Wang et al. *Zool. Res.* 2024, 45(6): 1371−1384
https://doi.org/10.24272/j.issn.2095-8137.2024.221
https://cstr.cn/32248.14.zr.2024.221

ZR | Zoological Research

Article Open Access

Whole-genome methylation reveals tissue-specific differences in non-CG methylation in bovine

Jing Wang[1], Wen Yuan[1], Fang Liu[1], Guang-Bo Liu[1], Xiao-Xiong Geng[1], Chen Li[1], Chen-Chen Zhang[1], Nan Li[1], Xue-Ling Li[1,*]

[1] *State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestocks*, *Inner Mongolia University*, *Hohhot*, *Inner Mongolia* 010070, *China*

ABSTRACT

DNA methylation at non-CG dinucleotides (mCH, H=A, C, T) widely occurs and plays an important role in specific cell types, including pluripotent, neural, and germ cells. However, the functions and regulatory mechanisms of mCH, particularly in species other than humans and mice, remain inadequately explored. In this study, we analyzed the distribution of mCH across different bovine tissues, identifying significantly elevated mCH levels in bovine embryonic stem cells (bESCs), as well as brain, spleen, and ileum tissues compared to other tissues. Marked differences in mCH patterns between somatic cells and bESCs were observed, reflecting distinct base preferences and the differential expression of DNA methyltransferases. We also identified exon methylation in both CG and non-CG contexts, resembling gene-associated methylation patterns observed in plants. To characterize tissue-specific variations in mCH, we developed a novel method for differential mCH analysis. Results indicated that mCH is not randomly distributed but tends to be enriched in tissue-specific functional regions. Furthermore, regression models demonstrated a positional correlation between CG methylation and mCH. This study enhances our understanding of mCH distribution and function in bovine somatic and stem cells, providing new insights into its potential roles across species and tissues. These findings advance knowledge of epigenetic mechanisms, shedding light on the potential involvement of mCH in development and disease processes.

Keywords: Epigenetics; Non-CG Methylation; CH-DMRs; Comparative Genomics; Developmental Biology

INTRODUCTION

DNA methylation is a common epigenetic modification in mammalian genomes, primarily occurring at CG dinucleotides. However, recent studies have shown that non-CG methylation (mCH, H=A, C, T) is also extensively present and exhibits non-random distribution across genomes, particularly in pluripotent stem cells (PSCs), neural cells, and germ cells, suggesting it may play important roles in specific cell types (Lister et al., 2009; Lister et al., 2013; He & Ecker, 2015; Schultz et al., 2016; He et al., 2020). mCH levels are initially very low in the prefrontal cortex of human fetuses, but increase markedly with brain maturation (Lister et al., 2013). Similarly, reprogrammed somatic cells display considerable mCH over large genomic regions that have not fully reverted to the pluripotent epigenetic state (Hu et al., 2012). Collectively, these observations suggest that mCH may contribute not only to the epigenetic regulation of pluripotent and neural cells but also to other cell types.

The sequence context of mCH varies significantly among cell and tissue types. For example, mCH is primarily enriched at 5’ CAG 3’ sites in embryonic stem cells (ESCs) (Lister et al., 2009; Laurent et al., 2010) but at 5’ CAC 3’ sites in neurons (Lister et al., 2013). Based on these sequence motifs, cells and tissues with mCH can be classified into mCAG and mCAC groups. The mCAG group includes pluripotent cells, such as ESCs, nuclear transfer cells, and induced pluripotent stem cells (iPSCs) (Haines et al., 2001; Laurent et al., 2010; Ziller et al., 2011), as well as oocytes (Guo et al., 2014a; Wang et al., 2014) and polar bodies, while the mCAC group includes the prefrontal cortex, neurons, glial cells, and various human organs (Lister et al., 2013; Schultz et al., 2015). In the mCAG group, mCH levels within genes are moderately associated with increased expression, while in the mCAC group, higher mCH levels correlate with decreased expression (Guo et al., 2014a; Guo et al., 2014b), indicating at least two distinct regulatory pathways.

The *de novo* methylation of mCH is mediated by DNA

Received: 21 August 2024; Accepted: 29 September 2024; Online: 30 September 2024

Foundation items: This work was supported by the STI 2030-Major Projects (2023ZD0407504) of China, Development Plan for Young Scientific and Technological Talents in Colleges and Universities of Inner Mongolia Autonomous Region of China (NMGIRT2204), National Natural Science Foundation of China (32160172), and Science and Technology Major Project of the Inner Mongolia Autonomous Region of China to the State Key Laboratory of Reproductive Regulation (2021ZD0048&2023KYPT0010)

*Corresponding author, E-mail: lixueling@imu.edu.cn

methyltransferase enzymes DNMT3A and DNMT3B (Ramsahoye et al., 2000; Liao et al., 2015). Although these enzymes preferentially methylate CG sites, they can also methylate non-CG sites at lower frequencies. The precise mechanisms by which DNMT3A and DNMT3B selectively target specific genomic regions to establish a non-random mCH landscape remain largely unclear. Certain transcription factors, such as MeCP2, are capable of binding to mCH-containing DNA, thereby linking mCH with CG methylation (mCG) in transcriptional repression (Lagger et al., 2017). MBD2b also enhances DNA binding through the coexistence of mCG and mCH (Guo et al., 2014b).

Despite recent advances in methylation research, significant gaps persist in our understanding of the functions and mechanisms of non-CG methylation in different cell types. Existing mCH studies have primarily focused on model species (Lister et al., 2009; Laurent et al., 2010; Lister et al., 2013; He et al., 2020), with relatively limited research conducted in other species, and the tissue-specific roles of mCH and their corresponding biological functions have not been fully characterized. This study addresses these gaps by mapping mCH distribution across various bovine tissues, introducing a novel differential analysis to identify tissue-specific mCH patterns, examining mCH developmental dynamics in the forebrain, and investigating spatial correlations between mCH and mCG. This work provides a framework for studying functional differences in mCH across species, offering a foundation for future studies on the roles of mCH in diverse biological contexts.

MATERIALS AND METHODS

Sample collection

All animal experiments were performed in accordance with the guidelines of the Animal Welfare Committee of Inner Mongolia University under approval number IMU-2020-CATTLE-045. Fetal bovine tissues were collected from four healthy pregnant Holstein cows at a slaughterhouse. Crown-rump length (CRL) was measured to estimate gestational age. Fetal sex was determined based on gonadal position within the abdominal cavity, anatomical structure, and proximity to adjacent organs (mesonephros and/or kidneys), identifying two male and two female fetuses (13–16 cm CRL). Collected tissues were diced, rapidly cooled in liquid nitrogen, and preserved at –80°C until further study.

Cell culture

Bovine embryonic stem cells (bESCs) were cultured using a previously published protocol (Bogliotti et al., 2018). This culture system included mTeSR-E6 basal medium (STEMCELL Technologies, USA), WNT signaling pathway inhibitor IWR-1 (2.5 μmol/L, Sigma Aldrich, USA), bovine serum albumin (BSA) with extremely low free fatty acids (≤ 0.05 mg/g total protein) (0.664 g/50 mL of culture medium, MP Biomedicals, USA), and recombinant human basic fibroblast growth factor (rh-bFGF, 20 ng/mL, PeproTech, USA) to establish bESCs from *in vitro* fertilized bovine embryos. The cell line generated was termed bESCs-F7.

Furthermore, bESCs were also cultured with MM-102, an inhibitor of mixed-lineage leukemia 1 (MLL1) (Li et al., 2023), under two different conditions. In the first condition, bESCs-F7 were exposed to a high concentration (50 μmol/L) of MM-102 (5307/10, Tocris Bioscience, UK) for 7 days, followed by five additional passages before characterization, yielding the bESCs-F7-50 cell line. In the second condition, bESCs-F7 cells were treated with a low concentration (5 μmol/L) of MM-102 over a longer period, with characterization following five passages, resulting in the bESCs-F7-5 cell line.

Bovine extended pluripotent stem cells (bEPSCs) were cultured using a previously published culture system (Zhao et al., 2021). In brief, 500 mL of medium was prepared by combining 485 mL of basic mTeSR1 medium (Stemcell, 85850, USA), 5.0 mL of 100×penicillin-streptomycin solution (Gibco, USA), 0.1 mmol/L 2-mercaptoethanol (Gibco, USA), 1 μmol/L GSK3β inhibitor CHIR99021 (Selleck Chemicals, S2924, USA), 0.3 μmol/L Lck/Src inhibitor WH-4-023 (Selleck Chemicals, S7565, USA), 5 μmol/L tankyrase inhibitor XAV939 (Sigma, X3004, USA), or 5 μmol/L classic WNT signaling pathway inhibitor IWR-1 (Selleck Chemicals, S7086, USA), 50 μg/mL vitamin C (Sigma, 49752-100G, USA), 10 ng/mL leukemia inhibitory factor (LIF) (Millipore, LIF1010, USA), and 20.0 ng/mL activin A (R&D, 338-AC, USA). This system was used to establish bEPSCs from bovine embryos, with the resulting cell line named bEPSCs-AGS.

Bisulfite sequencing (BS-seq)

Genomic DNA was extracted from tissues using a Magnetic Universal Genomic DNA Kit (DP705) (TIANGEN, China). Approximately 10–50 mg of tissue was finely diced, followed by the addition of 300 μL of tissue digestion solution GHA and 20 μL of protease K. The mixture was thoroughly ground and digested for 30 min. Subsequently, 300 μL of cracking solution GHL and 300 μL of isopropanol were added and mixed. Next, 15 μL of magnetic bead suspension was introduced, followed by shaking and standing. After magnetic bead adsorption, the liquid was removed, and genomic DNA was collected. Extracted DNA quality was assessed by agarose gel electrophoresis for degradation and contamination using a spectrophotometer (IMPLEN, USA). DNA concentration was determined using a Qubit Fluorometer 2.0 DNA Analysis Kit (Life Technologies, USA).

DNA samples were fragmented to 200–400 bp using the Covaris S220 system (USA). The resulting ssDNA fragments underwent bisulfite treatment using the Accel-NGS Methyl-Seq DNA Library Kit for Illumina (Swift, USA), which converted unmethylated cytosine to uracil, while leaving methylated cytosine unchanged. Library quality was assessed using the Agilent Bioanalyzer 2100 system (USA) and paired-end 150 bp (PE150) sequencing was performed using the Illumina platform (Illumina, USA).

Short reads (raw data) in FASTQ format, containing both sequence and sequencing quality information, were obtained. For effective bioinformatics analysis, sequencing artifacts were removed, including reads containing adapter contamination, low-quality nucleotides, and unrecognizable nucleotides (N). For robust downstream analysis, particularly for mCH calling where base quality is crucial, stringent quality control measures were implemented. Fastp (0.19.7) (Chen et al., 2018) was employed to conduct a preliminary quality assessment of raw reads, adopting the following strict quality control steps: paired reads were discarded if either read contained adapter contamination; paired reads were discarded if over 10% of bases in either read were ambiguous; and paired reads were discarded low-quality bases (Phred quality <5) exceeded 50% in either read. These rigorous quality control measures aimed to ensure the reliability and accuracy

of subsequent analyses.

Following quality control, the bisulfite-treated reads were aligned to the reference genome using bwa-meth (v.0.2.2) (Pedersen et al., 2014). The reference genome was first converted (C=>T, G=>A), then indexed using bwa (v.0.7.17) (Li & Durbin, 2009). Statistical measures, such as sample average coverage, were calculated using Qualimap (v.2.2.2-dev). The sequenced reads were subjected to the same conversion and aligned to the converted reference genome. The methylation status of cytosines was inferred by comparing the best alignment sequences generated from the original forward and reverse strands with the sequences from the normal genome.

Methylation levels for all cytosine positions were extracted with MethylDackel (https://github.com/dpryan79/MethylDackel, v.0.5.2). Cytosines were classified into CpG, CHG, or CHH (H=A, C, T). In cases where the reference sequence included N, it was assigned to either CHG or CHH. Cytosines located near chromosome ends, where site inferences were infeasible, were classified as CHH. Finally, the CHG and CHH methylation sites were aggregated to generate a comprehensive mCH level file for each sample.

RNA sequencing (RNA-seq)

RNA was first extracted from the samples, and its purity was measured using a spectrophotometer (IMPLEN, USA). RNA concentration was determined using a Qubit® RNA Assay Kit and RNA integrity was assessed using an RNA Nano 6000 Assay Kit on the Bioanalyzer 2100 system (Agilent Technologies, USA).

To prepare sequencing libraries, 1 μg of RNA from each sample was processed using a NEBNext® Ultra™ RNA Library Prep Kit (Illumina, USA). mRNA was purified from total RNA using magnetic beads with poly-T oligonucleotides. The synthesis of first-strand cDNA was performed using M-MuLV reverse transcriptase and random hexamer primers. Subsequently, second-strand cDNA was synthesized using DNA polymerase I and RNase H, with the remaining overhangs converted to blunt ends using exonuclease.

The 3’ ends of DNA fragments were adenylated and NEBNext adapters with hairpin loops were ligated to the cDNA fragments. Hybridization was then performed, and cDNA fragments of 250–300 bp were selected using the AMPure XP system (Beckman Coulter, USA). The ligated cDNA fragments were treated with USER enzyme at 37°C for 15 min, followed by heating at 95°C for 5 min. After completion of the reaction, polymerase chain reaction (PCR) amplification was carried out with Phusion high-fidelity DNA polymerase, universal PCR primers, and index primers. The PCR products were purified using the AMPure XP system, and library quality was evaluated using the Agilent Bioanalyzer 2100 system (USA). Sequencing was performed on the Illumina HiSeq platform (USA), generating 150 bp paired-end (PE) reads.

Fastp (v.0.20) was used to remove adapter sequences and low-quality reads, resulting in high-quality clean reads for alignment. The clean reads were then mapped to the UCSC bosTau9 reference genome using Hisat2 (v.2.0.4), which built an index for the reference genome and generated splice junction databases based on gene model annotations, offering superior accuracy to other non-splice-aware mapping tools.

For transcript quantification, the featureCounts module in Subread (v.2.0) was used to calculate read counts mapped to each gene. Fragments per kilobase million (FPKM) values were then computed for each gene based on both sequencing depth and gene length to account for their impact on read counts.

Clustering of mCH

The UCSC BT9 refGene gene annotation file was used to obtain the positions of all 14 179 genes and their promoters, 5’ UTR, exons, introns, and 3’UTR. Subsequently, k-means clustering was performed using the “kmeans” function in the R package, with all genes divided into six groups. The mCH coverage of all genes and their flanking regions was calculated using the “computeMatrix scale-regions” module in deepTools (v.3.4.3) (Ramirez et al., 2016) with the parameters: -a 10000 -b 10000 -m 10000 -bs 100 --skipZeros. The results were visualized using the “plotHeatmap” module. Gene Ontology (GO) enrichment analysis of the six gene sets was conducted using gprofile2 (v.0.2.1) (Reimand et al., 2007) and clusterProfiler (v.4.6.2) (Wu et al., 2021). A corrected *P*-value less than 0.05 was considered significantly enriched for signaling pathway associations.

Sequence context preference for mCH

In mCH-enriched samples, the sequence context preference for mCH was analyzed. A binomial test was applied to the non-conversion rate of all CH sites to identify those with methylation above background noise, allowing determination of methylation status for each site. After excluding sites with low sequencing coverage (<3) and unmethylated CH sites, the distribution of the two bases following each C site was recorded, with sites containing N excluded from analysis. The data were subsequently converted to PCM format and visualized using motifStack (v.1.42.0) (Ou et al., 2018).

Identification of CH differentially methylated regions (CH-DMRs)

CH-DMRs were identified based on previously established methods (Lister et al., 2011; Ma et al., 2014). First, the genome was divided into non-overlapping 5 kb windows, with every set of 10 consecutive windows grouped together. Groups containing more than five missing values in either the control or experimental samples were filtered out. Given the sparse distribution of mCH, mCH levels were normalized by adjusting the weighted methylation level for the bisulfite non-conversion rate and subsequently dividing by the median methylation level of the 5 kb window. To assess differential methylation, the Mann-Whitney *U* test was applied to the normalized mCH levels between samples, including fetal forebrain, adult forebrain, bEPSCs-AGS, adult spleen, adult ileum, bESCs-F7, bESCs-F7-5, and bESCs-F7-50. A window was considered differentially methylated if the absolute $\log_2$ fold-change (FC) between groups exceeded 1.5 and the adjusted *P*-value was below a 5% false discovery rate (FDR). Subsequently, windows within 100 kb of each other and sharing the same methylation status (hyper/hypo methylation) were merged, with regions smaller than 100 kb filtered out. The final set of large CH-DMRs were used for subsequent analysis.

Regression model

To examine the relationship between mCG/mCH levels and histone modification features in the adult forebrain, we quantified mCG/mCH levels and four histone modifications, including H3K4me1, H3K4me3, H3K27ac, and H3K27me3, across all detected regions. Histone modification levels were normalized to account for overall differences:

$$X_{ij} = \frac{log_2\left(X_{ij}+1\right)}{log_2\left(\max\left(X_j+1\right)\right)} \quad (1)$$

This normalization ensured that $X_{ij} \in [0, 1]$, where X represents the level of the j-th histone mark or mCG/mCH level in the i-th genomic region. Linear models were built using the "lm()" function, and the "anova()" function was applied to calculate *F*-values for each feature. Additionally, a random forest regression model was trained with the R package randomForest (v.4.2.2), using parameters "ntree=500" and "importance=TRUE" to assess the relative importance of each feature.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment

GO and KEGG (http://www.genome.jp/kegg/) enrichment analyses were performed using gprofile2 (v.0.2.1) (Reimand et al., 2007) and clusterProfiler (v.4.6.2) (Wu et al., 2021). Terms and pathways with adjusted *P*-values less than 0.05 were considered significantly enriched.

Quantitative real-time PCR (qPCR)

To extract total RNA, 500 μL of Trizol reagent was added to the cells, and the mixture was allowed to stand for 5 min, followed by the addition of 100 μL of chloroform and an additional 15 min of standing. The mixture was centrifuged (centrifugal radius 8 cm, 13 200 r/min, 15 min, 4°C), with the resulting supernatant mixed with an equal volume of isopropanol and left to stand for 5–10 min. A 70% ethanol solution was prepared. After centrifuging (centrifugal radius 8 cm, 13 200 r/min, 15 min, 4°C) the sample again, the supernatant was discarded, and 1 mL of 70% ethanol was added. The sample was again centrifuged by same condition, and the supernatant was discarded. The RNA pellet was allowed to air dry completely and was then resuspended in 20–50 μL of nuclease-free water. RNA concentration was measured using the NanoDrop 2000C and reverse-transcribed into cDNA.

Reverse transcription was performed using the TaKaRa Prime Script RT Reagent Kit with gDNA Eraser (Perfect Real Time) (TaKaRa, Japan). The first step involved DNA removal with a reaction volume of 20 μL, consisting of RNase-free water (2.0 μL), gDNA Eraser (2.0 μL), RNA template (14 μL), and 5×gDNA Eraser Buffer (4.0 μL). This reaction was carried out at 42°C for 2 min. For reverse transcription, a 40 μL system was prepared using the reaction solution from step 1 (20 μL), PrimeScript RT Enzyme Mix 1 (2 μL), PT Primer Mix (2.0 μL), RNase-free water (8.0 μL), and 5× PrimeScript Buffer 2 (for real-time) (8.0 μL). The reaction proceeded at 37°C for 15 min, followed by 85°C for 5 s. The resulting cDNA was either stored at –20°C or used directly in subsequent qPCR experiments.

For qPCR, a reaction volume of 20 μL was prepared, containing 10 μL of 2×GoTaq qPCR Master Mix, 0.2 μL of 100×CXR, 0.8 μL each of forward and reverse primers, 2 μL of cDNA template, and 6.2 μL of RNase-free water. The PCR program included the following steps: 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. To ensure accuracy, each reaction was conducted in three biological replicates, with each biological replicate run in three technical replicates. Data were normalized to GAPDH expression levels and analyzed using the relative Ct method ($2^{-\triangle\triangle Ct}$). Statistical analysis was performed in R (v.4.2.2) (https://www.r-project.org/), with significance defined at $P<0.05$.

RESULTS

Overall distribution of non-CG methylation in the bovine genome

Whole-genome methylation sequencing was performed on 18 samples, comprising nine fetal tissues (forebrain, hindbrain, muscle, heart, lung, liver, kidney, rumen, and gonads) from 12-week-old Holstein cow fetuses (Supplementary Figure S1A) and four PSC lines (bESCs-F7, bESCs-F7-5, bESCs-F7-50, and bEPSCs-AGS). These tissue and cell samples were integrated with previously sequenced adult tissues (Zhou et al., 2018; Zhou et al., 2020). The average bisulfite conversion rate across samples was 99.6%, with a mean coverage of 32.5X, and the correlation between duplicate samples exceeded 0.85. Details on samples are provided in Supplementary Table S1. Hierarchical clustering (ward.D2) of methylation profiles revealed three primary clusters: PSCs, fetal tissues, and adult tissues (Figure 1A), as anticipated. This clustering validated the consistency among biological replicates and underscored the dynamic methylation differences across developmental stages. Methylated cytosine sites were classified into CpG, CHG, and CHH contexts. The methylation level of CpG sites was generally high, with an average of 74% (except for fetal liver at 43%), while CHG and CHH methylation levels were generally low (<1%) (Figure 1A; Supplementary Table S1). However, specific samples, including three bESC lines (bESCs-F7, bESCs-F7-5, bESCs-F7-50) and adult brain, spleen, and ileum tissues, exhibited elevated levels of mCH, with average mCH exceeding 1% (Figure 1B; Supplementary Table S1). Among the samples accumulating mCH, mCA had the highest proportion (Supplementary Figure S1B).

We identified specific enrichment of mCH near the *SOX6* gene in the forebrain, a gene primarily responsible for the molecular differentiation and regulation of dorsal and ventral telencephalic progenitors and cortical interneurons (Pereira Luppi et al., 2021; Li et al., 2022). In tissues exhibiting mCH enrichment, elevated mCH levels were predominantly localized around the promoter and downstream regions of *SOX6* in the adult forebrain (Figure 1C). In addition, mCH enrichment was observed in the JUN family, including *JUN* and *JUNB*, within the adult forebrain (Supplementary Figure S1C). These genes can form AP-1 complexes with members of the FOS family, facilitating the regulation of cellular proliferation and inflammatory responses (Matsuo et al., 2000). This enrichment pattern suggests distinct mechanisms of mCH regulation across cell types, potentially linked to variations in DNMTs. We explored the expression levels of three DNA methyltransferases, including *DNMT3A*, *DNMT3B*, and *DNMT1*, in PSCs and somatic tissues (Supplementary Figure S2A). RNA-seq revealed a notably higher expression level of *DNMT3B* in PSCs compared to somatic tissues, indicating cell-type-specific *de novo* mCH patterns. Elevated mCH levels were also detected in the promoter region of *DNMT3A* within PSCs, while mCG levels remained unchanged compared to other samples (Supplementary Figure S2B), indicating a potential gene silencing role of mCH. Additionally, *DNMT1* transcription levels were higher in somatic tissues than in PSCs; however, PSCs displayed significantly increased read counts over the initial exons of the *DNMT1* gene (Supplementary Figure S2C), indicating

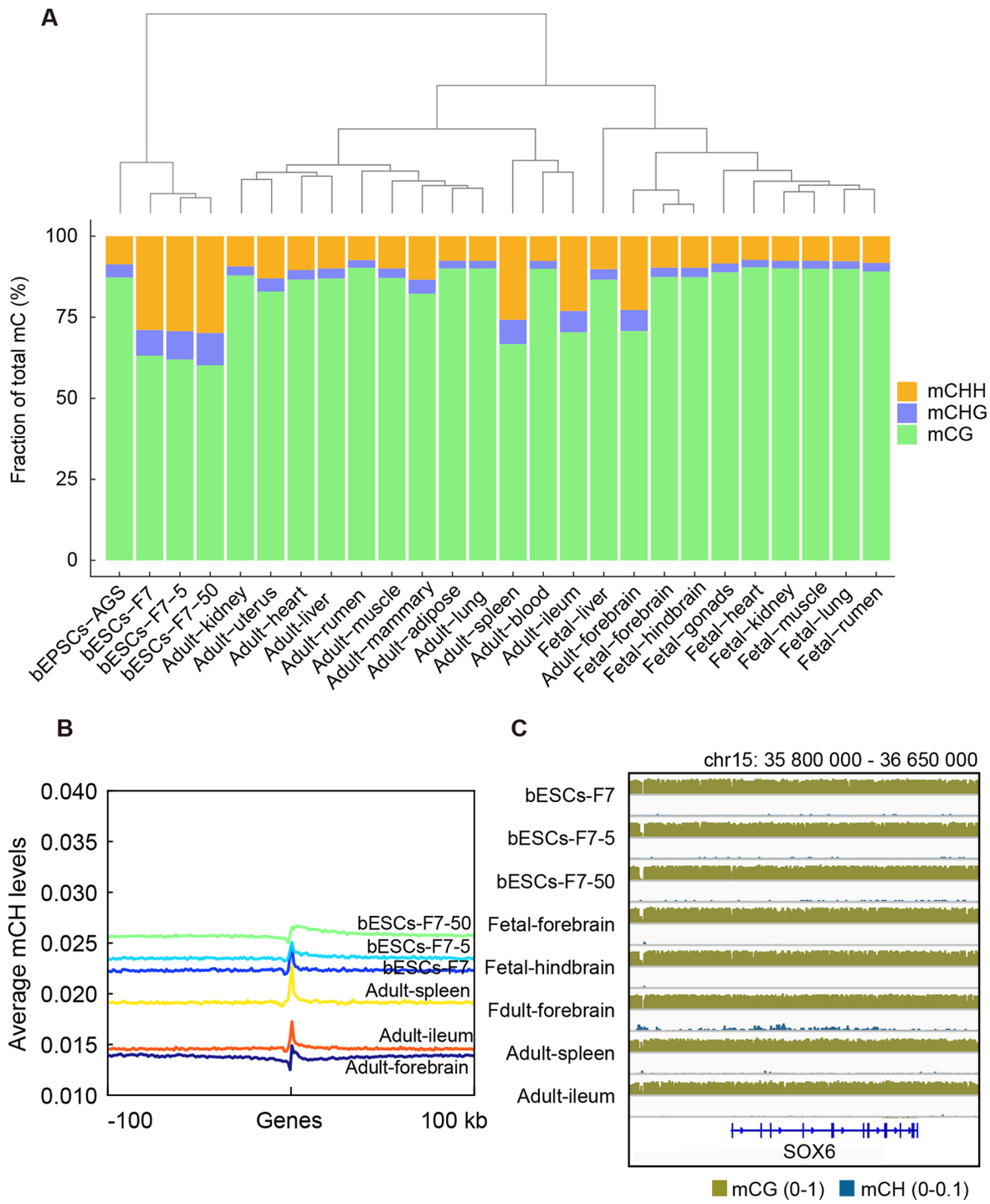

Figure 1 Global DNA methylation patterns in bovine tissues and pluripotent stem cells (PSCs) during development

A: Hierarchical clustering (ward.D2) of global methylation levels of 27 adult (Zhou et al., 2018; Zhou et al., 2020), fetal, and PSC samples after merging duplicates. From top to bottom, dendrogram represents clustering of all samples and global CpG, CHG, and CHH levels. PSCs included: primed bovine embryonic stem cells (bESCs-F7) derived from blastocysts through *in vitro* culture (Bogliotti et al., 2018); primed bESCs (bESCs-F7-5) derived from F7 and cultured long-term with low concentration MM-102 (5 μmol/L); and primed bESCs (bESCs-F7-50) derived from F7 and cultured for 7 days with high concentration MM-102 (50 μmol/L) (Li et al., 2023). Bovine extended pluripotent stem cells (bEPSCs-AGS) were derived from blastocysts through *in vitro* culture (Zhao et al., 2021). B: Average mCH levels in all genes and flanking regions (±100 kb) in bESCs-F7, bESCs-F7-5, bESCs-F7-50, bEPSCs-AGS, adult forebrain, adult spleen, and adult ileum. C: mCG and mCH in *SOX6* gene region, with height indicating mCG levels ranging from 0 to 1 and mCH levels ranging from 0 to 0.1.

differential alternative splicing. Comparative analysis of oocyte and pre-implantation embryo expression (GSE196484) (Zhu et al., 2022) showed a concordance with oocyte specific *DNMT1o* isoform expression (Russell & Betts, 2008; Mcgraw et al., 2013; Whidden et al., 2016). The qPCR results also confirmed *DNMT1o* (*DNMT1a*) expression in bESCs (Supplementary Figure S2D and Table S2). In conclusion, differential DNMT expression across cell types contributes to

distinct mCH patterns that influence critical biological functions.

Non-CG methylation enrichment patterns across different cell types

To investigate mCH patterns across various cell types, we analyzed mCH levels for all genes, including 10 kb upstream and downstream flanking regions, dividing gene-associated mCH patterns into six clusters (P1–P6) (Supplementary Table S3). Clustering revealed distinct mCH enrichment patterns, with notable differences between the forebrain and PSCs, spleen, and ileum (Figure 2A). Cluster P1 showed significantly elevated intragenic mCH levels across all cell types. In contrast, clusters P2 and P3 exhibited high mCH levels in bESCs and adult spleen and ileum, primarily concentrated on promoters and exons (Figure 2B). Genes in clusters P4–P6 displayed high mCH levels in the forebrain. For example, *SOX17*, a representative gene in cluster P1, exhibited consistently high mCH levels across all samples. In contrast, *ZIC2* and *MAFK*, both associated with neural functions, showed reduced mCH levels specifically in the forebrain. *MAFK*, associated with nerve growth, plays a role in regulating neurite outgrowth (Töröcsik et al., 2002), while *ZIC2* is involved in guiding the migration of specific neuronal groups essential for forebrain development (Figure 2C) (Murillo et al., 2015). Additionally, *JUND* demonstrated mCH accumulation in the forebrain, with a methylation pattern resembling that of other JUN family members, *JUN* and *JUNB* (Supplementary Figure S1C).

Given the inhibitory effect of mCG on gene expression, we analyzed RNA-seq data from corresponding samples to clarify the functional impact of mCH on gene expression. Results revealed significantly lower gene expression levels in clusters P2 and P3 in non-forebrain samples compared to the forebrain (Figure 2D). A significant negative correlation was observed between mCH and gene expression (cor=–0.22), indicating that mCH may exert a similar inhibitory effect as mCG. Genes affected by mCH repression included numerous transcription factors (TFs) (Supplementary Figure S3A). Clusters P2 and P3 contained many TFs related to forebrain development, while clusters P4–P6 were enriched in many metabolism-related genes (Supplementary Figure S3B and Table S4).

Certain regions with pronounced mCH enrichment emerged as particularly noteworthy (Figure 3A). Methylation was observed in both CG and non-CG contexts near exons of select genes (Figure 3B; Supplementary Table S5), resembling transposable element like methylation (teM) patterns (Bewick & Schmitz, 2017; Muyle et al., 2022; Zeng et al., 2023). Histone modification data for the corresponding samples (GSE158430) (Kern et al., 2021) showed no significant differences in histone modification levels between mCH-enriched regions and their flanking regions, although the higher levels of H3K27me3 and lower levels of H3K4me3 suggest potential gene repression within these regions (Figure 3C). Additionally, these regions exhibited higher GC content (Figure 3D) and lower expression levels (Figure 3E). For instance, mCH was enriched in the promoter region of the germ cell marker gene *DAZL* in bESCs, spleen, and ileum (Figure 3F) (Smorag et al., 2014). Although *DAZL* typically exhibits mCG loss in the promoter region of gonadal tissue, no corresponding mCH enrichment was observed. Similarly, the heat shock protein gene *HSPA1A* displayed mCH enrichment exclusively in the forebrain and ileum (Figure 3G). Notably, these mCH-enriched regions contained repetitive sequences, resembling teM patterns found in plants. This suggests a possible mechanism in animals where specific repetitive elements recruit DNMTs, maintaining these regions in a silenced state.

Tissue-specific differences in mCH across different cell types

Distinct mCH patterns were observed among the various tissues and cell types, particularly between PSCs and brain tissues. We identified extensive megabase-scale CH-DMRs (Figure 4A; Supplementary Table S6). Due to the large differences in mCH levels among groups, all CH-DMR analyses were conducted after normalization. Binomial tests were applied to each mCH site based on the conversion rate of each methylation group, with $P<0.01$ indicating significant methylation at that site. After excluding non-methylated and low-coverage sites, all mCH sites were analyzed. In bESCs, mCH predominantly followed a CAG pattern, while this preference was absent in bEPSCs-AGS. In brain tissues, the dominant mCH pattern changed to CAC, with this pattern becoming more pronounced with development (Figure 4B). Using bESCs-F7 as a control, CH-DMRs were compared across bEPSCs-AGS (C1), fetal hindbrain (C2), fetal forebrain (C3), adult forebrain (C4), ileum, and spleen. Overall, mCH distribution remained relatively stable in bESCs, spleen, and ileum, with no CH-DMRs significantly higher than adjacent regions. In contrast, brain tissues exhibited many CH-DMRs distinct from other samples, suggesting a link to tissue-specific function. A total of 929 (C1), 1 138 (C2), 1 130 (C3), and 1 016 (C4) differential regions were identified (Figure 4C), with an average length of 274 927 bp.

Notably, bESCs-F7-5 and bESCs-F7-50, treated with 5 μmol/L and 50 μmol/L MM-102, respectively, showed higher mCH levels (~2.5%) compared to untreated bESCs-F7. Additionally, bEPSCs-AGS exhibited slightly higher mCH levels in specific CH-DMRs compared to adjacent regions (Figure 4D). These regions primarily include genes involved in cell differentiation and cell cycle signaling pathways, indicating that MM-102 may promote bESC pluripotency (Figure 4E). We then examined the mCG and mCH levels near some key pluripotency genes. Further examination of mCG and mCH levels near key pluripotency genes across different PSCs showed consistent mCG levels and low mCH levels (<0.1) (Supplementary Figure S4). All differential regions were annotated, and gene expression levels were assessed in overlapping CH-DMRs between adult brain tissues and other regions. Compared with iPSCs, most CH-DMRs in somatic tissues were clustered on functional genes associated with tissue-specific functions (Lister et al., 2011; Ma et al., 2014). Fetal and adult brain tissues shared similar mCH patterns, resulting in substantial overlap among regions C2, C3, and C4. A total of 250 genes were common across these three clusters (Figure 4F), indicating a high level of mCH conservation during brain tissue development. The highly conserved, brain-specific TF LIM-only 3 (LMO3), known for its specific expression in the nervous system during development (Sang et al., 2014), exhibited elevated mCH levels near its promoter while maintaining stable transcription, a phenomenon not observed in other tissues (Figure 4G).

High mCH levels were frequently correlated with relatively low gene expression (two-sided *t*-test, $P<0.0001$) (Figure 4H) (Lagger et al., 2017; He et al., 2020), suggesting that mCH

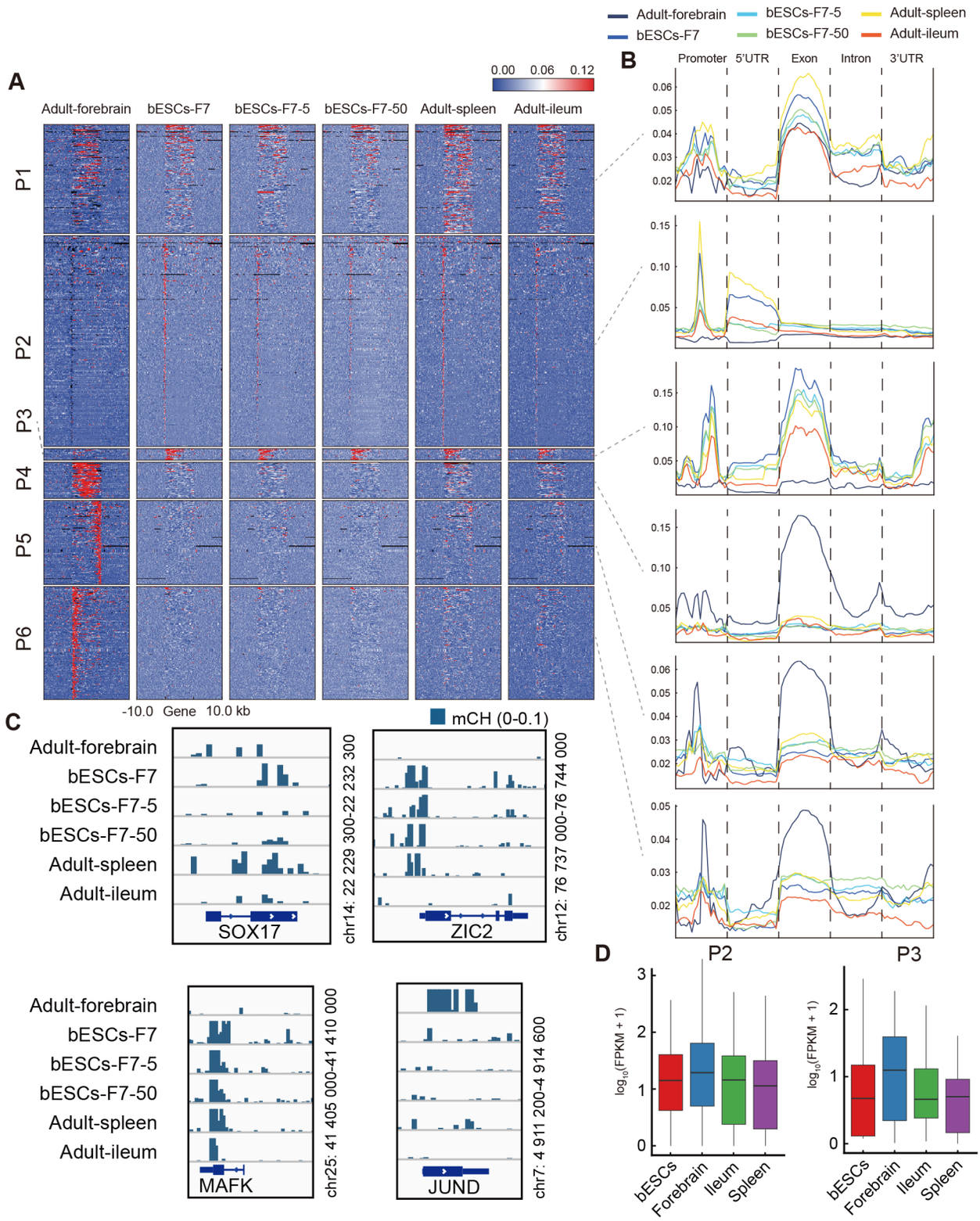

Figure 2 Specific mCH patterns in bovine embryonic stem cells (bESCs) and different tissues

A: K-means were used to cluster genes and flanking regions (±10 kb) in ESCs (bESCs-F7/F7-5/F7-50) and tissues (forebrain, spleen, ileum), with classification into six groups based on different mCH enrichment patterns (colors represent mCH levels). Right side displays average mCH levels of genes and their flanking regions (±10 kb) within each group. B: Average mCH levels of genes within the group in the promoter, 5'UTR, exon, intron, and 3'UTR regions. C: mCH levels in *SOX17*, *ZIC2*, *MAFK*, and *JUND* gene regions, ranging from 0 to 0.1. D: Gene expression levels in clusters P2 and P3, with levels transformed by $\log_{10}$ (FPKM+1). Boxplots include upper quartile, median, lower quartile, and outliers above top whisker.

may function as a negative regulator of gene expression. CH-DMR analysis revealed distinct mCH distribution patterns and base preferences between the forebrain and PSCs. This base preference appears to be influenced by structural differences in DNMT3A-DNA and DNMT3B-DNA complexes, with DNMT3A favoring CAC sites and DNMT3B targeting CAG

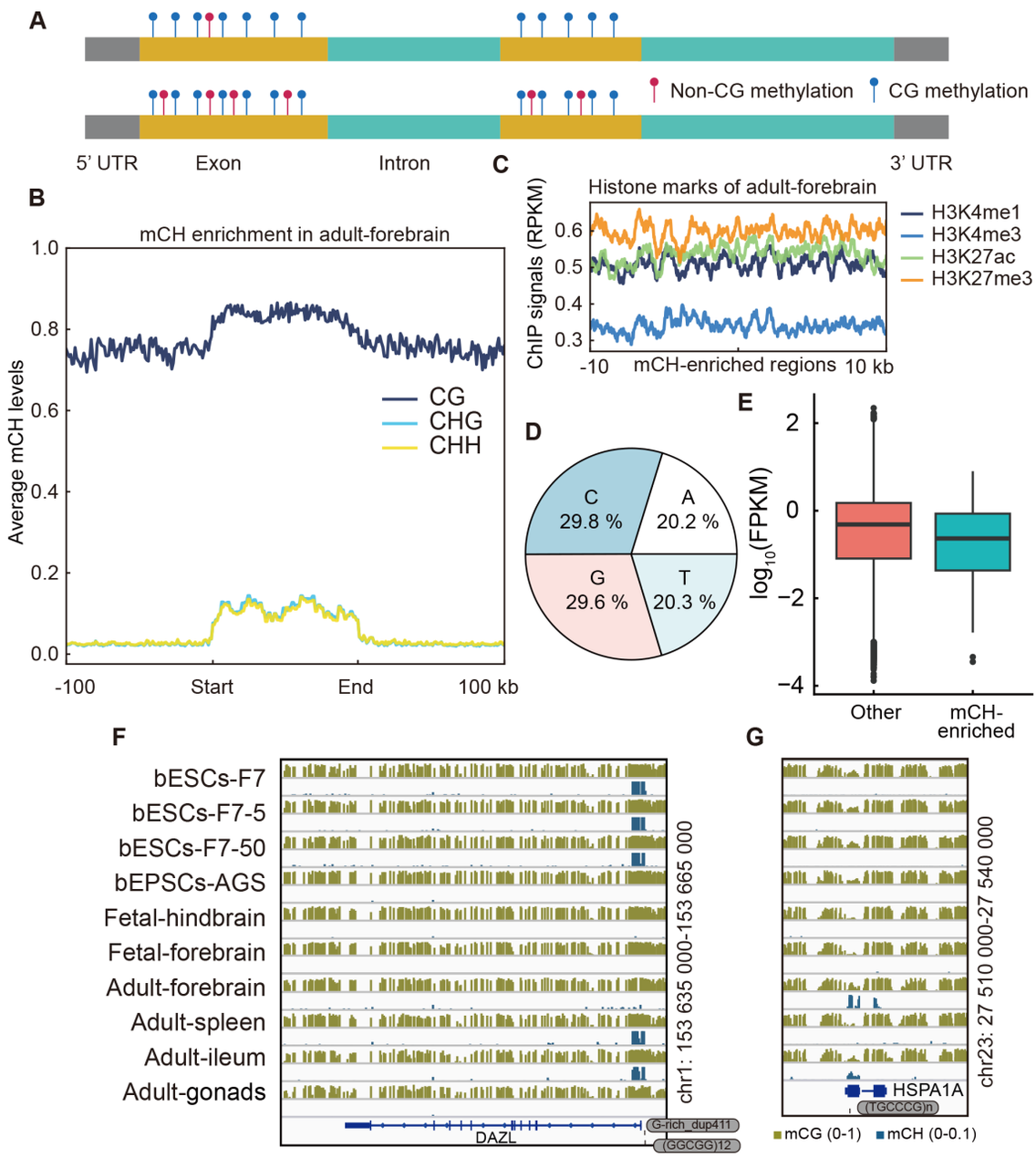

Figure 3 mCH-enriched regions in bovine genome

A: Schematic of enriched mCH exons and common exons. B: Average mCG and mCH levels in mCH-enriched genes and flanking regions (±100 kb) in adult-forebrain. C: Four histone marks in mCH-enriched regions and flanking regions (±10 kb) in adult-forebrain. D: GC content in mCH-enriched regions. E: Comparison of expression levels of genes in mCH-enriched regions with other genes. F: mCG and mCH in *DAZL* region. G: mCG and mCH in *HSPA1A* region.

sites (Gao et al., 2020; Zhang et al., 2018). Enzymology and enhanced reduced representation bisulfite sequencing data support this distinction (Gao et al., 2020). Differential expression patterns of *DNMT3A* and *DNMT3B* between bESCs and somatic tissues further corroborated these preferences. In PSC lines, *DNMT3B* expression was significantly higher than in brain tissues, whereas, in the forebrain, *DNMT3A* expression was higher than *DNMT3B* expression (Figure 4I; Supplementary Table S2). GO enrichment analysis indicated that cluster C1 was enriched in metabolic pathways, clusters C2 and C3 in the fetal forebrain and fetal hindbrain were enriched in pathways related to neurodevelopment and cell differentiation, and cluster C4 in the adult forebrain was enriched in pathways related to systemic development (Figure 4J; Supplementary Table S7). These findings suggest that mCH is not randomly distributed across the genome but is enriched in regions related to the functional roles of the tissues and cell types studied.

Dynamic changes in mCH during forebrain development

While DMR analysis showed broadly similar mCH patterns between the fetal and adult forebrain, notable differences were also identified. Comparing mCH differences in the fetal and adult forebrain, we identified 541 mCH-enriched regions in the fetal forebrain and 792 mCH-enriched regions in the adult forebrain (Figure 5A; Supplementary Table S8). Although most CH-DMRs were evenly distributed throughout the genome, the adult X chromosome exhibited a distinct up-regulation in CH-DMRs, possibly related to X chromosome

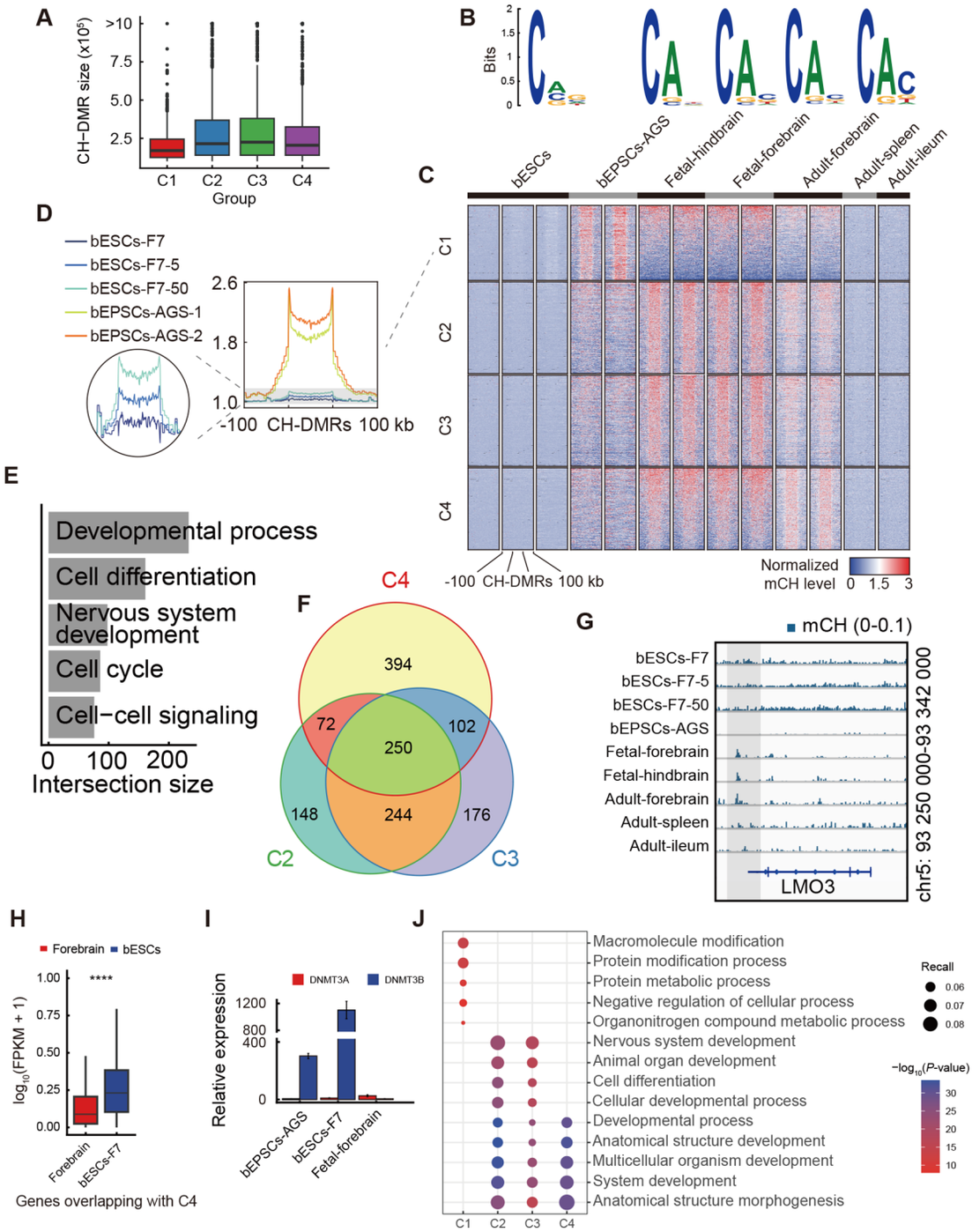

Figure 4 mCH accumulation across different samples

A: Distribution of CH differentially methylated regions (CH-DMRs) sizes in four groups. Boxplots include upper quartile, median, lower quartile, and outliers above top whisker. B: Sequence context preference for mCH, excluding cases containing unknown base N. C: mCH levels in four CH-DMR clusters and their flanking regions (±100 kb), with four groups representing bovine embryonic stem cells (bESCs) vs. bovine extended pluripotent stem cells (bEPSCs-AGS) (C1), bESCs vs. fetal-hindbrain (C2), bESCs vs. fetal-forebrain (C3), and bESCs vs. adult-forebrain (C4), respectively. All mCH levels were normalized by median. Forebrain was approximated using frontal cortex. D: mCH levels in four PSC samples (F7, F7-5, F7-50, and AGS) in AGS-specific CH-DMRs (C1) and its flanking regions (±100 kb), where F7-5 and F7-50 groups with added MM-102 exhibited higher mCH levels compared to F7. E: GO-BP enrichment results of genes overlapping with C1. F: Genes overlapping with CH-DMRs in clusters C2, C3, and C4. G: mCH profile of *LMO3*. Height represents mCG level. H: Expression levels of genes overlapping with C4 cluster in bESCs and adult forebrain. Statistical analysis was conducted using a two-sided *t*-test. ****: $P<0.0001$. Expression levels of genes are represented using $\log_{10}$-transformed FPKM values. I: Expression levels of *DNMT3A* (red) and *DNMT3B* (blue) in three samples (bEPSCs-AGS, bESCs-F7 and fetal-forebrain), based on qPCR. qPCR expression levels were calculated using standard relative Ct values ($2^{-\triangle\triangle Ct}$). Data represent mean±standard deviation (SD) from three independent experiments. J: GO-BP enrichment results of genes overlapping with CH-DMRs in first four clusters (C1, C2, C3, and C4), size of bubble represents the ratio of number of genes to total number of genes under the GO entry, and color represents *P*-value of the hypergeometric distribution test (converted by $-\log_{10}$).

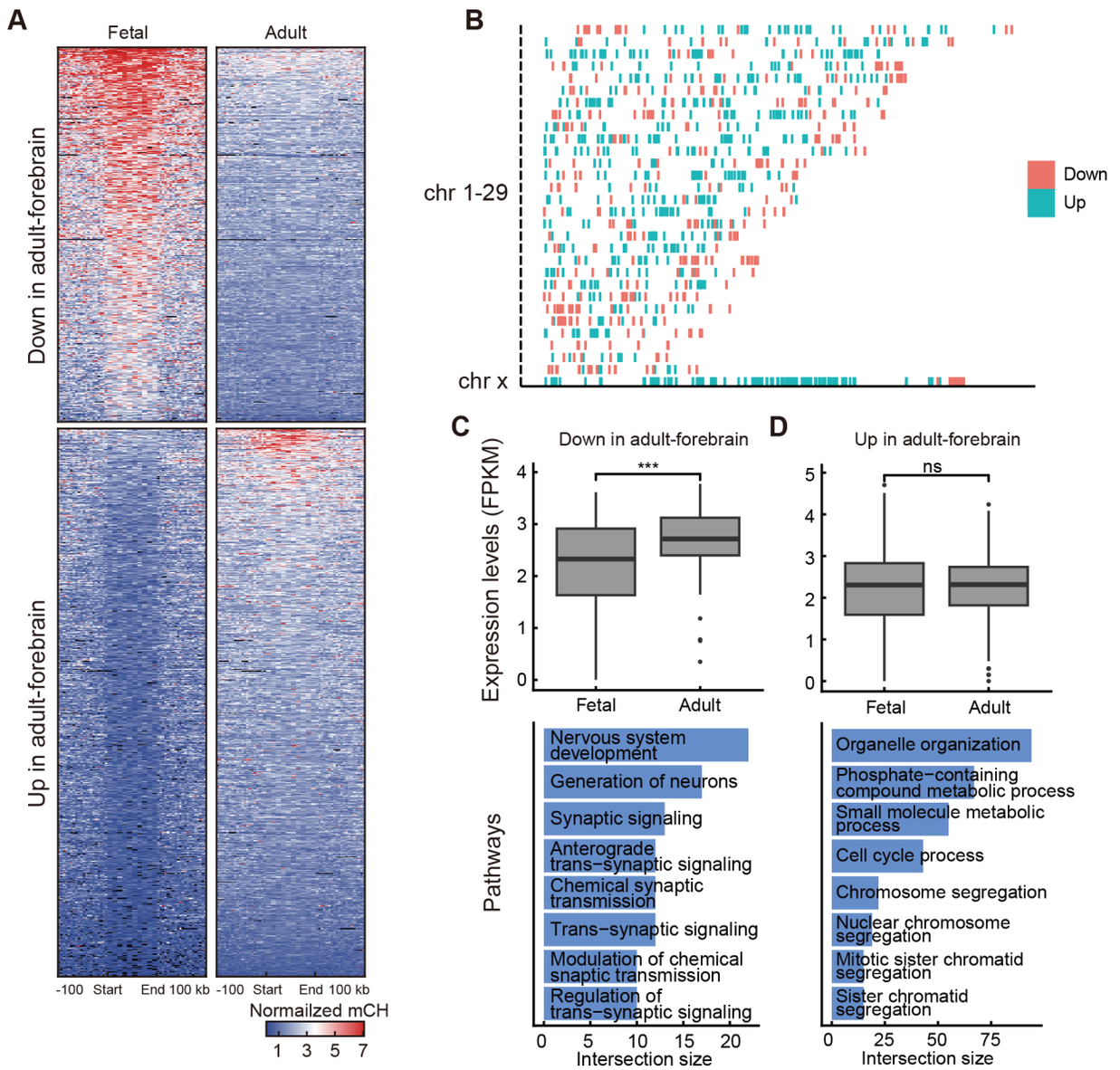

Figure 5 Dynamic changes in mCH in forebrain during development

A: mCH levels of CH-DMRs between fetal and adult forebrain and their flanking regions (±100 kb). All mCH levels were normalized by median. B: Distribution of CH-DMRs in chromosomes obtained by comparing fetal and adult forebrain. C: Gene expression levels near CH-DMRs down-regulated in adult forebrain, with expression levels $\log_{10}$ normalized (above). Corresponding gene GO-BP-enriched pathways (below). D: Gene expression levels near CH-DMRs up-regulated in adult forebrain, with expression levels $\log_{10}$ normalized (above). Corresponding gene GO-BP-enriched pathways (below). ns: Not significant; ***: P<0.001.

inactivation in adults (Figure 5B) (Cotton et al., 2015; Schultz et al., 2015). Genes enriched with mCH in the fetal forebrain displayed significantly lower expression levels compared to those in the adult forebrain. These genes are closely associated with forebrain specialization and neural development (Figure 5C). Correlation analyses for CHG and CHH methylation revealed a negative correlation between mCH in gene-proximal regions and gene expression (CHG: correlation=–0.26, P=0.001; CHH: correlation=–0.23, P=0.003). This negative correlation suggests a potential role for mCH in regulating tissue-specific gene expression during brain development, as these genes become active and contribute to functional differentiation at later stages. In contrast, genes enriched with mCH in the adult forebrain showed no significant differences compared to the fetal forebrain, with these genes primarily related to cell growth and development (Figure 5D).

Presence of mCH is correlated with mCG location

We next investigated the relationship between mCH and mCG in the forebrain. In the adult forebrain, regions with notably low mCH levels also displayed significantly reduced mCG signals compared to adjacent regions, a pattern not observed in the fetal forebrain (Figure 6A). This suggests that mCH may be contingent on mCG. The previously identified large-scale up-regulated CH-DMRs in the adult X chromosome reinforce this association. One plausible explanation is that both mCH and mCG are catalyzed by the *DNMT3A/B* enzymes, which show a marked preference for CpG dinucleotides. However, in mice, DNMT3A/B also promotes cytosine methylation in non-CpG contexts at a rate 40–500 times lower than that in CpG contexts (Ramsahoye et al., 2000; Jeltsch, 2006), leading to a positive correlation between mCH and mCG levels. In PSCs

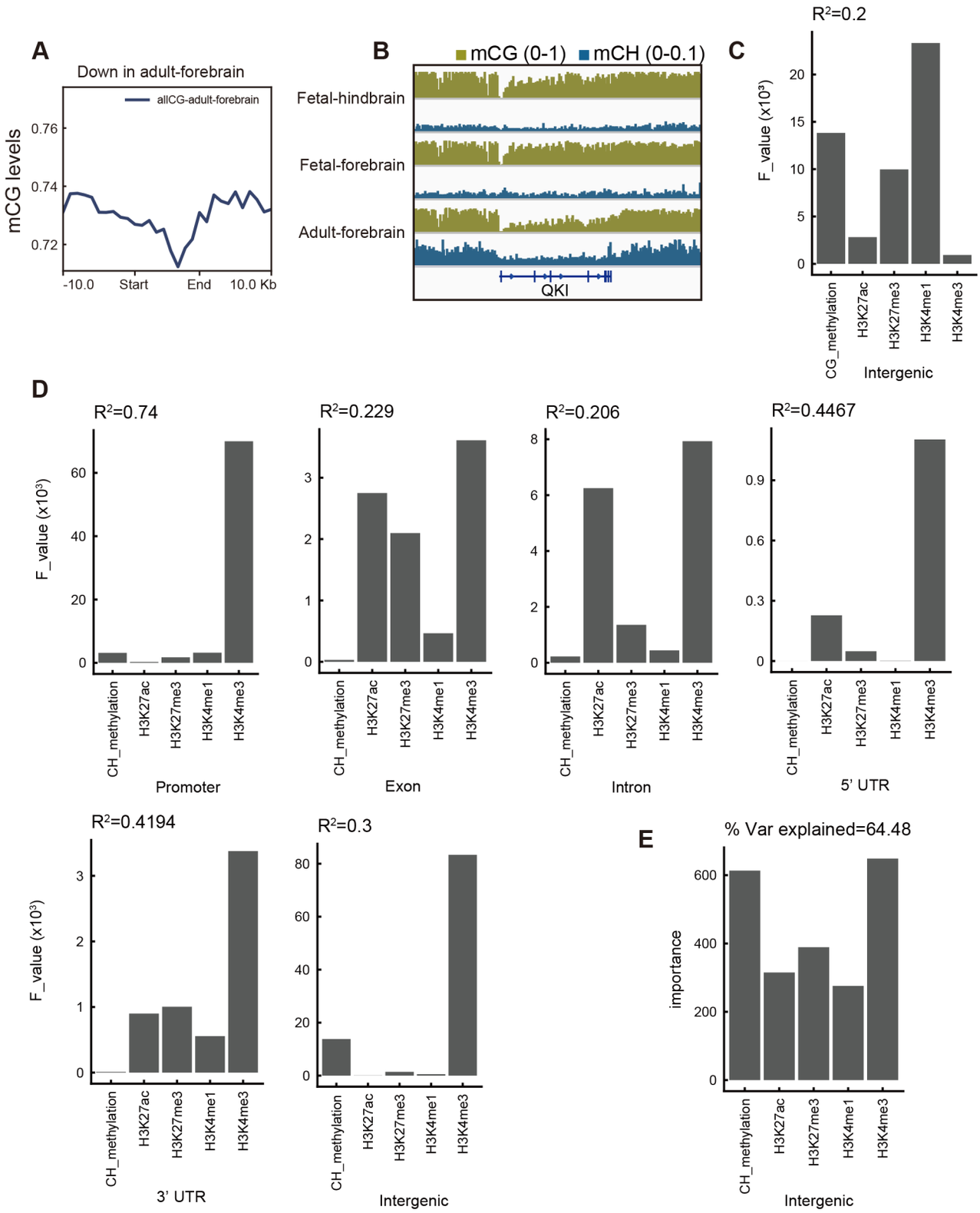

Figure 6 Positional relationship between mCH and mCG

A: Average mCG levels in all genes and flanking regions (±10 kb) in adult forebrain. B: mCG and mCH profiles of *QKI*. Khaki represents mCG sites, height represents mCG level, ranging from 0 to 1; dark blue represents mCH sites, height represents mCH level, ranging from 0 to 0.1. C: Contribution of features predictive for mCH in a linear model based on 5 kb tiling of intergenic regions in bovine genome. D: Contribution of features predictive of mCG in a linear model based on 5 kb tiling of promoters, exons, introns, 5' UTR, 3' UTR, and intergenic regions in bovine genome. E: Contribution of four histone marks and mCH levels in intergenic regions to predicting mCG levels using a random forest regression model.

undergoing global methylation remodeling, *DNMT3A/B* perform *de novo* methylation across most regions of the genome, which could result in the co-occurrence of both mCG and mCH in the genome. In the forebrain, extended low-methylation regions (spanning hundreds of kilobases) exhibit reduced mCH levels relative to adjacent regions. For instance,

QKI, a member of the signal transduction and activation of RNA (STAR) family of RNA-binding proteins, is crucial for oligodendrocyte differentiation and myelin synthesis. In the adult forebrain, our results revealed an extensive low-mCG methylation region spanning hundreds of kilobases near the *QKI* gene, with notably reduced mCH levels within the gene compared to adjacent regions (Figure 6B).

To quantitatively assess the relationship between mCH and mCG, we applied a linear model, with mCH as the dependent variable and mCG, H3K4me3, H3K27ac, H3K4me1, and H3K27me3 levels as independent variables. Models were constructed for different genomic regions, including promoters, 5'UTR, 3'UTR, exons, introns, and distal regions. However, proximal regions yielded low explanatory power, with *R*-squared values generally below 0.1. Only the model for distal regions achieved an *R*-squared value of 0.2 (Figure 6C). To evaluate feature contributions to the model, we calculated *F*-statistics for each variable using analysis of variance and ranked them by *F*-value. Results showed that mCG had the second-highest *F*-value, indicating a certain correlation between mCG and mCH in distal regions. We then reversed the approach to predict mCG levels using mCH and the four histone modifications. As expected, models for promoters, 3' UTR, and 5'UTR performed well, with H3K4me3 contributing the most to the model (Figure 6D). In distal regions, the *R*-squared value of the model was 0.3, showing some improvement, and mCH also had the second-highest *F*-value. To further examine this relationship, we applied random forest regression to the distal regions, achieving an overall explanatory power of 64.48%. Using the "importance=TRUE" parameter, we identified mCH as the second-most important variable in the model (Figure 6E). These findings demonstrate a correlation between mCH and mCG in distal regions.

DISCUSSION

Research on mCH has mainly focused on model species and humans (Lister et al., 2009; Laurent et al., 2010; Lister et al., 2013; He et al., 2020), with limited investigations on ruminants. Additionally, these studies have concentrated on mCH distribution in PSCs and somatic tissues, often overlooking its functional implications. Here, we observed notable mCH accumulation in the spleen, ileum, brain, and bESCs. Our results also revealed significant differences in mCH patterns between brain tissue and bESCs, characterized by enrichment in specific functional regions and an inverse relationship with functional gene expression, emphasizing the role of mCH in tissue-specific function.

Our results also showed significant differences in mCH patterns between bESCs and somatic tissues, including differences in enrichment and accumulation patterns. In bESCs, mCH predominantly localized at 5' CAG 3' sites, accompanied by high expression of *DNMT3B*, whereas, in the forebrain, mCH favored the 5' CAC 3' pattern. This suggests that different mCH patterns in bESCs and somatic tissues may result from selective activity by distinct different DNMTs. Structural differences between DNMT3A and DNMT3B play a key role in determining their substrate preferences for mCH. In the DNMT3B-DNA complex, the N779 residue interacts specifically with CpA bases, while K777 forms a side-chain hydrogen bond with the N7 atom of G7, explaining the preference of DNMT3B for G at the +1 position in non-CpG contexts (Gao et al., 2020; Zhang et al., 2018;). This structural divergence explains the observed methylation pattern differences between DNMT3A and DNMT3B, particularly on non-CG sites (Gao et al., 2020). Consequently, DNMT3B expression in bESCs drives mCH accumulation in line with its CH preference, while DNMT1 likely maintains mCH during self-renewal. High mCH levels are observed in oocytes (Shirane et al., 2013; Guo et al., 2014a; Wang et al., 2014), which gradually decrease as embryonic development progresses. In mice, three *Dnmt1* variants exist, *Dnmt1s*, *Dnmt1o*, and *Dnmt1p*, with *Dnmt1o* showing specificity for oocytes and pre-implantation embryos (Ko et al., 2005). Bovines contain two *DNMT1* variants, namely *DNMT1a* (*DNMT1o*) and *DNMT1b* (Russell & Betts, 2008). *DNMT1o* relies on the initial exons for expression and is active in oocytes and pre-implantation embryos, suggesting a potential role in non-CG methylation in response to the high mCH levels observed in oocytes. Based on RNA-seq, our results showed that bESCs had more reads at the first few exons of *DNMT1* compared to other samples, indicating possible alternative splicing for *DNMT1o*, with the qPCR results also confirming the expression of *DNMT1o* in bESCs. These findings suggest that *DNMT1o* may be involved in mCH maintenance in bESCs (Russell & Betts, 2008; Mcgraw et al., 2013; Whidden et al., 2016). All three bESC lines in this study (bESCs-F7/F7-5/F7-50) were derived from *in vitro* fertilization embryos (Bogliotti et al., 2018; Li et al., 2023), with the sustained expression of *DNMT1o* potentially contributing to the observed accumulation of mCH. Although bEPSCs are also derived from embryos (Zhao et al., 2021), *DNMT1o* expression in bESCs was significantly lower than in bEPSCs, which may explain the overall lower mCH level. In brain and other somatic tissues, mCH accumulation is driven by distinct mechanisms, where mature synaptic development of neural circuits is accompanied by large-scale reconstruction of the neuronal epigenome (Lister et al., 2013), coinciding with gradual mCH accumulation in brain tissue (He et al., 2020). Knockout of *Dnmt3a* in the central nervous system during late pregnancy can cause motor deficits and premature death in mice (Nguyen et al., 2007), highlighting the crucial role of *DNMT3A* in establishing a normal DNA methylation profile in the brain and supporting healthy brain development, thereby indirectly affecting mCH accumulation.

The distribution patterns of mCH vary among different cell types, with most mCH locating in intragenic and upstream/downstream regions. Based on clustering analysis, these patterns were classified into six distinct distribution types. Specifically, clusters P2 and P3 were enriched in bESCs, spleen, and ileum, whereas clusters P4–P6 were enriched in the forebrain. Many of these genes in these mCH-enriched clusters encode transcription factors and signaling pathways critical for tissue-specific functions. We also observed highly mCH-enriched regions within certain exons, with these genes displaying not only elevated mCH levels compared to adjacent regions but also slightly reduced expression levels relative to other genes. Internal GC content was significantly higher than that of the flanking regions, and related GC repeat sequences were identified. The high GC content led to methyltransferase activity, resulting in higher levels of mCG and mCH compared to the flanking region. A similar mechanism has been observed in plants, where double-stranded RNA recruits methyltransferases to methylate repetitive DNA and transposable elements, thereby inhibiting their transcription. We speculate that a comparable regulatory mechanism may exist in animals to regulate the expression of

repetitive sequences. Overall, our findings revealed substantial differences in mCH patterns between brain tissue and bESCs. These differences showed enrichment in specific functional regions and exhibited a negative correlation with the expression of functional genes, suggesting that mCH may play a role in tissue-specific gene regulation.

Research has identified megabase-scale CH-DMRs in iPSCs and ESCs in humans (Lister et al., 2011; Ma et al., 2014). Differential analysis of mCH across tissues and cells revealed that most CH-DMRs were associated with specific functions. In bEPSCs-AGS, mCH was predominantly enriched near metabolism-related genes, while in the forebrain, mCH was primarily enriched near genes related to cell differentiation and neural development. This forebrain-specific mCH distinction may indirectly affect behavior and stress responses in cattle, thereby impacting production performance. These findings indicated that mCH is not randomly distributed but instead shows targeted enrichment in specific genomic regions aligned with tissue functions. As a potential epigenetic regulation, mCH provides insights into both tissue-specific differences and developmental-stage differences. By comparing mCH distribution between fetal and adult brain tissues, we found that dynamic CH-DMRs include genes related to cell growth and development as well as tissue-specific genes related to brain functions. Additionally, during brain development, mCH levels show a strong negative correlation with the expression of neuronal development genes, suggesting that changes in mCH are major determinants of neuronal transcriptional differentiation (Jeong et al., 2021). The specific distribution patterns and regulatory functions of mCH highlight its role in regulating growth and development. Future research should explore how these epigenetic markers could be applied in breeding and production practices to improve bovine productivity and health.

Non-CG methylation exhibited a spatial correlation with mCG, particularly in the forebrain, where CH-DMRs frequently mirrored patterns of mCG variation. Within genes, mCH tended to cluster, forming large CH-DMRs spanning hundreds of kilobases. Using a regression model, we found that mCG was an important contributor to predicting mCH levels, with the correlation becoming stronger toward the distal ends of genes. When we applied a random forest model, the overall interpretability of mCH patterns improved, reaching 64%, with mCG playing a key role in the predictive power of the model. The high overlap of CH-DMR regions between the fetal and adult forebrain indicated high conservation of mCH at the kilobase-scale, suggesting shared methylation and demethylation enzymes for both mCH and mCG. Additionally, distal regions of genes require more precise regulation due to the presence of numerous regulatory elements, including enhancers and transcription factor binding sites, which may explain the higher correlation found between mCG and mCH in these distal regions.

This study reported several key findings: (1) Compared to previous studies on humans and mice, the existence of mCH was confirmed in both bESCs and somatic cells, with pronounced enrichment in brain tissues and bESCs, as well as the spleen and ileum; (2) While mCH distribution was relatively stable in PSCs, spleen, and ileum, brain tissues exhibited distinct mCH patterns in specific regions; (3) Differences in mCH patterns between bESCs and somatic tissues appeared to arise from DNMT3A/B base preferences, with DNMT1o contributing to mCH maintenance in bESCs during self-renewal; (4) Significant mCH enrichment was identified in specific exon regions, which inhibited the expression of corresponding genes, potentially as a mechanism for regulating repetitive sequences and transposable elements; and (5) mCH distribution in the bovine genome was not random but followed regular patterns according to developmental stages and tissue-specific functions. In contrast to mCG, mCH exhibited extensive positional correlation within the genome and possessed distinct, non-redundant functions. This study provides novel insights into the distribution and function of non-CG methylation in different tissues and cells, but further exploration of its specific biological functions and molecular mechanisms is necessary to provide a deeper understanding of epigenetic regulation complexities.

DATA AVAILABILITY

The RNA-seq, BS-seq and ChIP-seq data from fetal tissues are available in the Gene Expression Omnibus database (GSE264346), China National Center for Bioinformation database of Genome Sequence Archive (GSA) (PRJCA030053), and Science Data Bank database (doi: 10.57760/sciencedb.j00139.00098). All PSC data are available in the Gene Expression Omnibus database (GSE248161, GSE129760) and NCBI database (PRJNA693452).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

X.L.L. and J.W. designed the study and wrote the manuscript. J.W. analyzed the data. W.Y. and F.L. performed most of the experiments. G.B.L., X.X.G., C.C.Z., C.L., and N.L. performed the research. All authors read and approved the final version of the manuscript.

REFERENCES

Bewick AJ, Schmitz RJ. 2017. Gene body DNA methylation in plants. *Current Opinion in Plant Biology*, **36**: 103–110.

Bogliotti YS, Wu J, Vilarino M, et al. 2018. Efficient derivation of stable primed pluripotent embryonic stem cells from bovine blastocysts. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(9): 2090–2095.

Chen SF, Zhou YQ, Chen YR, et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884–i890.

Cotton AM, Price EM, Jones MJ, et al. 2015. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Human Molecular Genetics*, **24**(6): 1528–1539.

Gao LF, Emperle M, Guo YR, et al. 2020. Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. *Nature Communications*, **11**(1): 3355.

Guo HS, Zhu P, Yan LY, et al. 2014a. The DNA methylation landscape of human early embryos. *Nature*, **511**(7511): 606–610.

Guo JU, Su YJ, Shin JH, et al. 2014b. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*, **17**(2): 215–222.

Haines TR, Rodenhiser DI, Ainsworth PJ. 2001. Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. *Developmental Biology*, **240**(2): 585–598.

He YP, Ecker JR. 2015. Non-CG methylation in the human genome. *Annual Review of Genomics and Human Genetics*, **16**: 55–77.

He YP, Hariharan M, Gorkin DU, et al. 2020. Spatiotemporal DNA

methylome dynamics of the developing mouse fetus. *Nature*, **583**(7818): 752–759.
Hu SB, Cheng L, Wen B. 2012. Large chromatin domains in pluripotent and differentiated cells. *Acta Biochimica et Biophysica Sinica*, **44**(1): 48–53.
Jeltsch A. 2006. Molecular enzymology of mammalian DNA methyltransferases. *In*: Doerfler W, Böhm P. Current Topics in Microbiology and Immunology. Berlin: Springer, 203–225.
Jeong H, Mendizabal I, Berto S, et al. 2021. Evolution of DNA methylation in the human brain. *Nature Communications*, **12**(1): 2021.
Kern C, Wang Y, Xu XQ, et al. 2021. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nature Communications*, **12**(1): 1821.
Ko YG, Nishino K, Hattori N, et al. 2005. Stage-by-stage change in DNA methylation status of *Dnmt1* locus during mouse early development. *Journal of Biological Chemistry*, **280**(10): 9627–9634.
Lagger S, Connelly JC, Schweikert G, et al. 2017. MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. *PLoS Genetics*, **13**(5): e1006793.
Laurent L, Wong E, Li GL, et al. 2010. Dynamic changes in the human methylome during differentiation. *Genome Research*, **20**(3): 320–331.
Li C, Han XJ, Wang J, et al. 2023. Mixed-lineage leukemia 1 inhibition enhances the differentiation potential of bovine embryonic stem cells by increasing H3K4 mono-methylation at active promoters. *International Journal of Molecular Sciences*, **24**(15): 11901.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**(14): 1754–1760.
Li LL, Medina-Menéndez C, García-Corzo L, et al. 2022. *SoxD* genes are required for adult neural stem cell activation. *Cell Reports*, **38**(5): 110313.
Liao J, Karnik R, Gu HC, et al. 2015. Targeted disruption of *DNMT1, DNMT3A* and *DNMT3B* in human embryonic stem cells. *Nature Genetics*, **47**(5): 469–478.
Lister R, Mukamel EA, Nery JR, et al. 2013. Global epigenomic reconfiguration during mammalian brain development. *Science*, **341**(6146): 1237905.
Lister R, Pelizzola M, Dowen RH, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**(7271): 315–322.
Lister R, Pelizzola M, Kida YS, et al. 2011. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*, **471**(7336): 68–73.
Ma H, Morey R, O'Neil RC, et al. 2014. Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature*, **511**(7508): 177–183.
Matsuo K, Owens JM, Tonko M, et al. 2000. *Fosl1* is a transcriptional target of c-Fos during osteoclast differentiation. *Nature Genetics*, **24**(2): 184–187.
McGraw S, Oakes CC, Martel J, et al. 2013. Loss of DNMT1o disrupts imprinted X chromosome inactivation and accentuates placental defects in females. *PLoS Genetics*, **9**(11): e1003873.
Murillo B, Ruiz-Reig N, Herrera M, et al. 2015. Zic2 controls the migration of specific neuronal populations in the developing forebrain. *Journal of Neuroscience*, **35**(32): 11266–11280.
Muyle AM, Seymour DK, Lv YD, et al. 2022. Gene body methylation in plants: mechanisms, functions, and important implications for understanding evolutionary processes. *Genome Biology and Evolution*, **14**(4): evac038.
Nguyen S, Meletis K, Fu DD, et al. 2007. Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Developmental Dynamics*, **236**(6): 1663–1676.
Ou J, Wolfe SA, Brodsky MH, et al. 2018. motifStack for the analysis of transcription factor binding site evolution. *Nature Methods*, **15**(1): 8–9.
Pedersen BS, Eyring K, De S, et al. 2014. Fast and accurate alignment of long bisulfite-seq reads. arXiv: 1401.1129
Pereira Luppi M, Azcorra M, Caronia-Brown G, et al. 2021. *Sox6* expression distinguishes dorsally and ventrally biased dopamine neurons in the substantia nigra with distinctive properties and embryonic origins. *Cell Reports*, **37**(6): 109975.
Ramírez F, Ryan DP, Grüning B, et al. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, **44**(W1): W160–W165.
Ramsahoye BH, Biniszkiewicz D, Lyko F, et al. 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America*, **97**(10): 5237–5242.
Reimand J, Kull M, Peterson H, et al. 2007. g: profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research*, **35**(suppl_2): W193–W200.
Russell DF, Betts DH. 2008. Alternative splicing and expression analysis of bovine DNA methyltransferase 1. *Developmental Dynamics*, **237**(4): 1051–1059.
Sang M, Ma L, Sang MJ, et al. 2014. LIM-domain-only proteins: multifunctional nuclear transcription coregulators that interacts with diverse proteins. *Molecular Biology Reports*, **41**(2): 1067–1073.
Schultz MD, He YP, Whitaker JW, et al. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*, **523**(7559): 212–216.
Schultz MD, He YP, Whitaker JW, et al. 2016. Corrigendum: human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*, **530**(7589): 242.
Shirane K, Toh H, Kobayashi H, et al. 2013. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genetics*, **9**(4): e1003439.
Smorag L, Xu XB, Engel W, et al. 2014. The roles of DAZL in RNA biology and development. *WIREs RNA*, **5**(4): 527–535.
Töröcsik B, Angelastro JM, Greene LA. 2002. The basic region and leucine zipper transcription factor MafK is a new nerve growth factor-responsive immediate early gene that regulates neurite outgrowth. *Journal of Neuroscience*, **22**(20): 8971–8980.
Wang L, Zhang J, Duan JL, et al. 2014. Programming and inheritance of parental DNA methylomes in mammals. *Cell*, **157**(4): 979–991.
Whidden L, Martel J, Rahimi S, et al. 2016. Compromised oocyte quality and assisted reproduction contribute to sex-specific effects on offspring outcomes and epigenetic patterning. *Human Molecular Genetics*, **25**(21): 4649–4660.
Wu TZ, Hu EQ, Xu SB, et al. 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*, **2**(3): 100141.
Zeng YB, Dawe RK, Gent JI. 2023. Natural methylation epialleles correlate with gene expression in maize. *Genetics*, **225**(2): iyad146.
Zhang ZM, Lu R, Wang PC, et al. 2018. Structural basis for DNMT3A-mediated *de novo* DNA methylation. *Nature*, **554**(7692): 387–391.
Zhao LX, Gao XD, Zheng YX, et al. 2021. Establishment of bovine expanded potential stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, **118**(15): e2018505118.
Zhou Y, Connor EE, Bickhart DM, et al. 2018. Comparative whole genome DNA methylation profiling of cattle sperm and somatic tissues reveals striking hypomethylated patterns in sperm. *Gigascience*, **7**(5): giy039.
Zhou Y, Liu SL, Hu Y, et al. 2020. Comparative whole genome DNA methylation profiling across cattle tissues reveals global and tissue-specific methylation patterns. *BMC Biology*, **18**(1): 85.
Zhu LK, Zhou T, Iyyappan R, et al. 2022. High-resolution ribosome profiling reveals translational selectivity for transcripts in bovine preimplantation embryo development. *Development*, **149**(21): dev200819.
Ziller MJ, Müller F, Liao J, et al. 2011. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genetics*, **7**(12): e1002389.