2 binding sites (bottom). Dot plots illustrating the enrichment of hypo (loss) and hyper (gain) methylation patterns around TFBSs and genes within 10,000 bp of each target CpG for **e** liver, **f** eye, **g** blood and spleen, and **h** kidney

of CRX, a key regulator of photoreceptor differentiation [73, 74], and the hyper-5hmC marking of *Poc1b*, associated with retinopathy [75] (Fig. 6f). Similarly, *Dmbt1*, a modulator gene of the gastrointestinal functions [76–78], is associated with pancreas-specific

hyper-5hmC marking (Additional file 1: Fig. S6e). These observations extend to multiple tissues, such as blood, spleen, kidney, and thymus (Fig. 6g, h, Additional file 1: Fig. S6f).

Some observed regulatory patterns show an association of cytosine modifications between the TF gene locus and its corresponding binding site. For instance, immune-specific modifications are enriched in *Pax5*, a key regulator of B cell development (Fig. 6g). The *Pax5* gene exhibits hyper-5modC, while binding sites for this TF show hypo-5modC, reflecting a regulatory circuit where TF expression and its genomic binding activity are coordinately controlled [79–81].

A close examination suggests that while some observed regulatory patterns are influenced by a single cytosine modification form (Additional file 1: Fig. S6e, f), others are coordinated through mechanisms involving both 5mC and 5hmC dynamics. For example, MECP2, a key methylation reader, shows cerebellum-specific patterns of both 5modC and 5hmC (Additional file 1: Fig. S6d). When the MECP2 binding sites are marked by 5hmCs, the nearby gene expression tends to be activated compared to when the binding sites are marked by 5mCs (Additional file 1: Fig. S6g), consistent with a previous study [82]. In sum, our analysis of tissue-specific OvR signatures shows that 5modC is directly associated with transcription factor binding, whereas 5hmC tends to localize near binding sites and genes with tissue-specific functions. Together, these patterns reveal complex regulatory networks where tissue-specific functions are captured by combinations of 5mC and 5hmC dynamics, sometimes acting independently and sometimes coordinately.

Cytosine modification alterations are linked to gene expression changes

To explore the transcriptional ramifications of 5mC and 5hmC variations, we next analyzed the correlation between published transcriptome profiles of 16 tissue types and the gene-specific 5mC and 5hmC levels quantified in those matched tissues [66]. In most tissues, a global negative correlation was observed between 5mC and gene expression (Fig. 7a). The main exception to this was in Polycomb-repressed regions and heterochromatin, where 5mC showed a positive correlation with the expression of nearby genes (within ~ 10 kb). The positive association between 5mC and gene expression at H3K27me3-marked regions has been reported previously, likely representing a switch between two repressive epigenetic marks [83, 84]. In contrast, 5hmC levels showed opposite, positive correlations with expression, particularly at gene bodies and enhancers (Fig. 7b). Unexpectedly, in tissues with higher global 5hmC, such as the brain tissues, 5hmC is less correlated with gene expression (Fig. 7b, Additional file 1: Fig. S7a). This result suggests that in non-proliferating tissues, much of the 5hmC does not play a regulatory role in gene expression (Fig. 7b), instead reflecting a steady accumulation of 5hmC at non-specific 5mC sites not counteracted by cellular replication in post-mitotic cells. Despite being rare in heterochromatin, 5hmC remains positively correlated with RNA expression (Fig. 7b, Additional file 1: Fig. S7b), whereas 5mC showed a negative correlation (Fig. 7a, Additional file 1: Fig. S7c). This suggests that the presence of 5hmC can be even more predictive of gene expression in broad heterochromatic neighborhoods.

We also focused on whether base-resolution distributions of 5hmC and 5mC are linked with tissue-specific expression. As it is well established that 5modC levels are inversely correlated with transcription, and 5hmC levels are generally less prevalent,
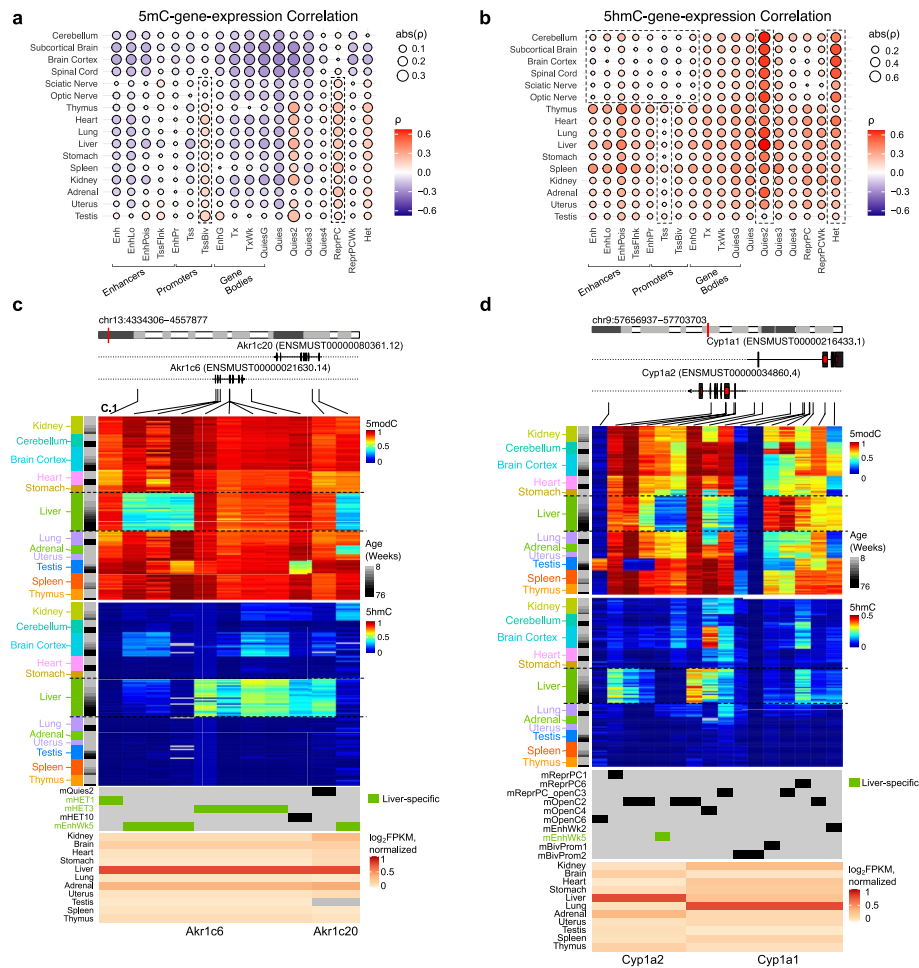
Fig. 7 Cytosine modification alterations are linked to gene expression changes. Dot plot showing the correlation between RNA expression (from 6-week samples) and **a** 5mC and **b** 5hmC levels (from 8-week samples) across genes. The size of each dot reflects the absolute value of Spearman's rho, while the color represents Spearman's rho score. A track view for **c** *Akr1c6* and **d** *Cyp1a2*. The heatmap was organized from top to bottom, displaying 5modC, 5hmC, full-stack chromHMM chromatin state, and RNA expression levels

we primarily analyzed the 5modC loss (gain of unmodified C) and the 5hmC gain at specific CpGs across genes of interest. Multiple tissue-specific 5modC/5hmC showed correlation with tissue-specific gene expression. Example genes, e.g., *Akr1c6* and *Cyp1a2* in the liver (Fig. 7c and d), showed coordinated variations in their epigenomic and transcriptomic profiles, with regional change in 5modC and 5hmC being liver-specific. Similar patterns can be observed in other representative genes, such as *Cd8b1* in the thymus (Additional file 1: Fig. S7d), *Cdh16* in the kidney (Additional file 1: Fig. S7e), and *Rbfox1* and *Prkce* in the brain (Additional file 1: Fig. S7f, g).

Integrating across these examples, when associated with gene expression, tissue-specific 5hmCs tend to be more pervasively distributed around the gene bodies (Fig. 7c, d, Additional file 1: Fig. S7d–g). In contrast, tissue-specific 5modC absence remains more focal and localized. Exceptions to this rule are CpGs in quiescent genomic territories, which are heavily modified but depleted in 5hmC (Fig. 7c, C.1).

Lee *et al. Genome Biology*   (2025) 26:343

Page 16 of 30

The total modification and 5hmC differences take place at different CpGs, constituting complementary molecular definitions of the epigenetic identity (Fig. 7c, d).

It is worth noting that marker genes do not always carry tissue-specific changes in both 5modC or 5hmC (Additional file 1: Fig. S7h). For example, *Adcy7* and *Acox3*, marking spleen and kidney, respectively, carry tissue-specific 5modC, but the presence of 5hmC is not obvious from our profiles (Additional file 1: Fig. S7i, j). *Adamts20* marking brain lacks tissue-specific 5modC but exhibits clear tissue-specific 5hmC (Additional file 1: Fig. S7k). Additionally, tissue-specific 5hmC is not always indicative of gene expression. This is best epitomized by the brain and aged liver tissues, which often exhibit 5hmC at non-expressing genes, again likely to reflect the global accumulation of 5hmC in long-lived cells (Fig. 7d, Additional file 1: Fig. S7e, i, j).

Here, we found that, in general, 5mC negatively and 5hmC positively correlate with gene expression. However, 5hmC accumulation can occur at non-expressing genes and is uncoupled from transcription in long-lived tissues. We confirmed that 5modC and 5hmC provide complementary tissue characterizations.

Ternary-code epigenetic aging implicates tissue functional maturation and immune dynamics

Age is known to shape the epigenome, as demonstrated by epigenetic aging clocks that can track both chronological and biological age [85, 86]. In our mouse cohort, tissues were collected from a wide age range of mice, including at least one 76-week-old mouse per sex. This diverse age range enabled us to investigate the effects of aging on DNA modification patterns, with our approach allowing us to distinguish the roles of 5mC from 5hmC (Fig. 8a). The established paradigm of DNA methylation dynamics associated with mitotic aging is epitomized by the gain of modifications at Polycomb repressive complex targets and their loss at late-replicating genome [87, 88]. We first validated this established paradigm as an orthogonal check on our

(See figure on next page.)

Fig. 8 Ternary-code epigenetic aging via tissue functional maturation and immune dynamics. **a** A ternary plot for global C, 5mC, and 5hmC levels of samples with circle sizes indicating mouse ages. **b** ChromHMM enrichment analysis of gain of 5modC levels (left) and gain of 5hmC levels (right). Log$_2$ odds ratios less than −2 are capped. **c** Aging-related enrichment analysis for the gain of 5hmC in the liver during aging. (bottom) The x-axis represents age-related gain 5modC probes, and the probes are organized based on the age-related slope coefficient values. To examine the age-related contribution of 5hmC, only CpGs with low 5modC (~ 10%) in the youngest (8-week-old) mice were selected. The y-axis shows each probe's age-related slope coefficients from linear regression (unit: week). (top) The x-axis aligns with the bottom plot, and the y-axis exhibits the enrichment score (ES). Each bar in the figure depicts the location of age-associated 5hmC gains. **d** Scatter plots illustrating the correlation between levels of 5hmC (left) and 5modC (right) in the brain cortex (top) and liver (bottom) at two different developmental stages: 8 weeks (a representative sample) and 76 weeks (a representative sample). The color gradient indicates the density of data points. The red diagonal line represents a 1:1 correlation, and the percentages in the upper left and lower right corners indicate the proportion of data points falling above and below this line, respectively. **e** Scatter plots illustrating the correlation between changes in RNA expression and 5hmC levels and GO term enrichment analysis for four tissue types (liver, heart, brain cortex, and skin). The red dashed lines represent the axes with no change in RNA expression and 5hmC. The numbers in each quadrant's corners denote the percentage of data points within that quadrant. The y-axis in scatter plots represents age-associated RNA expression (log$_{10}$ of (normalized counts + 1)), while the x-axis shows estimates from a linear regression model examining changes in 5hmC levels with aging. Bar charts illustrate the enrichment of GO terms associated with metabolic processes in the four tissue types
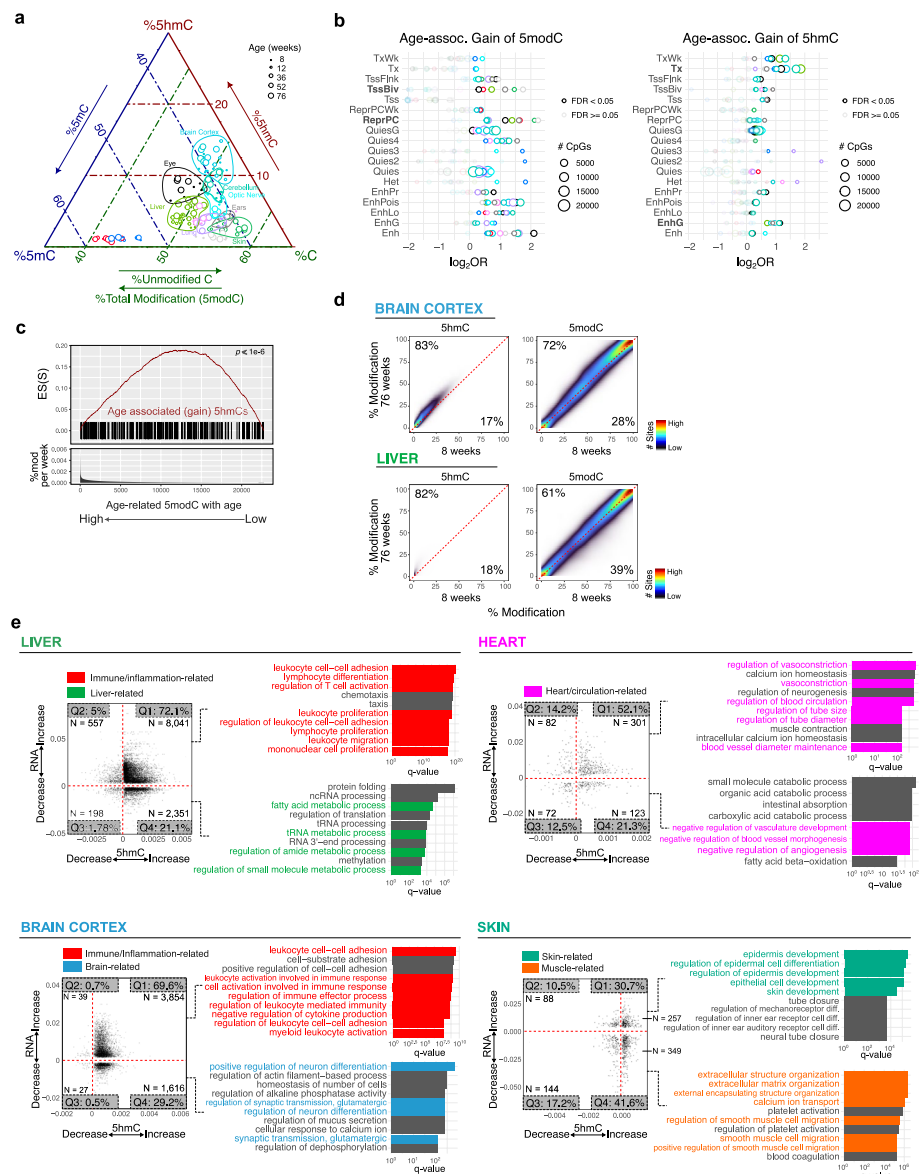
Fig. 8 (See legend on previous page.)

data. Indeed, we observe gains of 5modC at Polycomb targets (Fig. 8b) and global loss of 5modC in aged tissues with high turnover rates, e.g., the skin (horizontal axis in Fig. 8a, Additional file 1: Fig. S8a) [89]. Notably, the global 5modC trend is reversed in tissues with slow proliferation, such as the brain cortex, cerebellum, and eye, where 5modC increases with age (Fig. 8a).

To explain the tissue discrepancies, we dissected the two modification forms in 5modC and queried whether age-associated changes in 5modC can be driven by 5hmC. First, we observed that in tissues with slow proliferation, global 5hmC levels indeed increase with age (Fig. 8a, the 5hmC dimension), with this 5hmC enriched in gene bodies (Fig. 8b, right). To rigorously test whether increased age-associated 5hmC drives age-associated gain of 5modC, we performed an enrichment analysis

of CpGs with both modification changes in liver samples, which provided the most extensive sampling across the different ages. Our results confirmed that the age-related increases in 5modC significantly overlap with sites showing an increase in 5hmC during aging (Fig. 8c). However, Polycomb targets (ReprPC) showed low 5hmC enrichment, suggesting a limited contribution of 5hmC gain to 5modC gain in these regions (Fig. 8b, Additional file 1: Fig. S8b, c). Interestingly, sites of higher 5hmC levels in young tissue are linked to even greater increases in 5hmC in older tissues, a pattern more pronounced when examining 5hmC alone than the combined 5modC signal (Fig. 8d).

To query links to tissue function, we compared age-associated changes in cytosine modifications with age-associated gene expression changes across various tissue types, focusing on the less well-characterized 5hmC. Using public transcriptome data from mouse brain, liver, heart, dorsal skin, and lung tissues [90], we compared the rate of 5hmC changes per CpG, as measured by the linear regression slope coefficients, with the expression changes in the linked genes during aging. In all tissues aside from the lung, we observed that 5hmC correlates positively with gene expression, with concordant increases or decreases with age (Fig. 8e, Additional file 1: Fig. S8d). The changes frequently reflect tissue development and functions. For example, age-associated 5hmC gains decreased gene expression involved in hepatocyte differentiation in the liver and neuronal function in the brain, while supporting epidermis development in the skin (Fig. 8e). In the heart, 5hmC gains influence genes involved in blood circulation in both directions. The few reduced 5hmCs in liver aging can also be associated with liver function change, as seen with the *Tdo2* gene, where both RNA expression and 5hmC levels decline with age (Additional file 1: Fig. S8e). In bulk tissues, 5hmC dynamics may also reflect shifts in tissue composition. For example, gene sets that show gained 5hmCs and lost gene expression (Fig. 8e, the fourth quadrant of the skin panel) are linked to muscle development, likely due to increased subcutaneous stromal contamination in samples [91, 92]. Similarly, liver aging is associated with increased 5hmC levels and RNA expression in immune-related genes (Fig. 8e, the first quadrant of the liver panel). In fact, immune system hallmarks are also observed across brain tissue and lungs, potentially mediated by chronic tissue inflammation, a universal aging hallmark (Fig. 8e, Additional file 1: Fig. S8d).

In sum, our analysis revealed age-dependent 5hmC dynamics across multiple tissues, suggesting its contribution to the established aging paradigm of total cytosine modifications. These changes correlate with transcriptional programs linked to tissue development, function, and inflammation, underscoring the complex epigenetic regulation of aging trajectories.

Discussion

In this study, we profiled the genomic landscape of C, 5mC, and 5hmC to generate a comprehensive atlas of the ternary code spanning different tissues across ages and in both sexes. Our study builds on the distinctive strengths of the Infinium arrays and a combined chemical and enzymatic sequencing approach. Our atlas suggests that in proliferative tissues where global 5hmC is low, enhancer 5hmCs can still be high and biologically meaningful. It highlights the value of dissecting 5modC signals into their

component 5mC and 5hmC parts in all tissues, informing on the regulatory roles of each modification in gene expression and their utility as molecular barcodes of both cell identity and tissue aging.

Strength of Infinium BeadChips in profiling ternary methylomes

In contrast to most previous profiling efforts to parse 5hmC based on high-throughput sequencing methods—such as TAB-seq [32] and ACE-seq [31] for measuring 5hmC alone, or OxBS-seq [28] and DM-seq [30] for measuring 5mC alone—this study employs Infinium arrays. Early studies on glioblastoma, cord blood, and fetal brain development have demonstrated the compatibility of Infinium arrays with oxidative bisulfite (OxBS) conversion methods [20, 93, 94]. This study utilized a combined chemical and enzymatic conversion strategy with the array, offering several advantages for ternary-code methylome profiling. First, the array offers high effective depth. Unlike 5modC, 5hmC does not exhibit the bimodal distribution characteristic in homogeneous cell populations. Given that 5hmC is more of a continuous variable, the Infinium arrays provide better quantitative resolution, effectively achieving an equivalent sequencing depth of approximately $100\times$, which then permits the resolution of modification components at base resolution [54]. Second, the sites selected on the Infinium arrays offer an advantage. The array is enriched in enhancer and gene body CpGs, which are particularly suitable for interrogating 5hmC dynamics as these regions are among the most biologically variable [37, 95]. Lastly, the array offers a distinctive advantage in terms of throughput. Prior studies highlight that the ternary code can vary in response to numerous biological factors such as cell type and age [11–13, 96]. Dissecting the contributions of multiple factors and their interactions requires large sample sizes, especially given the subtle changes observed (Fig. 8d). Our approach using the Infinium BeadChip offers a scalable solution for population-scale ternary-code methylome profiling, enabling a comprehensive dataset to investigate how factors, such as tissue and age, influence the ternary methylome and its regulatory implications.

Tight biochemical coupling of 5hmC and 5mC in complex grammars

Resolving 5modC signals into their component signals helps to reveal more about the biogenesis of 5hmC. As an oxidation product of 5mC, 5hmC can revert to unmodified cytosine or further oxidize, depending on the cellular context. Our analysis reveals that 5hmC consistently arises at sites where 5mC is also present. Genomic regions devoid of 5mC are also devoid of 5hmC, with promoters and bivalent regions exemplifying this substrate dependence. While further oxidation of 5hmC can lead back to unmodified C, the substrate effect of 5hmC on the production of 5mC is not evident, reflecting the higher chemical stability of 5mC than 5hmC. Notably, although our approach does not focus on 5fC/5caC, the levels of these modifications are known to be orders of magnitude below those of 5hmC and thus negligible on the scale of our array-based analysis [5–7].

Furthermore, while the processivity of DNA methyltransferases (DNMTs) is well-studied in generating contiguous 5mC patterns, the mechanisms governing 5hmC distribution remain less clear. We observed spatial correlations of 5hmCs that occur on smaller genomic scales relative to 5mCs (Fig. 2b). Although smaller in scale, the observation of

a degree of clustering of 5hmCs raises questions about whether this reflects intrinsic enzyme processivity or localized TET activity due to its recruitment to specific loci. The observed patterns emphasize the necessity of studying 5mC and 5hmC together to fully understand their interplay and the importance of analyzing the ternary-code methylome to uncover nuanced regulatory mechanisms governing how these modifications are introduced.

Ternary-code methylome augments epigenetic cell identity definition

Our tissue analysis revealed that 5hmC and 5mC collectively contribute to defining cell identity across diverse tissue types. Hierarchical clustering and t-SNE analyses identified the separation of 5mC and 5hmC profiles by tissue type, even in tissues with globally low 5hmC levels, such as blood and thymus. In neuronal tissues, subtype-specific patterns were especially prominent, showcasing the ability of 5hmC to delineate cellular identity independently of 5mC levels. *Globally*, 5hmC levels correlate with tissue turnover rates, reflecting its status as a non-inheritable marker that needs to be reestablished by TET enzymes in newly replicated DNA after installation of 5mC. In proliferative tissues, 5hmC can be diluted due to a lag in establishing 5hmC relative to the maintenance of 5mC marks. Conversely, higher 5hmC levels are observed in tissues with higher levels of post-mitotic cells or in quiescent tissues such as the aged liver [13]. *Locally*, 5mC and 5hmC establish a synergistic regulatory framework. Tissue-specific one-vs-rest (OvR) methylation enrichment analysis revealed that 5hmC is enriched in gene bodies, while low 5mC is found at TFBSs, with 5hmC often flanking these regions. These complementary distributions illustrate how 5mC and 5hmC collaboratively regulate gene networks and augment cell identity definition.

Transcriptional implications of 5hmC dynamics

Overall, 5hmC is positively associated with nearby gene expression, in contrast to 5mC, which is often negatively associated with transcriptional activity at enhancers and promoters (Fig. 7a, b). However, the relationship between cytosine modifications and gene expression is more nuanced as 5hmC biogenesis is governed by multiple factors, including cell division rates, chromatin states, and age. For instance, in long-lived cells and tissues, such as the brain, 5hmC is often broadly present, appearing to be less correlated with gene expression than more proliferative tissues (Fig. 7b). This reveals that transcription-related 5hmC presence was being weighed against other factors influencing 5hmC biogenesis. Similarly, at genomic territories typically depleted of 5hmCs, such as heterochromatin and some quiescent regions (Additional file 1: Fig. S3a, likely due to the high methylation levels that recruit MBDs and restrict TET1 binding [65]), the little 5hmC can exhibit robust correlations with the expression of the few genes these regions harbor (Fig. 7b), again reflecting this signal-over-noise balance.

Ternary-code methylome-based delineation of epigenetic aging

Compared to 5modC, 5hmC shows greater sensitivity to aging (Fig. 4d), with gains contributing to overall modification increases more at gene bodies than at Polycomb targets (Fig. 8). Similar to 5modC, age-related changes in 5hmC are also tissue-specific and associated with transcriptional consequences (Fig. 8). These findings emphasize the

importance of resolving 5hmC dynamics in the study of epigenetic aging. For example, 5hmC levels accumulate in the aged liver while they decrease in the aged lung. These variations are likely driven by shifts in cell type proportions, tissue differentiation, and developmental processes, which collectively reflect tissue functional maturation or decline [97]. Also, 5hmC alterations are implicated in age-dependent chronic inflammation in our analysis (Fig. 8e). A causal link may exist. For instance, prior research found that intermittent hypoxia exposure increased TET activity and total 5hmC levels at Wnt pathway genes in the hippocampus of mice, and Tet1 knockdown reduced hypoxia-induced neuroinflammation [98]. These findings highlight the unique potential of parsing 5hmC and 5mC in delineating biological aging, augmenting epigenetic clocks, and advancing our understanding of the physiological and cellular impacts of aging.

Limitations of the study

This study's small sample size may reduce statistical power for certain analyses. Additionally, while robust, the Infinium BeadChip's coverage is limited to approximately 1% of genome-wide CpG sites [23], potentially missing important regions. Furthermore, this study profiled bulk tissues, which may obscure cell-type-specific variations in 5hmC and 5mC levels. Recent single-cell and spatial studies [38–40, 99, 100] have begun to uncover these finer-scale differences, but they are limited to small scales, often focusing on brain tissues and neuronal cell types in a few individuals. Expanding single-cell approaches to larger populations and diverse tissue types will be critical to complement bulk-tissue analyses and fully understand the complexity of the ternary-code DNA methylome.

Conclusions

In summary, our study advances the understanding of the ternary-code DNA methylome by providing a high-resolution, base-level atlas of 5mC and 5hmC across diverse mouse tissues, ages, and sexes. The findings highlight the distinct and complementary roles of 5mC and 5hmC in epigenetic cell identity, aging, and gene expression. These insights lay a foundation for future research into epigenetic mechanisms underlying development, aging, and disease, with broad implications for cancer and regenerative biology.

Methods

Sample preparation and DNA extraction

A total of 265 samples representing 29 tissue types were obtained from 32 C57BL/6 J mice, which included 22 mice from the Jackson lab and 6 mice from the Jin lab. Genomic DNA was extracted from these mouse tissues using a phenol/chloroform/isoamyl alcohol protocol with minor modifications [101]. Tissues were incubated with lysis buffer (10 mM Tris pH 8.0, 300 mM NaCl, 5 mM EDTA pH 8.0, 0.5% SDS, and autoclaved ultrapure water). The specific lysis conditions for each tissue are described in Additional file 2: Table S1. In a fume hood, an equal volume of phenol/chloroform/isoamyl alcohol (Sigma-Aldrich, 77,617) was added to the Phase Lock Gel tube containing the lysed mixture. The aqueous phase was then transferred to a new 1.5-ml centrifuge tube (Eppendorf, 05414203). Five hundred microliters of 100% isopropanol (Millipore-Sigma, EM1.09634.1011), 1.7 μl of GlycoBlue (Invitrogen™, AM9515), and 33 μl of 7.5 M ammonium acetate solution (Sigma-Aldrich, A2706) were added. The mixture was

Lee *et al. Genome Biology*   (2025) 26:343

Page 22 of 30

vortexed briefly and incubated at $-20\,°C$ for 30 min to overnight. The samples were subsequently centrifuged at 16,000 g for 30 min at 4 °C. Following centrifugation, the samples were washed twice with 1 ml of 70% ethanol (MilliporeSigma, EM1.00983.1011). After the final wash and removal of 70% ethanol, the samples were air-dried for 10 min and then resuspended in 46 to 200 μl of Tris buffer pH 8.0 (VWR, 97,062–674) at 55 °C for 10 min. If the DNA was not fully dissolved, it was further incubated at 4 °C for 1 to 2 days. DNA concentration was measured using a Qubit 4.0 Fluorometer (Invitrogen) with the dsDNA HS Assay Kit (Invitrogen, Q33231). For DNA samples that were either not transparent or showed inadequate quantities, an additional bead purification step was performed using AMPURE XP (Beckman Coulter, A63881). Two times the volume of bead reagents was added to the sample, mixed thoroughly, briefly centrifuged, and incubated at room temperature for 5 min. The mixture was then placed on a magnetic stand and washed twice with 500 μl of freshly prepared 80% ethanol, followed by 3 to 5 min drying. Final elution was carried out with 46 μl to 200 μl of TE buffer. One microliter of the eluate was used to measure DNA concentration using the Qubit 4.0 Fluorometer.

Bisulfite conversion of genomic DNA

DNA bisulfite conversion was performed using the EZ DNA Methylation kit (Zymo Research, D5001) or EZ-96 DNA Methylation MagPrep (Zymo, D5040). Samples were bisulfite converted using the EZ DNA Methylation kit, and the conversion was performed according to the manufacturer's instructions, with the specified modifications for the Illumina Infinium Methylation Assay. The final elution volume was 25 μl of the customized buffer (0.07 × of the supplied elution buffer). Ten microliters of eluate was used for the BS array, and the remaining 15 μl was used for A3A conversion. Samples bisulfite converted by the EZ-96 DNA Methylation MagPrep kit (Zymo, D5040) followed the same process as the above, but with a final elution of 20 μl using the elution buffer. Ten microliters of the eluted solution was used for the BS array. Ten microliters of elution buffer was added to each well to the remaining 10 μl of eluate, and 15 μl of this was then used for A3A conversion.

CpG-methylated lambda DNA preparation: 1 μl of CpG methylase (Zymo, E2010) was combined with 1 μg of unmethylated lambda DNA (Promega, D1521), along with 2 μl of 10X CpG buffer, 1 μl of 20X SAM, and autoclaved ultrapure water added to achieve a total volume of 20 μl. The mixture was incubated at 30 °C overnight, but 1 μl of the CpG methylase was added after 2 and 4 h of incubation. The clean-up was performed using Genomic DNA Clean & Concentrator-10 (Zymo, D4010), followed by the manufacturer's protocol with elution by autoclaved ultrapure water.

Evaluation CpG methylation of the CpG-methylated lambda DNA: A mixture was prepared by combining 20 ng of methylated lambda DNA with 0.2 μl of HpaII (NEB, R0171S), 10X buffer, and autoclaved ultrapure water, resulting in a total volume of 10 μl. The mixture was incubated at 37 °C for 50 min, followed by inactivation at 80 °C for 20 min. Subsequently, 10 μl of the samples and 2 μl of 6X loading dye were loaded onto a gel electrophoresis using a 2% gel (VWR, 0710-500G) at 120 mA for 45 min. We used samples without evidence of fragmented DNA for further use.

The CpG-methylated lambda, amounting to 100 ng, underwent bisulfite conversion using the procedure outlined in this Methods section, and the elution was carried out

with 100 μl using the customized elution buffer. Subsequently, a mixture was prepared by combining 90 μl of the customized elution buffer with 10 μl of the eluted bisulfite-converted DNA (BCD), and this mixture was designated as ML1.

A3A deamination of BCD for bACE array

A3A enzyme purification was conducted according to the method described in [102]. The bACE conversion was carried out with minor modifications based on the method described in [103]. Fifteen microliters of BCD, 1 μl of ML1, 5 μl of DMSO (Sigma-Aldrich, D2650-100ml), 5 μl of A3A reaction buffer (35 mM 2:7:7 succinic acid: sodium dihydrogen phosphate: glycine), and autoclaved ultrapure water were combined, resulting in a final volume of 48.5 μl for the 60 μM A3A or 49 μl for the 83.7 μM A3A. The mixture was then denatured at 95 °C for 5 min and rapidly cooled by transferring to a cool rack pre-chilled on dry ice. The deamination reaction was incubated at 30 °C for 2 h with either 1 μl of A3A at 83.7 μM or 1.5 μl of A3A at 60 μM. After incubation, purification was carried out using OLIGO CLEAN AND CONC-5 (Zymo, D4060) according to the manufacturer's instructions, with the final elution performed using 12 μl of the customized elution buffer.

Quantitative real-time PCR (qPCR) for checking of A3A deamination

qPCR standards were prepared by serially diluting ML1 to concentrations of $1 \times$, $0.1 \times$, $0.01 \times$, and $0.001 \times$ using the customized elution buffer. For the qPCR reactions, 1 μl of each A3A-converted sample or qPCR standard was combined with 5 μl of Platinum II PCR Master Mix (2X) (Thermo Fisher, 14,000,012), 0.5 μl of 10% ROX (Invitrogen, 12–223-012), 0.5 μl of EvaGreen (EMSCO, NC0521178), 0.2 μl of 10 μM M1_F primer (AGGAGGTAATTAGTCGGATTGGC, IDT, see Additional file 2: Table S2), 0.2 μl of 10 μM M1_R primer (GAACCTATCTACCCGTTCGTACCGT, IDT), and autoclaved ultrapure water to reach a total volume of 10 μl. The mixtures were prepared on ice in a 384-well PCR plate (Applied Biosystems, A36931). After vortexing and brief centrifugation, the plate was placed into a QuantStudio 5 (Thermo Fisher). The amplification protocol consisted of an initial step at 94 °C for 2 min, followed by 40 cycles of 15 s at 94 °C and 45 s at 60 °C. Only A3A-converted samples with Ct values higher than the $0.001 \times$ ML1 standard were used.

BS array and bACE array by Illumina MM285 array and data pre processing

We used the maximum array input of 10 μl, as specified in [52]. BS array and bACE array were processed using the Infinium Mouse Methylation BeadChip kit (referred to as the MM285 array) according to the manufacturer's protocol. The MM285 array interrogates ~ 285,000 CpG sites, covering ~ 1.3% of the set of CpGs in the mouse genome. Despite its limited coverage, the array captures diverse methylation features genome-wide, prioritizing 5hmC-enriched genomic regions, such as enhancers [23]. The array IDAT files were subjected to preprocessing, quality control, and analysis using the *SeSAMe* R package [104]. Beta values were extracted from raw IDAT following the *openSesame* function with default parameters. GRCm38/mm10 manifest files (probe mapping information, gene annotation, CGI, chromatin state, enhancer, TFBSs) were obtained from https://zwdzwd.github.io/InfiniumAnnotation#mouse.

ACE-seq of cerebellum, heart, kidney, and blood tissue DNA

Ten nanograms of purified mouse tissue gDNA was mixed with 1% of unmethylated λ gDNA spike-in and 1% 5mCpG methylated plasmid spike-in, in a total volume of 50 μl in low TE buffer. This input DNA was then sheared to 350 bp (Covaris ultrasonicator), end-repaired, and ligated to modified Y-shaped adaptors (forward: ACACTCTTTCCCTAC ACGACGCTCTTCCGATC*T, *indicates phosphorothioate, reverse: GATCGGAAG AGCACACGTCTGAACTCCAGTCA, all Cs are 5pyC, IDT) using an NEBNext UltraII DNA library prep kit (NEB, E7645) according to the manufacturer's instructions. The sample was purified using a $1.2 \times$ left-sided SPRI (Beckman Coulter, B23317) according to the manufacturer's instructions and eluted in 17.6 μl of nuclease-free H_2O (nfH_2O). 16.6 μl of the resulting eluent was then treated with 1 μl bGT (NEB, M0357) in a reaction containing 0.4 μl UDP-glucose (NEB, S2200) and 2 μl of CutSmart Buffer (NEB), which was incubated at 37 °C for 1.5 h. Following conversion, the samples were purified by $1.2 \times$ left-sided SPRI and eluted in 17 μl of nfH_2O. Sixteen microliters of eluent was then snap-cooled by adding 4 μl of formamide (Thermo Fisher, 17,899), heating to 85 °C for 10 min, and then immediately placing the sample on ice. The snap-cooled mixture was then combined with 68 μl of nfH_2O, 10 μl of APOBEC $10 \times$ buffer (NEB, E7134), 1 μl of BSA (NEB, E7135), and 1 μl of APOBEC (NEB, E7133) and incubated at 37 °C for 3 h. The deaminated mixture was then purified by 1.2 left-sided SPRI and eluted in 23.5 μl of nfH_2O. 22.5 μl of eluent was then combined with 25 μl of 2X KAPA HiFi U^+ Polymerase (Roche, 50–196–5287) and 2.5 μl of NEBNext Multiplex Oligos for Enzymatic Methyl-seq (NEB, E1720) and amplified using 8 PCR cycles. The amplified libraries were then purified using a $0.8 \times$ left-sided SPRI purification, with elution in 11 μl nfH_2O to yield final libraries. Final libraries were quantified with a Qubit HS kit (Invitrogen, Q32851), assessed for quality using a High Sensitivity D1000 ScreenTape (Agilent, NC1786959), and sequenced on an Illumina MiSeq.

Methylome of Illumina MM285 array interpolation and evaluation

As highlighted in two previous studies [23, 95], systematic deviations observed in Infinium arrays may arise due to signal background effects, which could introduce slight measurement biases. In this study, CpG levels were interpolated using DNA methylation standards from EpigenDX Mouse DNA (with methylation levels of 0%, 5%, 10%, 25%, 50%, 75%, and 100% CpG-methylated genomic DNA; GSM5587118–GSM5587126, GSM5587128 in GEO under accession GSE184410) through the approx function with the linear option from the stats R package (version 4.3.3). The interpolated values were validated by comparing them with results from the same DNA samples (cerebellum, kidney, heart, and blood) in the ACE-seq dataset (GSE308297, Additional file 1: Fig. S1b). These interpolated beta values were then used for the downstream analysis. The MM285 probe annotation for chromHMM state is downloaded from the KnowYourCG database https://github.com/zhou-lab/KYCG_knowledgebase_MM285/tree/main/mm10.

Tissue prediction

To develop a tissue prediction model using a random forest classifier, we assessed its performance with 8 tissue types—liver, brain cortex, blood, skin, eye, femur, heart, and

lung—each having more than 5 samples for both sexes, resulting in a total of 105 samples. Seventy percent ($n = 73$) were used as the training set, and the remaining 30% were used as the test set. We selected 2000 5hmCpGs from the training set, which was identified as the most variable in 5hmC levels across the samples. We applied tenfold cross-validation with a 70:30 split of training to test data using the caret R package (version 6.0.94). The model was configured with 150 trees (ntree = 150). The model's accuracy was evaluated by calculating the proportion of correctly classified samples in the test set. SHAP values of the selected 2000 5hmCpGs were assessed using the iml R package (version 0.11.4).

Tissue-specific hyper- and hypomethylation signatures

We used a one-vs-rest approach to identify CpGs uniquely hypo- or hypermethylated in each tissue. We first computed the area under the curve (AUC) to discriminate the target tissue from tissues of other tissue types. CpGs with > 0.3 missingness in both the target/rest groups were excluded. Only CpGs with a delta beta > 0.05 were selected as cell markers. For visualization, the top 50 hypo- and hypermethylated CpGs sorted by AUC and delta beta were selected for each tissue type. For 5hmC samples, the same analysis was performed, but only CpGs with zero missingness in the target group were considered, and a delta beta > 0.02 was used as a threshold to be considered a marker.

Tissue-specific enrichment analysis of CpG sets with TFBSs and genes

CpG sets linked to hypo- and hypermethylated tissue signatures were analyzed for enrichment with TFBSs [105] using the testEnrichment function and for associated genes using KYCG_buildGeneDBs, both from the *SeSAMe* R package and utilizing the MM285 manifest (available at https://github.com/zhou-lab/KYCG_knowledgebase_MM285/tree/main/mm10).

DNA methylation-gene expression correlations

We assessed 11 tissue types (adrenal, brain, heart, kidney, liver, lung, spleen, stomach, testis, thymus, and uterus) of the fragments per kilobase million (FPKM) dataset [66] with our methylome dataset. Genes were included if their FPKM values exceeded 0.1 in at least one sample, and 5hmCpGs were selected if their beta values were greater than 0.1 in at least one sample. The brain data from the FPKM dataset were compared with various neuronal tissue types (cerebellum, subcortical brain, brain cortex, spinal cord, sciatic nerve, and optic nerve) from the methylome dataset, focusing on selected CpG probes.

The RNA raw count data used for analyzing the relationship between DNA methylation and gene expression with aging were sourced from https://twcstanford.shinyapps.io/maca/ [90]. This raw count data was processed and normalized using the DESeq2 R package (version 1.42.1) [106]. We conducted a linear regression analysis to assess changes in gene expression with aging, accounting for each tissue type and calculated the slope coefficients. Genes with a linear regression p value less than 0.05 were selected for further analysis. For comparison with the ternary methylome dataset, we selected CpG probes with a beta value greater than 0.1 in at least one sample. We then performed a linear regression analysis of 5hmC, 5mC, and 5modC levels against tissue type and

Lee *et al. Genome Biology*   (2025) 26:343

Page 26 of 30

aging and calculated the slope coefficients. CpG probes with a linear regression p value less than 0.05 were subsequently used for further analysis.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03808-y.

Additional file 1: Figs. S1–S8. Fig. S1 Distribution of cytosine modifications in various tissue contexts. Fig. S2 Interplay between 5mC and 5hmC in local genomic contexts. Fig. S3 The correlation between global cytosine modifications. Fig. S4 The tissue specificity analysis of 5hmC and 5mC. Fig. S5 Tissue prediction model evaluation and tissue signatures of 5mCs. Fig. S6 Tissue-specific cytosine modifications in various tissue types. Fig. S7 Analysis of tissue-specific signatures of 5hmC and 5mC in relation to gene expression. Fig. S8 Analysis of aging-related cytosine modifications.

Additional file 2: Tables S1 and S2. Table S1 Tissue DNA extraction details. Table S2 Primers for methylation-specific PCR (MSP).

Peer review information
Claudia Feng was the primary editor 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
W.Z. and R.M.K. conceived and supervised the project. S.M.L., W.Z., and R.M.K. wrote the main manuscript text, and S.M.L. and W.Z. prepared figures. S.M.L, D.C.G., C.C., E.K., and I.Z. produced data. J.B.P., C.K., C.E.L., C.J., and R.P. produced and/or provided materials. M.S.B. contributed to conceptualization and interpretation. All authors contributed to the discussion of the results and reviewed the manuscript.

Data availability
The generated mouse BS and bACE methylome profiles ($N = 530$) are available in the Gene Expression Omnibus with accession GSE290585 [107], and the ACE-seq dataset is available under GSE308297 [108]. Informatics for array data pre-processing and functional analysis is available in the R/Bioconductor package *SeSAMe* (version 3.22 +): (https://bioconductor.org/packages/release/bioc/html/sesame.html) (https://bioconductor.org/packages/release/bioc/html/sesame.html). Custom scripts created for these analyses are available in a public GitHub repository (https://github.com/varamos/DNAm_Atlas) [109] and are archived in Zenodo [110]. Accession codes for the published data in GEO used in this study are as follows: DNA methylation standards, GSE184410 [111]; transcriptomes, PRJNA375882 [112]; GSE132040 [113].

Declarations

Ethics approval and consent to participate
This study was approved by the Children's Hospital of Philadelphia Institutional Animal Care and Use Committee.

Consent for publication
Not applicable.

Competing of interests
W.Z. received Infinium BeadChips from Illumina Inc. R.P. is an Illumina employee.

References
1. Laird A, Thomson JP, Harrison DJ, Meehan RR. 5-hydroxymethylcytosine profiling as an indicator of cellular state. Epigenomics. 2013;5(6):655–69.
2. Pfeifer GP, Xiong W, Hahn MA, Jin SG. The role of 5-hydroxymethylcytosine in human cancer. Cell Tissue Res. 2014;356(3):631–41.
3. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929–30.
4. Wigler M, Levy D, Perucho M. The somatic replication of DNA methylation. Cell. 1981;24(1):33–40.
5. Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 5-hydroxymethylcytosine is a predominantly stable DNA modification. Nat Chem. 2014;6(12):1049–55.

6. 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(6047):1300–3.
7. Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, Fung HL, et al. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell. 2013;153(3):692–706.
8. Doege CA, Inoue K, Yamashita T, Rhee DB, Travis S, Fujita R, et al. Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature. 2012;488(7413):652–5.
9. Costa Y, Ding J, Theunissen TW, Faiola F, Hore TA, Shliaha PV, et al. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature. 2013;495(7441):370–4.
10. Parry A, Rulands S, Reik W. Active turnover of DNA methylation during cell fate decisions. Nat Rev Genet. 2021;22(1):59–66.
11. Cui XL, Nie J, Ku J, Dougherty U, West-Szymanski DC, Collin F, et al. A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. Nat Commun. 2020;11(1):6161.
12. 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(1):4249.
13. Occean JR, Yang N, Sun Y, Dawkins MS, Munk R, Belair C, et al. Gene body DNA hydroxymethylation restricts the magnitude of transcriptional changes during aging. Nat Commun. 2024;15(1):6357.
14. Wagner M, Steinbacher J, Kraus TF, Michalakis S, Hackner B, Pfaffeneder T, et al. Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. Angew Chem Int Ed Engl. 2015;54(42):12511–4.
15. Cheng Y, Bernstein A, Chen D, Jin P. 5-hydroxymethylcytosine: a new player in brain disorders? Exp Neurol. 2015;268:3–9.
16. Sun W, Zang L, Shu Q, Li X. From development to diseases: the role of 5hmC in brain. Genomics. 2014;104(5):347–51.
17. Ficz G, Gribben JG. Loss of 5-hydroxymethylcytosine in cancer: cause or consequence? Genomics. 2014;104(5):352–7.
18. An J, Rao A, Ko M. TET family dioxygenases and DNA demethylation in stem cells and cancers. Exp Mol Med. 2017;49(4):e323.
19. Chen R, Zhang Q, Duan X, York P, Chen GD, Yin P, et al. The 5-hydroxymethylcytosine (5hmC) reader UHRF2 is required for normal levels of 5hmC in mouse adult brain and spatial learning and memory. J Biol Chem. 2017;292(11):4533–43.
20. Johnson KC, Houseman EA, King JE, von Herrmann KM, Fadul CE, Christensen BC. 5-hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients. Nat Commun. 2016;7:13177.
21. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS ONE. 2010;5(1):e8888.
22. Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature. 2015;523(7559):212–6.
23. Zhou W, Hinoue T, Barnes B, Mitchell O, Iqbal W, Lee SM, et al. DNA methylation dynamics and dysregulation delineated by high-throughput profiling in the mouse. Cell Genom. 2022;2(7):100144.
24. He Y, Hariharan M, Gorkin DU, Dickel DE, Luo C, Castanon RG, et al. Spatiotemporal DNA methylome dynamics of the developing mouse fetus. Nature. 2020;583(7818):752–9.
25. Loyfer N, Magenheim J, Peretz A, Cann G, Bredno J, Klochendler A, et al. A DNA methylation atlas of normal human cell types. Nature. 2023;613(7943):355–64.
26. 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(7991):366–77.
27. 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(6667):eadf5357.
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(6083):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(4):424–9.
30. Wang T, Fowler JM, Liu L, Loo CE, Luo M, Schutsky EK, et al. Direct enzymatic sequencing of 5-methylcytosine at single-base resolution. Nat Chem Biol. 2023;19(8):1004–12.
31. 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;36(11):1083–90.
32. 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(6):1368–80.
33. Zeng H, He B, Xia B, Bai D, Lu X, Cai J, et al. Bisulfite-free, nanoscale analysis of 5-hydroxymethylcytosine at single base resolution. J Am Chem Soc. 2018;140(41):13190–4.
34. Wen L, Tang F. Genomic distribution and possible functions of DNA hydroxymethylation in the brain. Genomics. 2014;104(5):341–6.
35. Li W, Liu M. Distribution of 5-hydroxymethylcytosine in different human tissues. J Nucleic Acids. 2011;2011:870726.
36. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, et al. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. 2011;25(7):679–84.
37. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. 2011;12(6):R54.
38. 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(6):960–74.
39. 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(49):e2310367120.

40. 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(1):85–96.
41. De Ridder K, Che H, Leroy K, Thienpont B. Benchmarking of methods for DNA methylome deconvolution. Nat Commun. 2024;15(1):4134.
42. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature. 2018;555(7697):469–74.
43. Gaiti F, Chaligne R, Gu H, Brand RM, Kothen-Hill S, Schulman RC, et al. Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. Nature. 2019;569(7757):576–80.
44. Yehya N, Till JE, Srivastava N, Zhang D, Christie JD, Carpenter EL, et al. Cell-free DNA methylomics identify tissue injury patterns in pediatric ARDS. JCI Insight. 2025.
45. Walker NJ, Rashid M, Yu S, Bignell H, Lumby CK, Livi CM, et al. Hydroxymethylation profile of cell-free DNA is a biomarker for early colorectal cancer. Sci Rep. 2022;12(1):16566.
46. Wang Z, Du M, Yuan Q, Guo Y, Hutchinson JN, Su L, et al. Epigenomic analysis of 5-hydroxymethylcytosine (5hmC) reveals novel DNA methylation markers for lung cancers. Neoplasia. 2020;22(3):154–61.
47. Zhang S, Zhang J, Hu X, Yin S, Yuan Y, Xia L, et al. Noninvasive detection of brain gliomas using plasma cell-free DNA 5-hydroxymethylcytosine sequencing. Int J Cancer. 2023;152(8):1707–18.
48. Li W, Zhang X, Lu X, You L, Song Y, Luo Z, et al. 5-hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. Cell Res. 2017;27(10):1243–57.
49. Yu F, Li K, Li S, Liu J, Zhang Y, Zhou M, et al. CFEA: a cell-free epigenome atlas in human diseases. Nucleic Acids Res. 2020;48(D1):D40–4.
50. Goldberg DC, Cloud C, Lee SM, Barnes B, Gruber S, Kim E, et al. Scalable screening of ternary-code DNA methylation dynamics associated with human traits. Cell Genom. 2025;5(9):100929.
51. Putiri EL, Tiedemann RL, Thompson JJ, Liu C, Ho T, Choi JH, et al. Distinct and overlapping control of 5-methylcytosine and 5-hydroxymethylcytosine by the TET proteins in human cancer cells. Genome Biol. 2014;15(6):R81.
52. Lee SM, Loo CE, Prasasya RD, Bartolomei MS, Kohli RM, Zhou W. Low-input and single-cell methods for Infinium DNA methylation BeadChips. Nucleic Acids Res. 2024;52(7):e38.
53. Ding W, Kaur D, Horvath S, Zhou W. Comparative epigenome analysis using Infinium DNA methylation beadchips. Brief Bioinform. 2023;24(1):bbac617.
54. Zhou L, Ng HK, Drautz-Moses DI, Schuster SC, Beck S, Kim C, et al. Systematic evaluation of library preparation methods and sequencing platforms for high-throughput whole genome bisulfite sequencing. Sci Rep. 2019;9(1):10383.
55. Caldwell BA, Liu MY, Prasasya RD, Wang T, DeNizio JE, Leu NA, et al. Functionally distinct roles for TET-oxidized 5-methylcytosine bases in somatic reprogramming to pluripotency. Mol Cell. 2021;81(4):859–69 e8.
56. Wang T, Loo CE, Kohli RM. Enzymatic approaches for profiling cytosine methylation and hydroxymethylation. Mol Metab. 2022;57:101314.
57. Stewart SK, Morris TJ, Guilhamon P, Bulstrode H, Bachman M, Balasubramanian S, et al. OxBS-450k: a method for analysing hydroxymethylation using 450k beadchips. Methods. 2015;72:9–15.
58. Houseman EA, Johnson KC, Christensen BC. OxyBS: estimation of 5-methylcytosine and 5-hydroxymethylcytosine from tandem-treated oxidative bisulfite and bisulfite DNA. Bioinformatics. 2016;32(16):2505–7.
59. Seim I, Ma S, Gladyshev VN. Gene expression signatures of human cell and tissue longevity. NPJ Aging Mech Dis. 2016;2:16014.
60. 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(5929):930–5.
61. Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE. 2010;5(12):e15367.
62. Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem. 2004;279(46):48350–9.
63. Xing X, Sato S, Wong NK, Hidaka K, Sugiyama H, Endo M. Direct observation and analysis of TET-mediated oxidation processes in a DNA origami nanochip. Nucleic Acids Res. 2020;48(8):4041–51.
64. van der Velde A, Fan K, Tsuji J, Moore JE, Purcaro MJ, Pratt HE, et al. Annotation of chromatin states in 66 complete mouse epigenomes during development. Commun Biol. 2021;4(1):239.
65. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol Cell. 2011;42(4):451–64.
66. Li B, Qing T, Zhu J, Wen Z, Yu Y, Fukumura R, et al. A comprehensive mouse transcriptomic bodymap across 17 tissues by RNA-seq. Sci Rep. 2017;7(1):4200.
67. Reizel Y, Spiro A, Sabag O, Skversky Y, Hecht M, Keshet I, et al. Gender-specific postnatal demethylation and establishment of epigenetic memory. Genes Dev. 2015;29(9):923–33.
68. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 2017;356(6337):eaaj2239.
69. Vu H, Ernst J. Universal chromatin state annotation of the mouse genome. Genome Biol. 2023;24(1):153.
70. Xue Y, Guo C, Hu F, Zhu W, Mao S. PPARA/RXRA signalling regulates the fate of hepatic non-esterified fatty acids in a sheep model of maternal undernutrition. Biochim Biophys Acta Mol Cell Biol Lipids. 2020;1865(2):158548.
71. Velica P, Davies NJ, Rocha PP, Schrewe H, Ride JP, Bunce CM. Lack of functional and expression homology between human and mouse aldo-keto reductase 1C enzymes: implications for modelling human cancers. Mol Cancer. 2009;8:121.
72. Enomoto N, Ikejima K, Yamashina S, Enomoto A, Nishiura T, Nishimura T, et al. Kupffer cell-derived prostaglandin E(2) is involved in alcohol-induced fat accumulation in rat liver. Am J Physiol Gastrointest Liver Physiol. 2000;279(1):G100–6.
73. Furukawa T, Morrow EM, Cepko CL. Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. Cell. 1997;91(4):531–41.

74. Swain PK, Chen S, Wang QL, Affatigato LM, Coats CL, Brady KD, et al. Mutations in the cone-rod homeobox gene are associated with the cone-rod dystrophy photoreceptor degeneration. Neuron. 1997;19(6):1329–36.

75. Hayashi T, Mizobuchi K, Kameya S, Ueno S, Matsuura T, Nakano T. A mild form of POC1B-associated retinal dystrophy with relatively preserved cone system function. Doc Ophthalmol. 2023;147(1):59–70.

76. Mollenhauer J, Herbertz S, Helmke B, Kollender G, Krebs I, Madsen J, et al. Deleted in malignant brain tumors 1 is a versatile mucin-like molecule likely to play a differential role in digestive tract cancer. Cancer Res. 2001;61(24):8880–6.

77. Takeshita H, Sato M, Shiwaku HO, Semba S, Sakurada A, Hoshi M, et al. Expression of the *DMBT1* gene is frequently suppressed in human lung cancer. Jpn J Cancer Res. 1999;90(9):903–8.

78. De Lisle RC, Xu W, Roe BA, Ziemer D. Effects of Muclin (Dmbt1) deficiency on the gastrointestinal system. Am J Physiol-Gastrointest Liver Physiol. 2008;294(3):G717-27.

79. Horcher M, Souabni A, Busslinger M. Pax5/BSAP maintains the identity of B cells in late B lymphopoiesis. Immunity. 2001;14(6):779–90.

80. Eberhard D, Jimenez G, Heavey B, Busslinger M. Transcriptional repression by Pax5 (BSAP) through interaction with corepressors of the Groucho family. EMBO J. 2000;19(10):2292–303.

81. Inagaki Y, Hayakawa F, Hirano D, Kojima Y, Morishita T, Yasuda T, et al. PAX5 tyrosine phosphorylation by SYK co-operatively functions with its serine phosphorylation to cancel the PAX5-dependent repression of BLIMP1: a mechanism for antigen-triggered plasma cell differentiation. Biochem Biophys Res Commun. 2016;475(2):176–81.

82. 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(7):1417–30.

83. Li Y, Zheng H, Wang Q, Zhou C, Wei L, Liu X, et al. Genome-wide analyses reveal a role of polycomb in promoting hypomethylation of DNA methylation valleys. Genome Biol. 2018;19(1):18.

84. O'Geen H, Mihalovits A, Brophy BD, Yang H, Miller MW, Lee CJ, et al. De-novo DNA methylation of bivalent promoters induces gene activation through PRC2 displacement. bioRxiv. 2025.

85. Huidobro C, Fernandez AF, Fraga MF. Aging epigenetics: causes and consequences. Mol Aspects Med. 2013;34(4):765–81.

86. Zipple MN, Zhao I, Kuo DC, Lee SM, Sheehan MJ, Zhou W. Ecological realism accelerates epigenetic aging in mice. Aging Cell. 2025;24(6):e70098.

87. Zhou W, Reizel Y. On correlative and causal links of replicative epimutations. Trends Genet. 2025;41(1):60–75.

88. Zoller JA, Parasyraki E, Lu AT, Haghani A, Niehrs C, Horvath S. DNA methylation clocks for clawed frogs reveal evolutionary conservation of epigenetic aging. Geroscience. 2024;46(1):945–60.

89. Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. Nat Genet. 2018;50(4):591–602.

90. Schaum N, Lehallier B, Hahn O, Palovics R, Hosseinzadeh S, Lee SE, et al. Ageing hallmarks exhibit organ-specific temporal signatures. Nature. 2020;583(7817):596–602.

91. Miliani de Marval PL, Kim SH, Rodriguez-Puebla ML. Isolation and characterization of a stem cell side-population from mouse hair follicles. Methods Mol Biol. 2014;1195:259–68.

92. Petrides G, Clark JR, Low H, Lovell N, Eviston TJ. Three-dimensional scanners for soft-tissue facial assessment in clinical practice. J Plast Reconstr Aesthet Surg. 2021;74(3):605–14.

93. Solomon O, Macisaac JL, Tindula G, Kobor MS, Eskenazi B, Holland N. 5-hydroxymethylcytosine in cord blood and associations of DNA methylation with sex in newborns. Mutagenesis. 2019;34(4):315–22.

94. Spiers H, Hannon E, Schalkwyk LC, Bray NJ, Mill J. 5-hydroxymethylcytosine is highly dynamic across human fetal brain development. BMC Genomics. 2017;18(1):738.

95. Kaur D, Lee SM, Goldberg D, Spix NJ, Hinoue T, Li HT, et al. Comprehensive evaluation of the Infinium Human MethylationEPIC v2 BeadChip. Epigenetics Commun. 2023;3(1):6.

96. Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, et al. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res. 2012;22(3):467–77.

97. Sun ED, Zhou OY, Hauptschein M, Rappoport N, Xu L, Navarro Negredo P, et al. Spatial transcriptomic clocks reveal cell proximity effects in brain ageing. Nature. 2025;638(8049):160–71.

98. Kong Y, Ji J, Zhan X, Yan W, Liu F, Ye P, et al. Tet1-mediated 5hmC regulates hippocampal neuroinflammation via wnt signaling as a novel mechanism in obstructive sleep apnoea leads to cognitive deficit. J Neuroinflammation. 2024;21(1):208.

99. Lee CN, Fu H, Cardilla A, Zhou W, Deng Y. Spatial joint profiling of DNA methylome and transcriptome in tissues. Nature. 2025.

100. Iqbal W, Zhou W. Computational methods for single-cell DNA methylome analysis. Genomics Proteomics Bioinformatics. 2023;21(1):48–66.

101. Molbert N, Ghanavi HR, Johansson T, Mostadius M, Hansson MC. An evaluation of DNA extraction methods on historical and roadkill mammalian specimen. Sci Rep. 2023;13(1):13080.

102. Wang T, Luo M, Berrios KN, Schutsky EK, Wu H, Kohli RM. Bisulfite-free sequencing of 5-hydroxymethylcytosine with APOBEC-coupled epigenetic sequencing (ACE-Seq). Methods Mol Biol. 2021;2198:349–67.

103. Foong YH, Caldwell B, Thorvaldsen JL, Krapp C, Mesaros CA, Zhou W, et al. TET1 displays catalytic and non-catalytic functions in the adult mouse cortex. Epigenetics. 2024;19(1):2374979.

104. Zhou W, Triche TJ Jr., Laird PW, Shen H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. Nucleic Acids Res. 2018;46(20):e123.

105. Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, et al. Cistrome data browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res. 2019;47(D1):D729–35.

106. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

107. Lee SM, Zhou W. A ternary-code DNA methylome atlas of mouse tissues. Gene Expression Omnibus. 2025. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE290585.

108. Lee SM, Goldberg DC, Cloud C, Parker JB, Krapp C, Loo CE, et al. A ternary-code DNA methylome atlas of mouse tissues. Gene Expression Omnibus. 2025. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE308297.
109. Lee SM. DNAm_Atlas. GitHub. 2025. https://github.com/varamos/DNAm_Atlas.
110. Lee SM. A ternary-code DNA methylome atlas of mouse tissues. 2025. Zenodo. https://doi.org/10.5281/zenodo.17109847.
111. Zhou W, Shen H, Laird PW. Mouse DNA methylation atlas using Infinium Mouse Methylation Beadchips. Gene Expression Omnibus. 2022. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE184410.
112. Li B, Qing T, Zhu J, Wen Z, Yu Y, Fukumura R, et al. A comprehensive mouse transcriptomic bodymap across 17 tissues by RNA-seq. Sci Rep. 2017. https://doi.org/10.1038/s41598-017-04520-z.
113. Schaum N, Hosseinzadeh S, Hahn O, Pisco AO, Darmanis S, Wyss-Coray T, et al. Tabula Muris Senis: bulk sequencing. Gene Expression Omnibus. 2019. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132040.

Publisher's Note