



ORIGINAL RESEARCH

Dynamic DNA 5-hydroxymethylcytosine and RNA 5-methylcytosine Reprogramming During Early Human Development



Xiao Han^{1,#}, Jia Guo^{1,#}, Mengke Wang^{2,3,4,#}, Nan Zhang^{1,#}, Jie Ren^{2,#}, Ying Yang^{2,3,4,5}, Xu Chi², Yusheng Chen^{2,3}, Huan Yao^{2,3}, Yong-Liang Zhao^{2,3,4}, Yun-Gui Yang^{2,3,4,5,*}, Yingpu Sun^{1,*}, Jiawei Xu^{1,*}

¹ Center for Reproductive Medicine, Henan Key Laboratory of Reproduction and Genetics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China

² Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

⁴ Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100049, China

⁵ Institute of Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing 100101, China

Received 6 December 2021; revised 18 April 2022; accepted 5 May 2022

Available online 26 May 2022

Handled by Chengqi Yi

KEYWORDS

Human foetus;
Foetal organogenesis;
DNA 5hmC;
RNA m⁵C;
Post-transcriptional
regulation

Abstract After implantation, complex and highly specialized molecular events render functionally distinct organ formation, whereas how the epigenome shapes organ-specific development remains to be fully elucidated. Here, nano-hmC-Seal, RNA bisulfite sequencing (RNA-BisSeq), and RNA sequencing (RNA-Seq) were performed, and the first multilayer landscapes of DNA 5-hydroxymethylcytosine (5hmC) and RNA 5-methylcytosine (m⁵C) epigenomes were obtained in the heart, kidney, liver, and lung of the human foetuses at 13–28 weeks with 123 samples in total. We identified 70,091 and 503 organ- and stage-specific differentially hydroxymethylated regions (DhMRs) and m⁵C-modified mRNAs, respectively. The key transcription factors (TFs), T-box transcription factor 20 (TBX20), paired box 8 (PAX8), krueppel-like factor 1 (KLF1), transcription factor 21 (TCF21), and CCAAT enhancer binding protein beta (CEBPB), specifically contribute to the formation of distinct organs at different stages. Additionally, 5hmC-enriched Alu elements may participate in the regulation of expression of TF-targeted genes. Our integrated studies reveal a

* Corresponding authors.

E-mail: fccxujw@zzu.edu.cn (Xu J), fccsunyp@zzu.edu.cn (Sun Y), ygyang@big.ac.cn (Yang YG).

Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China

<https://doi.org/10.1016/j.gpb.2022.05.005>

1672-0229 © 2023 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

putative essential link between DNA modification and RNA methylation, and illustrate the epigenetic maps during human foetal organogenesis, which provide a foundation for an in-depth understanding of the epigenetic mechanisms underlying early development and birth defects.

Introduction

Epigenetic modification contains DNA methylation, histone modifications, chromatin accessibility, RNA modifications, *etc.* It not only acts as a heritable and stable indication for the specification of chromatin organization and structure, but also participates in the regulation of transcriptional states during mammalian early development [1]. The epigenomes of human embryos and the adult organs in both physiological and pathological conditions have been investigated extensively [2–8]; however, the epigenetic network implicated in the normal development of human foetal organs remains unclear.

As an essential intermediate product of targeted, active DNA demethylation, 5-hydroxymethylcytosine (5hmC) is oxidized from 5-methylcytosine (5mC) by the ten eleven translocation (TET) family of enzymes (TET1/2/3) [9]. This modification is widely distributed in the locations of promoters, enhancers, and gene bodies, regulating target gene expression [10]. Now accumulating evidence has shown that abundant 5hmC and TET enzymes are involved in mediating developmental progression through regulation of the chromatin state and/or transcription in various embryonic and adult cell types, including zygotes, primordial germ cells, Purkinje neurons, and embryonic stem cells (ESCs) [11–13]. In fertilized oocytes, paternal 5mC is converted to 5hmC, allowing drastic reprogramming of sperm DNA in mice [14]. However, 5hmC accumulation is not a requirement for the initial loss of paternal 5mC. This phenomenon depends on TET3 activity and the deficiency of zygotic DNA methyltransferase 3 alpha (DNMT3A) and DNA methyltransferase 1 (DNMT1) [15]. Although the regulatory power and plasticity of 5hmC have been revealed in human adults [13,16], its function and reprogramming are still unclear during human foetal organ development after implantation.

In addition to DNA modifications, RNA methylations, especially N^6 -methyladenosine (m^6A) and 5-methylcytosine (m^5C), contribute to mammalian development at post-transcriptional regulation level [17,18]. The RNA m^6A profile in human adults [8,19] and mixed foetal samples from 18 to 25 weeks [5] has been investigated, and the tissue-specific differences were found to be highly correlated with developmental processes [5]. Another prevalent modification, m^5C , is reported to regulate mRNA nuclear export, mRNA stability, and disease pathogenesis [18,20]. The conserved, organ-specific, and dynamic features across mammalian transcriptomes of m^5C indicate its potential importance during mammalian development [21].

Here, we have established the first profiles of genome-wide DNA 5hmC and single-base resolution RNA m^5C of various samples, including heart, kidney, liver, and lung, from human foetuses at 13–28 weeks. Importantly, we found that both DNA 5hmC and RNA m^5C not only show organ- and stage-specific characteristics in human foetuses, but also exhibit differential and potential synergistic roles during human foetal development.

Results

Genome-wide profiling of DNA 5hmC in human foetal organs

We collected samples of four major organs, including heart, kidney, liver, and lung, from 8 human foetuses at 13–15, 18–19, 21–23, and 25–28 weeks, with 2–4 replicates for each stage (Figure 1A; Table S1). Nano-hmC-Seal [22] was first conducted for DNA 5hmC profiling and displayed high reproducibility between replicates in different periods (Figure S1A). To investigate genome-wide 5hmC features, we first examined the distributions of normalized reads of several development-related genes (Figure 1B). The 5hmC enrichment of these genes prompted us to consider its organ-specific features in human foetuses at different stages.

A previous study has found that DNA 5hmC is widely distributed on the whole genome [23]. Thus, we first calculated the 5hmC enrichment across all regions and found that most of the 5hmC peaks were localized on vital functional elements (Figure S1B). It preferred to localize on promoters and gene bodies, rather than intergenic regions (Figure S1C, two-sided Student's *t*-test, $P < 3.53 \times 10^{-2}$). This indicated that as functional elements, promoters and gene bodies are DNA 5hmC preferential regions that may play critical roles in human foetal organs. In addition, to evaluate 5hmC profiles and global changes, we then illustrated 5hmC distributions along gene bodies (± 3 kb) (Figure 1C). Meanwhile, the 5hmC distributions of heart, kidney, liver, and lung of human adults from nano-hmC-Seal data further revealed that they were conserved in human foetuses and adults, similar to other mammals [10,24] (Figure S1D). Moreover, the distinct signal intensity of 5hmC among stages in each organ indicated that 5hmC may play different roles during organ development (Figure 1C). For instance, in the foetal heart, the 5hmC signals were relatively more stable than those of other three organs, probably because the foetal heart's anatomical structure was completely matured at 8 weeks and displayed fundamental biological functions. In the kidney, the 5hmC signals reached a maximum at 18–19 weeks belonging to the period of corticomedullary differentiation of the kidney, and the nephron frequently increased simultaneously. In the liver, the 5hmC signals were higher when haematopoiesis began, and even higher when the liver showed a strong haematopoietic ability (Figure 1A and C). Together, these results indicate that DNA 5hmC may dynamically regulate organ development by altering the activity of functional elements in an organ-specific manner.

We next identified the differentially 5hmC-modified bins among the four organs using analysis of variance (ANOVA, fold change > 1.5 , $P < 0.01$) (Figure S1E), and the majority of these bins showed higher 5hmC levels in one organ than in other three organs (84.71%–99.80%). Then adjacent bins were merged into one region called differentially hydroxymethylated region (DhMR) and 173,330–485,961 organ-specific DhMRs were obtained with around 100–200 nt in length (Figure S1F and G). No obvious preference of 5hmC

modification was detected between gene bodies and intergenic regions, which was the same as for different chromosomes (Figure S1H and I). Consistent with the annotation results of the 5hmC peaks in each organ, most of the organ-specific DhMRs were localized in functional elements (Figure 1D). The merged top 1% least significant bins were defined as common hydroxymethylated regions (ChMRs) acting as controls. As expected, 5hmC enrichment in DhMRs was different across the four types of organs. Promoter regions are critical for transcriptional regulation, and epigenetic modifications have been reported to be involved in the regulation of promoter activity [25,26]. For example, active promoters are highly associated with demethylation of the corresponding CpG islands (CGIs), histone H3 lysine 27 acetylation (H3K27ac), and open chromatin. Additionally, 5hmC on bivalent promoters protects from *de novo* methylation for lineage-specific transcription upon differentiation in human embryonic stem cells (hESCs) [27]. In ESCs, 5hmC deficiency on the promoter region of *Nanog* leads to its down-regulation and affects trophectoderm differentiation [28]. Building on the aforementioned results, we compared the distance between organ-specific DhMRs and transcription start sites (TSSs) among four organs to further explore the differences of 5hmC on promoters in human foetal organs (Figure 1E). Most of the organ-specific DhMRs were located near TSSs, with prominent in the liver (Figure 1E, two-sided Wilcoxon and Mann–Whitney test, $P < 9.81 \times 10^{-6}$).

We further identified 10,658–20,419 (70,091 in total) organ- and stage-specific DhMRs (ANOVA, fold change > 1.5, $P < 0.01$), and found that samples from the same organ clustered more closely, whereas that samples from the same stage were scattered on the periphery (Figure 1F; Table S2). The results indicate that 5hmC could serve as a potential marker specific for human organs with different developmental stages. Moreover, cluster analysis showed that 5hmC signal intensity and the corresponding functions of the DhMRs were both organ- and stage-specific (Figure 1G and H). Taken together, these results suggest that organ- and stage-specific 5hmC modification is associated with morphogenesis and organ formation during early human development.

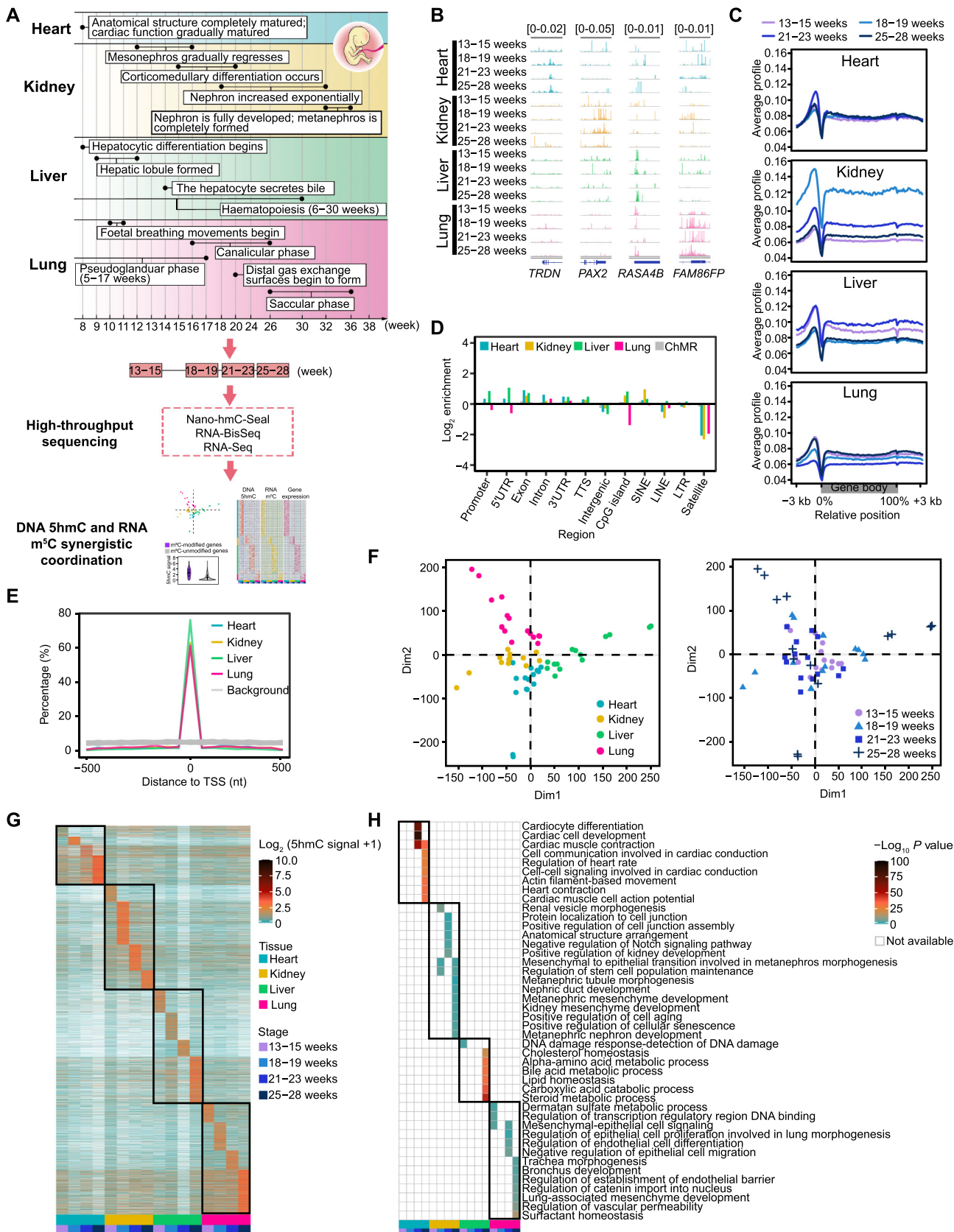
The reprogramming of DNA 5hmC during organ development

To further explore the dynamic sequential changes of 5hmC, we first classified the DhMRs into six clusters in each organ

according to the overall changes of 5hmC signals in different stages, considering that it contains almost all types of non-repetitive trends for six clusters (Figure 2A, Figure S2A). Most of the DhMR-containing genes belonged to at least two clusters, suggesting that the same gene may be regulated through multiple pathways in a 5hmC-dependent manner (Figure S2B). Therefore, we separated these DhMRs from different clusters into two groups: a single-regulation group (a gene set with DhMRs only belonging to one cluster) and a mixed-regulation group (a gene set with DhMRs belonging to at least two clusters) (Figure 2B). To explore how transcription factors (TFs) involved in 5hmC-dependent organ developmental processes, we performed RNA-Seq on samples used for nano-hmC-Seal and TF prediction on the DhMRs from both groups (Table S3). There was a strong correlation between replicates (Figure S2C). The results illustrated that among all the TFs enriched in single- and mixed-regulation groups, 93.08% and 38.77% of them were organ-specific TFs, respectively (Figure S2D; Table S4). Interestingly, 81.82% and 81.25% of these organ-specific TFs also showed organ-specific RNA abundance ($P < 0.01$, fold change > 1.5) in the two regulation groups (Figure S2E and F). Specifically, the top 15 TFs significantly related to development and the top 3 commonly enriched TFs are shown in Figure 2B. In both the single- and mixed-regulation groups, development-related TFs, including T-box transcription factor 20 (TBX20), paired box 8 (PAX8), Krueppel-like factor 1 (KLF1), and transcription factor 21 (TCF21), were highly enriched on DhMRs with an organ-specific expression feature (Figure 2B). Notably, high 5hmC signals were detected on the DhMRs from Cluster 5 in the heart at 21–28 weeks, Cluster 3 in the kidney at 18–23 weeks, Cluster 1 in the liver at 13–15 weeks, and Cluster 3 in the lung at 18–23 weeks, which were specifically enriched by TBX20, PAX8, KLF1, and TCF21, respectively. For example, TBX20 is a transcriptional activator/repressor required for cardiac development [29]. The significant enrichment of this TF suggests its key role in the later stage of development, conferring full maturity of foetal heart. PAX8 has been reported to be highly associated with the mesenchymal to epithelial transition involving metanephros morphogenesis [30]. From 18–23 weeks, corticomedullary differentiation is ongoing, and the nephron frequently increases simultaneously. PAX8 may mainly function during this period. A significantly higher binding of most TFs was also found in human adults, but the

Figure 1 DNA 5hmC landscape of early human foetal development

A. Developmental timelines of human foetal organs and schematic illustration of the organ samples and strategy used in this study. **B.** IGV views of organ-specific genes (*TRDN*, *PAX2*, *RASA4B*, and *FAM86FP*). **C.** Metagene profiles of DNA 5hmC across different developmental stages in heart, kidney, liver, and lung. **D.** Normalized enrichment score of DhMRs and ChMRs across distinct genomic regions relative to that expected in different organs, with positive values indicating enriched more than expected. **E.** Distribution of DhMRs at TSSs of different organs. Genome fragments in length of the average length of DhMRs serve as background ($P < 9.81 \times 10^{-6}$). The P values were calculated using two-sided Wilcoxon and Mann–Whitney tests. **F.** PCA of 5hmC signals in organ- and stage-specific DhMRs from different organ samples. **G.** Heatmap showing the 5hmC signals of organ- and stage-specific DhMRs. **H.** GREAT analysis results for organ- and stage-specific DhMRs. The color represents the significance of each biological process. 5hmC, 5-hydroxymethylcytosine; IGV, Integrative Genomics Viewer; DhMR, differentially hydroxymethylated region; ChMR, common hydroxymethylated region; UTR, untranslated region; TTS, transcriptional terminal site; SINE, short interspersed element; LINE, long interspersed nuclear element; LTR, long terminal repeat; PCA, principal component analysis; GREAT, Genomic Regions Enrichment of Annotations Tool; *FAM86FP*, family with sequence similarity 86 member F, pseudogene; *PAX2*, paired box 2; *RASA4B*, RAS p21 protein activator 4b; *TRDN*, triadin.



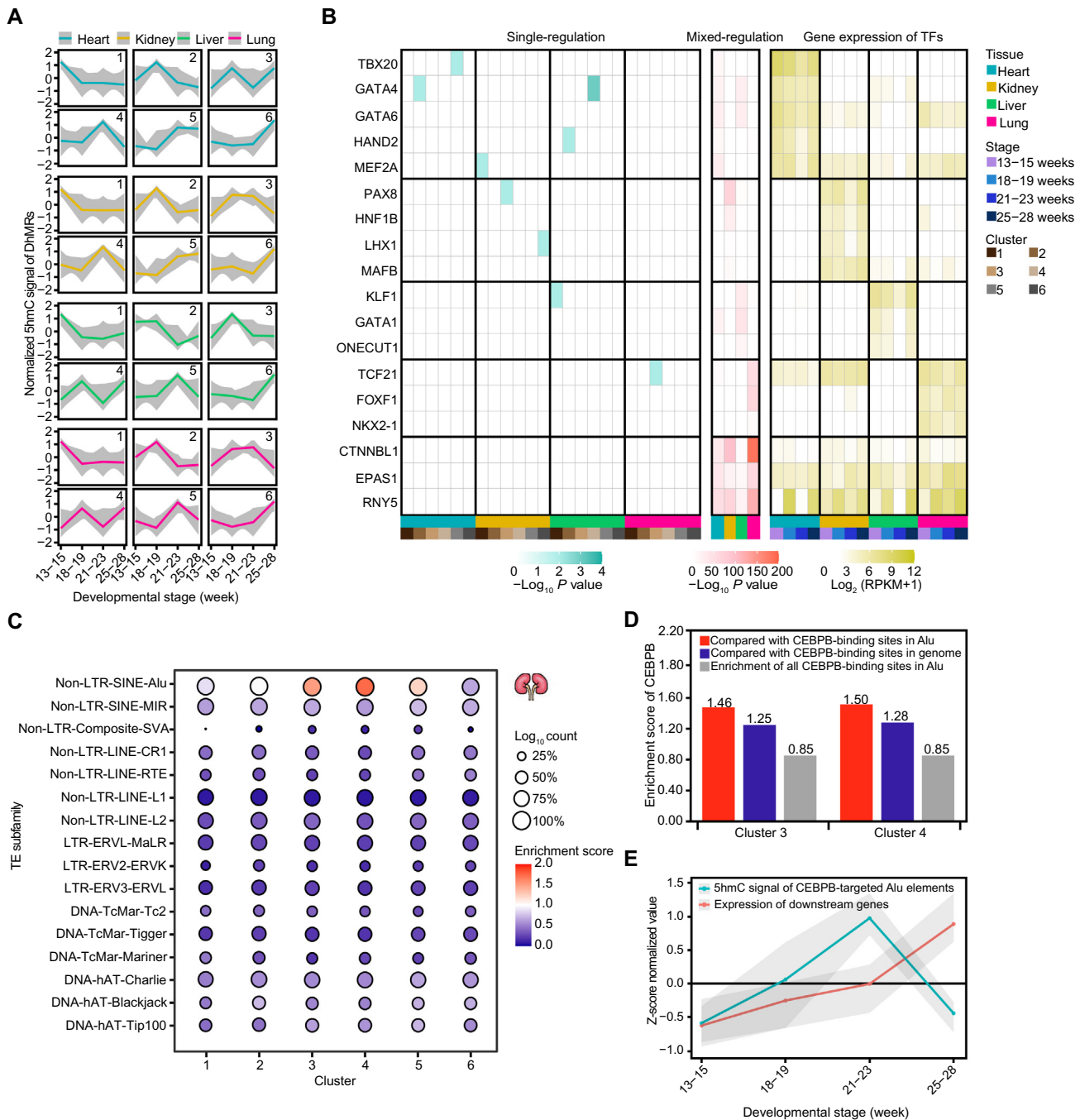


Figure 2 DNA 5hmC dynamically shapes foetal organ development by TFs as well as TEs

A. K-means clustering analysis of organ-specific DhMRs from heart, kidney, liver, and lung. DhMRs in each organ were separated into six clusters. **B.** Heatmap showing the TF-binding motifs identified by DhMRs. The TF-binding significance of single- and mixed-regulation gene groups and the expression level of each TF are shown from left to right. **C.** Enrichment of specific TE families in organ- and stage-specific DhMRs of foetal kidney. **D.** Bar plots showing the enrichment score of CEBPB-binding sites in Alu elements in Cluster 3 and Cluster 4 compared with their overall enrichment in all Alu elements (red) and their genome-wide distribution (blue). The overall enrichment of all CEBPB-binding sites in all Alu elements compared with their genome-wide distribution is also shown (gray). **E.** The z-score normalized 5hmC signals of CEBPB-targeted Alu elements as in (D) and the expression levels of downstream genes during human foetal kidney development. TF, transcription factor; TE, transposable element; CTNNB1, catenin beta like 1; CEBPB, CCAAT enhancer binding protein beta; EPAS1, endothelial PAS domain protein 1; FOXF1, forkhead box f1; GATA1, GATA binding protein 1; GATA4, GATA binding protein 4; GATA6, GATA binding protein 6; HAND2, heart and neural crest derivatives expressed 2; HNF1B, HNF1 homeobox B; KLF1, kruppel like factor 1; LHX1, LIM homeobox 1; MAFB, MAF BZIP transcription factor B; MEF2A, myocyte enhancer factor 2A; NKX2-1, NK2 homeobox 1; ONECUT1, one cut homeobox 1; PAX8, paired box 8; RNY5, RNA, Ro60-associated Y5; TBX20, T-box transcription factor 20; TCF21, transcription factor 21.

expression levels of those TFs in foetuses were obviously higher than those in adults (Figure S2G and H). Collectively, these results illustrate that TFs may participate in specific organ development with the assistance of sequential changes in 5hmC in foetuses.

DNA methylation is one of the major mechanisms adopted by the host to suppress the expression of transposable elements (TEs) [31], some of which, however, could be used by their host genome as regulatory elements in certain organs during development [32–34]. Therefore, we investigated whether 5hmC modifications are preferentially installed on certain families of TEs in an organ- and stage-specific manner, which potentially facilitates the switch of the repressive state to the accessible state of the chromatin for TF embedding and subsequent regulation of gene expression in *cis*. We found that members of the Alu family of short interspersed elements (SINEs) showed a significant 5hmC enrichment on the DhMRs within Clusters 3 and 4 in the foetal kidney (Figure 2C, Figure S2I–L). In particular, the binding sites of two TFs, namely CCAAT enhancer binding protein beta (CEBPB) and transcriptional repressor GATA binding 1 (TRPS1), were highly enriched in TSS-proximal Alu elements within these DhMRs (Figure 2D, Figure S2M), indicating that certain TFs preferentially bind to these Alu DhMRs, and their binding and *cis*-regulatory effects on target genes may be affected by differential 5hmC modifications of these Alu elements. The pattern of expression level changes of CEBPB-targeted genes was consistent with its role as an activator in transcriptional regulation (Figure 2E). Indeed, the expression levels of almost all CEBPB-targeted Alu-proximal genes (12/13) followed the 5hmC signals of the corresponding Alu DhMRs (Figure 2E). Interestingly, approximately half of these CEBPB-targeted Alu-proximal genes (7/13) have been previously reported to be tightly associated with kidney development or responses to injury (Table S5). In addition to CEBPB, six other TFs also showed primed elevation of 5hmC levels of Alu elements upstream of their target genes (Figure S2N). This result indicates that 5hmC can epigenetically mark a transition state of Alu elements, through which to affect the binding of TFs and further downstream gene expression.

DNA 5hmC is correlated with transcriptome homeostasis across human organs

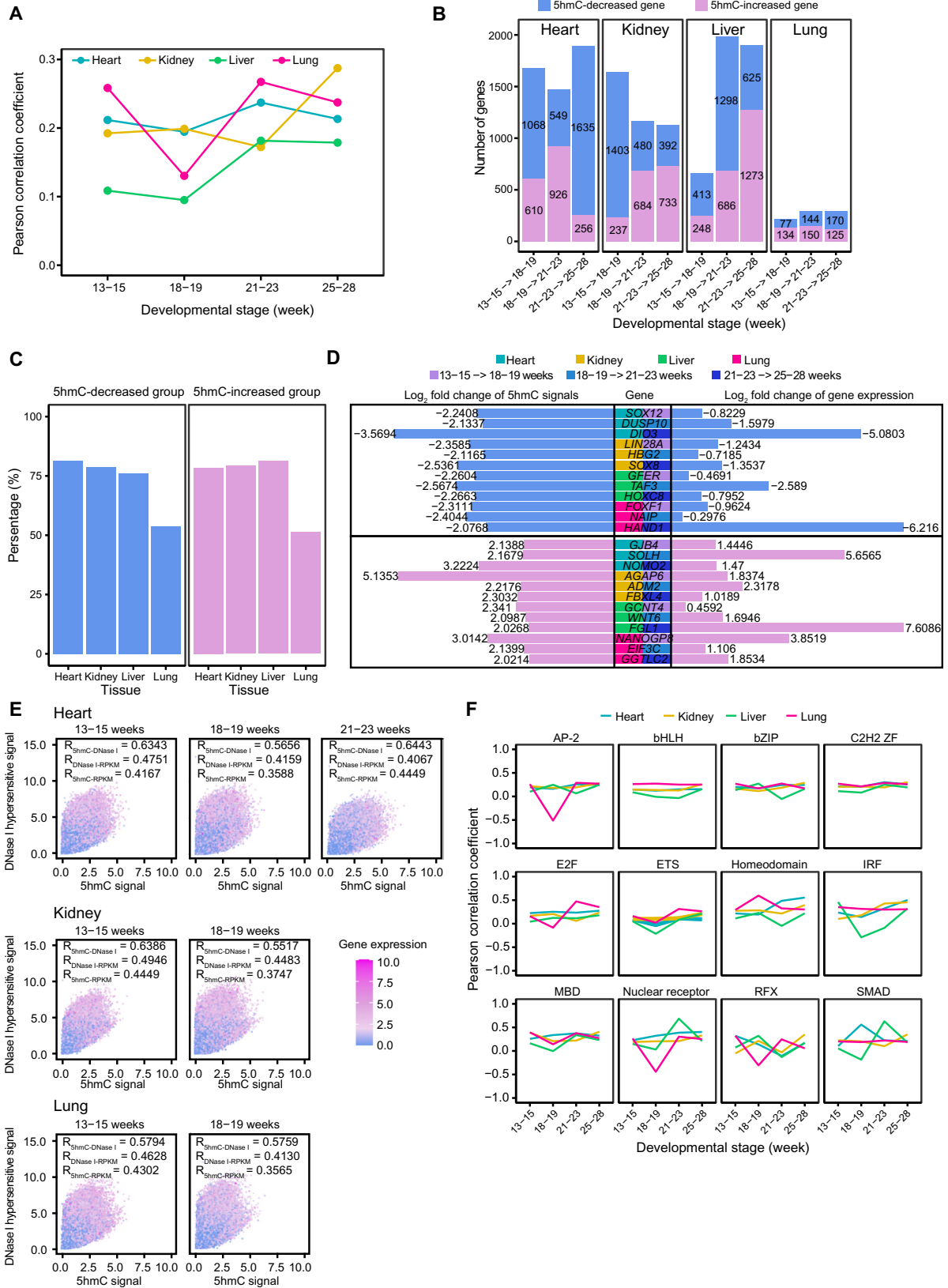
Most DhMRs were located around the TSSs in all foetal organs (Figure 1E) and the sequential changes of 5hmC signals on DhMRs showed a strong association with RNA abundance during organ development (Figure 2B and E). We then asked whether the dynamic reprogramming of DNA 5hmC on promoters regulates gene expression at developmental stages. In support, a positive correlation between transcriptional activity and the 5hmC signal on promoters was observed in foetal organ samples (Figure 3A), and most of the promoters contained 1–2 DhMRs (Figure S3A). Importantly, their association was not set in stone during the development of each organ (Figure 3A). Then we identified genes with significant 5hmC alterations in promoters across adjacent stages (changes $\geq 25\%$, fold change ≥ 2 , two-sided Student's *t*-test, $P < 1.03 \times 10^{-6}$) (Figure 3B, Figure S3B and C), and most of them were organ-specific genes with obvious 5hmC changes in promoters (Figure 3C). In particular, more than

half of the differentially expressed genes, such as those in the heart at 18–23 weeks (77.06%), liver at 18–23 weeks (74.09%), and liver at 21–28 weeks (74.70%) (Figure 3D, Figure S3D), were subjected to the 5hmC-decreased or -increased group, and genes which are highly related to development are shown in Figure 3D.

Additionally, we downloaded public DNase sequencing (DNase-Seq) data from the foetal heart, kidney, and lung at 13–23 weeks [35] and found that there was a high correlation among the 5hmC signals, chromatin openness on promoters, and gene expression levels, for which genes could be detected from all three omics datasets (Figure 3E). This finding suggests that 5hmC overlaps with open chromatin and is positively associated with transcriptional activity. However, we did not observe the aforementioned similar results between adjacent stages (Figure S3E). Then, we wondered whether TFs regulate the process across different organs through specific binding to 5hmC-enriched promoters. To discover corresponding 5hmC peaks based on known TF-binding motifs, we performed pattern matching to reveal their one-to-one potential correspondence. To increase the reliability and specificity of the binding sequences, we used 30% of the single base proportion to obtain the optimal combinatorial sequences (Figure S3F and G). Random peaks and genome fragments were used to eliminate the effects of background noise (two-sided paired Student's *t*-test, $P < 0.05$) (Figure S3H). According to the results, the number of 5hmC-modified target genes was different among distinct TF families. Cys2His2 zinc finger (C2H2 ZF), basic helix-loop-helix (bHLH), basic leucine zipper (bZIP), nuclear receptor, *etc.*, are the main TF families in eukaryotes [36], with more target genes than others in foetal organs (Figure S3I). We further calculated the correlation of the 5hmC signal and the gene expression level of target genes from different TF families (Figure 3F). The regulatory relationship of DNA 5hmC and transcription remained almost stable in C2H2 ZF, bHLH, bZIP, *etc.*, whereas there were probably different regulatory roles in specific TF families, such as adaptor protein 2 (AP-2), early 2 factor (E2F), and homeodomain (Figure 3F). All these findings indicate that 5hmC-modified promoters are highly associated with gene expression homeostasis during foetal organ development.

RNA m⁵C acts as a downstream regulator during foetal organogenesis

In the setting of organ-specific DNA 5hmC in the regulation of upstream gene expression, we wondered how RNA m⁵C drives organ-specific post-transcriptional regulatory networks, since it is one of the prevalent RNA post-transcriptional modifications associated with RNA stability [37,38]. RNA bisulfite sequencing (RNA-BisSeq) was performed on four foetal organs at three periods from 13 to 28 weeks. The samples had a dependable C-to-T conversion rate with good repeatability (Figure S4A and B; Table S6). Consistent with the findings in other vertebrates [18,21,37,38], the majority of m⁵C sites were enriched near the translation initiation sites (Figure 4A). A total of 2679–15,655 m⁵C sites within 316–1622 mRNAs were identified at each stage during organ development (Figure 4B). Eliminating the effect of continuous cytosines (Cs), most m⁵C sites were strongly enriched in CG-rich and coding sequence (CDS) regions (Figure S4C and D). Considering that



both the m⁵C methylation level and the number of m⁵C sites may affect the functions of corresponding mRNAs, we calculated the total methylation level across different organs according to the previous study [37] (Figure 4C, Figure S4E). Globally, m⁵C levels dynamically change during the development of the kidney, liver, and lung.

The global distributive features of m⁵C in human foetal organs prompted us to explore the potential role of m⁵C during development. Upon comparing RNA abundance changes between m⁵C-modified and -unmodified mRNAs across adjacent stages, we found that the m⁵C-modified mRNAs were unstable before 18–19 weeks, whereas they became stabilized afterwards ($P < 5.40 \times 10^{-7}$, Figure 4D). To eliminate the effect of RNA abundance on the evaluation of m⁵C sites, the correlation between methylation level and read coverage was calculated, which illustrated that the association between m⁵C and gene expression was independent of coverage changes (Figure S4F). Additionally, consistent with that of DNA 5hmC, organ-specifically m⁵C-modified mRNAs were significantly enriched in the pathways related to cell differentiation, proliferation, and adhesion, and participated in the regulation of gene expression and translation during embryogenesis (Figure S4G). Importantly, the organ- and stage-specific m⁵C-modified mRNAs were closely related to RNA abundance in each organ (Figure 4E; Table S7). These genes revealed a high level of enrichment for organ physiology-related processes. Thus, these results indicate that RNA m⁵C serves as a post-transcriptional regulator during human foetal organ development.

Individual and synergistic roles of DNA 5hmC and RNA m⁵C

We next proposed that DNA 5hmC and RNA m⁵C may cooperate to spatiotemporally modulate the regulatory networks in foetal organs. Overall, promoters of genes with high transcriptional activity presented relatively higher 5hmC signals than the m⁵C methylation levels on their corresponding mRNAs, suggesting a dominant role of DNA 5hmC at the transcriptomic level during development (Figure S5A). To further investigate the respective functions of DNA 5hmC and RNA m⁵C, we divided the organ-specifically modified genes into four groups: differentially expressed genes organ-specifically regulated by

both DNA 5hmC and RNA m⁵C (group A), differentially expressed genes organ-specifically regulated only by DNA 5hmC (group B), differentially expressed genes organ-specifically regulated only by RNA m⁵C (group C), and commonly expressed genes organ-specifically regulated by DNA 5hmC and RNA m⁵C (group D) (ANOVA, fold change > 1.5, $P < 0.01$; Figure 5A, Figure S5B and C; Table S8). Overall, 99.16% differentially expressed genes were organ-specifically regulated by both or single modifications, which suggested that the majority of organ-specifically expressed genes were modulated by DNA 5hmC and RNA m⁵C during human organ development (Figure S5B). As group D only accounts for 0.84% of the genes, these two types of modifications indeed affect the expression of the vast majority of organ-specific genes (Figure S5B). Specifically, in group A, there were close associations among DNA 5hmC signal intensity, RNA m⁵C methylation level, and gene expression level (Figure 5A, Figure S5C and D).

Importantly, the genes in all three groups (A, B and C) were involved in the processes of organ formation and functional maturity (Figure 5B; Table S9). In the foetal heart, the genes from all three groups were significantly enriched in heart contraction, cardiomyocyte differentiation, and heart rate-related cardiomyocytes. The co-regulatory genes were mainly enriched in angiogenesis. In contrast to DNA 5hmC mainly regulating cardiac myofibril assembly, related ion transportation, and cell communication through electrical coupling, m⁵C modification was involved in atrial action potential and action potential depolarization, especially for sino-atrial and atrioventricular node cells. Nevertheless, both ventricular action potential and action potential repolarization need the involvement of both modifications. Although the foetal heart performs basic functions earlier than other organs, its gradual functional maturity and maintenance of normal development also require these two modifications. Additionally, RNA m⁵C was even more critical in cell differentiation of the kidney, such as the regulation of mesenchymal epithelial cell transformation during metanephric morphogenesis, glomerular development, and renal vesicle morphogenesis. DNA 5hmC was the key driver of the organic acid metabolic process in the liver. In the lung, both DNA 5hmC and RNA m⁵C synergistically regulated respiratory gaseous exchange, and they separately modi-

Figure 3 DNA 5hmC participates in regulating gene expression across human foetal organs

A. Line chart showing the median Pearson's correlation coefficient of 5hmC signal and gene expression level at different stages on promoters. **B.** Graphs showing the number of significant 5hmC-decreased and -increased genes during foetal organ development. **C.** Bar plots showing the percentage of organ-specific genes with significant 5hmC changes in each organ. **D.** Bar plots showing the fold changes (\log_2) of 5hmC signals and gene expression levels of 18 organ-specific genes from 5hmC-decreased and -increased groups. **E.** Biplots comparing changes in 5hmC signals (x axis), changes in DNase I hypersensitive signals (y axis) with the RNA abundance changes (color) between the adjacent stages. Each dot represents a single gene. Public DNase-Seq data were from GSE18927. **F.** Dynamic of Pearson's correlation coefficients of different TF family groups from heart, kidney, liver, and lung. Pearson's correlation coefficients were calculated based on the 5hmC signals and the gene expression levels of target gene sets from corresponding TF families. AP-2, assembly polypeptide 2; bHLH, basic helix-loop-helix; bZIP, basic region/leucine zipper motif; C2H2 ZF, Cys2His2 zinc finger; E2F, early 2 factor; ETS, erythroblast transformation specific; IRF, interferon regulatory factor; MBD, methyl-CpG-binding domain; RFX, regulatory factor X; SMAD, SMAD family member; *ADM2*, adrenomedullin 2; *AGAP6*, ArfGAP with GTPase domain, ankyrin repeat and PH domain 6; *DIO3*, iodothyronine deiodinase 3; *DUSP10*, dual specificity phosphatase 10; *EIF3C*, eukaryotic translation initiation factor 3 subunit C; *FBXL4*, F-Box and leucine rich repeat protein 4; *FGL1*, fibrinogen like 1; *FOXF1*, forkhead box F1; *GCNT4*, glucosaminyl (N-acetyl) transferase 4; *GFER*, growth factor, augmentor of liver regeneration; *GGTLC2*, gamma-glutamyltransferase light chain 2; *GJB4*, gap junction protein beta 4; *HAND1*, heart and neural crest derivatives expressed 1; *HBG2*, hemoglobin subunit gamma 2; *HOXC8*, Homeobox C8; *LIN28A*, Lin-28 homolog A; *NANOGP8*, nanog homeobox retrogene P8; *NAIP*, NLR family apoptosis inhibitory protein; *NOMO2*, NODAL modulator 2; *SOLH*, Calpain 15; *SOX9*, SRY-box transcription factor 8; *SOX12*, SRY-box transcription factor 12; *TAF3*, TATA-box binding protein associated factor 3; *WNT6*, Wnt family member 6.

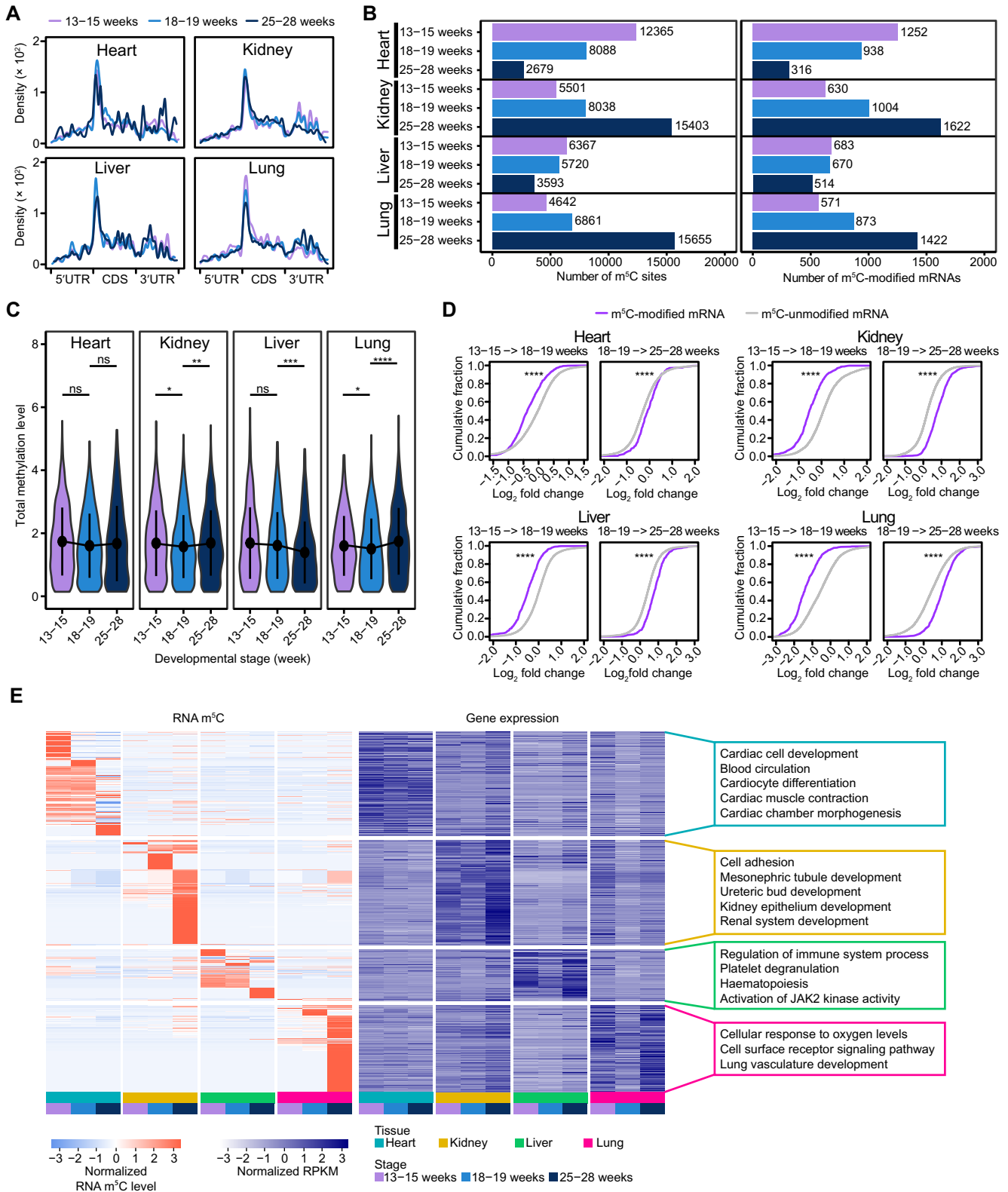


Figure 4 The RNA m⁵C profile during foetal organ development

A. The m⁵C distributions within different regions in different samples from heart, kidney, liver, and lung. **B.** Bar charts showing the numbers of m⁵C sites and m⁵C-modified mRNAs. **C.** Total RNA m⁵C methylation level across different developmental stages in heart, kidney, liver, and lung. The methylation level has been log₂ transformed. **D.** Cumulative distribution displaying the expression level change of m⁵C-modified and -unmodified mRNAs from RNA-Seq data comparing samples from adjacent stages. **E.** Heatmap showing the strong positive correlation between the m⁵C methylation levels of organ- and stage-specific genes and the expression levels of corresponding genes during foetal organ development. GO biological processes for each organ- and stage-specific gene set are listed on the right. The *P* values were calculated using two-sided Wilcoxon and Mann-Whitney tests. *, *P* < 0.05; **, *P* < 0.01, ***, *P* < 0.001, ****, *P* < 0.0001. CDS, coding sequence; GO, Gene Ontology.

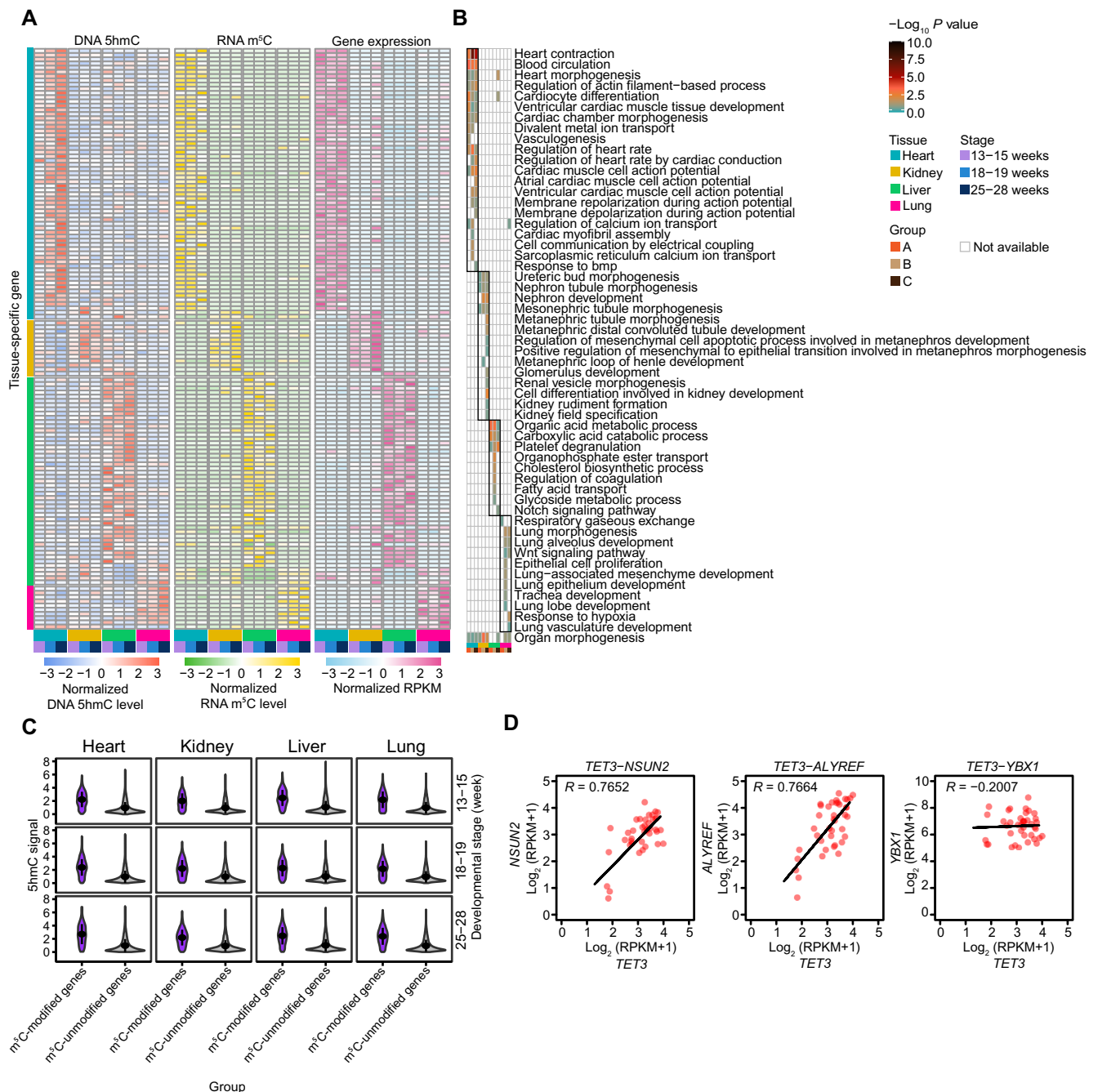


Figure 5 DNA 5hmC and RNA m⁵C consistently regulate foetal organ development

A. Heatmaps showing the dynamics of DNA 5hmC signals on promoters (left), m⁵C methylation levels (middle), and expression levels (right) of DNA 5hmC–RNA m⁵C co-regulated specifically-expressed gene (group A). DNA 5hmC signals, m⁵C methylation levels, and gene expression levels were normalized by z-score. **B.** GO biological processes of organ-specific genes in group A, group B, and group C. The color represents the significance of each biological process. **C.** The 5hmC signals on promoters of corresponding genes with or without m⁵C modification ($P < 2.20 \times 10^{-16}$). The P values were calculated using two-sided Wilcoxon and Mann–Whitney tests. **D.** Scatterplots showing the correlations between the expression levels of 5hmC methyltransferase *TET3* and m⁵C “writer” *NSUN2*, “reader” *ALYREF*, and *YBX1* in foetal organs. *ALYREF*, Aly/REF export factor; *NSUN2*, NOP2/Sun RNA methyltransferase 2; *TET3*, ten eleven translocation 3; *YBX1*, Y-box binding protein 1.

fied different gene sets to modulate lung morphogenesis (Table S9). These results suggest that in the foetal kidney and liver, DNA 5hmC and RNA m⁵C may exhibit differential and specific functions.

Interestingly, we found that 93.43%–99.60% of m⁵C-modified mRNAs were also modified by DNA 5hmC on the corresponding genes (Figure S5E). Moreover, 5hmC enrichment on promoters with m⁵C modification on their corresponding mRNAs was much higher than that without m⁵C

(Figure 5C), indicating a coordinated role of 5hmC and m⁵C in foetal organs. TET3 was reported to be the only existed DNA 5mC oxidase after fertilization, mediating the oxidation of most paternal nuclear Cs and a large-scale removal of 5mC [14,39]. To determine the potential mechanisms underlying for the cross-talk between the two types of modifications, we calculated the correlation coefficients between the expression of *TET3* and RNA m⁵C regulators. Apparently, the correlation coefficients of the expression levels of *TET3* and NOP2/Sun RNA methyltransferase 2 (*NSUN2*) or Aly/REF export factor (*ALYREF*) were high ($R = 0.7652$ for *NSUN2* and $R = 0.7664$ for *ALYREF*) (Figure 5D). However, no significant relationship was found between *TET3* and Y-box binding protein 1 (*YBX1*). These results suggest that DNA 5hmC and RNA m⁵C might drive organ-specific regulatory networks through pre- and post-transcriptional mechanisms, respectively, which are mediated by *TET3* and *NSUN2/ALYREF*. However, how they mechanistically contribute to the developmental processes warrants further detailed investigation.

Discussion

The importance of epigenetic modifications in embryogenesis is well known, but the global functions and essential links between DNA 5hmC and RNA m⁵C at pre- and post-transcriptional levels during foetal organ development are currently poorly understood. Our study provides the first overview of DNA 5hmC and RNA m⁵C during human foetal development, and reveals their potential cooperation as key molecular events in the development of foetal organs, including heart, kidney, liver, and lung. We found that the key TFs, *TBX20*, *PAX8*, *KLF1*, *TCF21*, and *CEBPB*, specifically contribute to the formation of distinct organs at different stages. Additionally, 5hmC-enriched Alu elements may participate in the regulation of expression of TF-targeted genes. Both DNA 5hmC and RNA m⁵C dynamically modulate the changes in RNA abundance and are highly associated with the developmental functions of the corresponding organs. The high correlation between *TET3* and *NSUN2/ALYREF* suggests the potential direct/indirect cross-talk between DNA 5hmC and RNA m⁵C. Collectively, the modifications in the layers of DNA and RNA perform their functions and may work cooperatively during foetal organ development (Figure S6).

A previous study has shown that the correlation coefficient between 5hmC and gene expression 2 kb upstream of the TSS is approximately 0.21 in the adult liver and lung [13], suggesting its positive regulatory roles in human organs. In fetuses, we found that the relationship between the 5hmC signal on promoters and transcriptional activity was highly dynamic and that the 5hmC modification localized on open chromatin was strongly associated with the gene expression level (Figure 3A and E). Chromatin accessibility of *cis*-regulatory sequences drives cell differentiation and zygotic genome activation (ZGA) by regulating gene expression in early mammalian embryos [6,40,41]. Studies have shown that post-translational modifications, such as histone acetylation and ubiquitination, enhance chromatin accessibility by disrupting chromatin folding and globule formation [42,43]. Additionally, as a 5hmC binding protein, methyl-CpG binding protein 2 (MeCP2) competes with H1 for nucleosomal binding sites leading to chromatin higher-order structure changes [43,44]. According to the

existing theoretical basis, we reasonably speculate that chromatin accessibility could be modulated by specific binding of 5hmC-related regulators in a histone modification-dependent manner, resulting in high transcriptional activity.

TEs provide evolutionary opportunities to generate functional *de novo* elements during speciation, and indeed the majority of primate-specific regulatory sequences are derived from TEs [45,46]. Although it has been found that organ-specific TE subfamilies function as enhancers in human adults [47], we still don't know whether and how human-specific TEs serve as regulatory elements to control gene expression for distinct biological functions and organ morphology in early human embryonic development. Here we observed correlation of gene expression with organ- and stage-specific 5hmC modification of Alu elements harboring specific TF-binding sites proximal to target genes during foetal kidney development. It indicates that 5hmC marks the change of chromatin accessibility of these TE elements that have been rewired for new functions by the host genome [34]. Mis-regulation of these target genes leads to diabetic kidney disease [48], irregular kidney morphology [49], renal fibrosis [50], etc. *CEBPB* itself is also involved in the regulation of renal development, as the prenatal caffeine exposure (PCE), which causes abnormalities in adrenal structure and function development in female foetal rats, decreases *CEBPB*'s expression and increases *CEBPA*'s expression in the so-called glucocorticoid activation system (11 β -HSD/GR/C/EBP) [51].

RNA m⁵C modification has been recently demonstrated to be an important post-transcriptional regulator of mRNA stability during maternal-to-zygotic transition (MZT) in zebrafish and the pathogenesis of bladder cancer in humans [37,38]. In human fetuses, the involvement of RNA m⁵C was inferred to be divided into two periods with 18–19 weeks as the boundary. Compared with those in the former stages, m⁵C-modified mRNAs were more stable than unmodified mRNAs after 18–19 weeks in all the four organs. Generally, the fate of m⁵C-modified transcripts is determined by their regulators: “writers”, “readers”, and “erasers” [17]. In some cases, RNA methylation or its regulators play dual roles in cancer progression by affecting the stability of different target genes [20], such as the m⁶A “writer” protein methyltransferase like 3 (*METTL3*) in colorectal cancer (CRC). We hypothesize that the differential roles of RNA m⁵C may result from the abundance/binding affinity changes of m⁵C regulators on different target genes to maintain normal development. Another possible reason is that the dynamic expression of m⁵C regulators, such as *NSUN2* (writer), *ALYREF* (reader) and *YBX1* (reader), may lead to differential regulation events of RNA m⁵C during organogenesis, which warrants further detailed investigation.

We speculate that DNA 5hmC and RNA m⁵C might drive organ-specific regulatory networks synergistically through the following three possible mechanisms: (1) TET family members participate in the regulation of both DNA 5hmC and RNA m⁵C methylation. DNA 5hmC is produced through 5mC oxidized by TET family members (*TET1*, *TET2*, and *TET3*) [52]. In HEK293T cells, *TET2* mediates m⁵C oxidation in tRNA promoting translation *in vitro* [53]. Thus, it is logical to speculate that TET may modulate both the DNA and RNA methylation spontaneously. However, there is still a lack of evidence that whether TET could mediate m⁵C oxidation on mRNA in mammalian. Therefore, this hypothesis remains to be further validated. (2) The interaction of methyltransferase and/or

binding proteins contributes to the cross-talk of DNA 5hmC and RNA m⁵C [54–58]. It is worth noting that histone H3 trimethylation at Lys9 (H3K9me3) has been reported to directly reinforce DNA methylation maintenance via the ubiquitin-like with PHD and ring finger domains 1 (UHRF1)–H3K9me3 axis [54]. Moreover, RNA m⁶A also directs the demethylation of histone H3 trimethylation at Lys9 (H3K9me2) modification which is targeted by lysine demethylase 3B (KDM3B) and thus promoting overall gene expression [55]. A previous study has demonstrated that histone H3 trimethylation at Lys36 (H3K36me3) and H3 trimethylation at Lys27 (H3K27me3) could guide or impede RNA m⁶A deposition in mice and humans [56–58]. The high correlation coefficients of the expression levels of *TET3* and RNA m⁵C regulators *NSUN2* or *ALYREF* (Figure 5D) prompted us to propose another possible mechanism that TET family members not only mediate 5hmC formation, but also might cooperate with *NSUN2/ALYREF* directly or indirectly, inducing mRNAs to obtain m⁵C modification. However, digging the direct evidence of the interactions between those regulators currently still under way. (3) The indirect association exists by an upstream mechanism. The essential links between DNA methylation and RNA m⁶A have been reported to influence the fruit ripening in tomato, where *SIALKBH2* is negatively regulated by DNA 5mC thereby modulating the mRNA m⁶A level, and m⁶A facilitates *SIDML2* mRNA decay in turn affecting DNA methylation to promote fruit ripening [59]. During foetal organogenesis, we found that 5hmC enrichment on promoters with m⁵C modification on their corresponding mRNAs was much higher than that without m⁵C (Figure 5C). It is possible that a similar feedback regulation manner might exist for DNA 5hmC and RNA m⁵C in human fetuses. Even though there are some possible synergistic regulatory ways mentioned above, the exact molecular mechanisms need to be further studied.

Breakthroughs in single-cell instruments and techniques bring off multi-omics research at high-resolution levels, interrogating cellular heterogeneity and achieving refined tissue classification [60,61]. Advances in single-cell epigenomic sequencing, such as single-cell assay for transposase-accessible chromatin sequencing (scATAC-Seq), single-cell chromatin immunocleavage sequencing (scChIC-Seq), single-cell reduced representation bisulfite sequencing (scRRBS), and single-cell Hi-C (scHi-C), make it possible to concisely and comprehensively explore DNA accessibility, histone modifications, DNA methylation, and chromatin interactions in developmental and disease models [62]. Despite the dramatic advances, single-cell sequencing still needs to be improved due to some limitations and challenges: (1) technological noise during library building (e.g., the methods of digesting tissue); (2) lowly expressed genes cannot be detected; (3) numerous downstream computational methods with vary performances already exist whereas lack unified standards and may introduce artificial biases; and (4) the results of sequencing only a small fraction of cells may not obtain the characteristics of bulk tissues [63]. Thus, to explore the overall regulation roles of these two types of modifications during human embryogenesis, we used bulk tissue samplings for obtaining the epigenetic profiling with aggregate or average signals, which contributes to further enable a deeper understanding of the embryonic development process. In the future, what deserves expecting is that more optimized single-cell sequencing methods will

indeed unravel the cell-to-cell differences and epigenetic networks of different tissues during human foetal development at a more precise level.

A previous study has shown that aberrant epigenomes are highly associated with multiple diseases, especially cancers [64]. TET-mediated DNA 5hmC was reported to act as an important regulator in kidney cancer, breast cancer, colon cancer, prostate cancer, and so on [65]. In most cases, a global reduction in 5hmC was detected in cancer samples compared to the normal one. TET mutation is one of the main reasons for pathogenesis. Its deregulation resulting from aberrant transcription, protein instability, and inappropriate localization leads to abnormal 5hmC enrichment. For example, the altered levels of TET2 promoting 5hmC variation are likely to be relevant in the Sjögren's syndrome etiopathogenesis [66]. Additionally, as another prevalent modification, RNA m⁵C also involves in regulating cancer progression through reshaping the transcriptome which is dependent on *NSUN2/YBX1* to affect RNA stability, including breast cancer, bladder cancer, and gastric cancer [20]. Thus, aberrant distributions of DNA 5hmC and/or RNA m⁵C highly possibly impede normal foetal organogenesis. In this work, we profile the first landscapes of DNA 5hmC and RNA m⁵C during human foetal development, which provide an important data resource and research basis for in-depth excavation of the mechanisms of early foetal congenital diseases.

Conclusion

Taken together, this study not only uncovered the dynamic DNA 5hmC and RNA m⁵C reprogramming during organogenesis in post-implanted embryos, but also revealed a putative cross-talk between DNA modification and RNA methylation, which adds another layer by which the transcriptome can be spatiotemporally co-regulated to form a well-coordinated network at specific developmental stages ensuring normal foetal development. Furthermore, a comprehensive description of the characteristics and functional importance of modifications at both the DNA and RNA levels provides available epigenome datasets of human foetuses and a foundation for understanding the in-depth epigenetic mechanisms for early human development.

Materials and methods

Tissue collection

Foetal heart, liver, kidney, and lung were obtained from foetal tissues after selective induction of labor (gestational age range, 13–28 weeks). Foetal tissues with pathological pregnancies, abnormal karyotype, or developmental defects were excluded from this study. All the normal organs were isolated after a mechanical dissection, serially washed with medical saline for three times, and immediately frozen in liquid nitrogen.

Genomic DNA and total RNA preparation

A maximum amount of 20 mg of tissue ground in a 100:1 mixture of 600 μ l Buffer RLT Plus and 6 μ l 14.3 M β -mercaptoethanol for genomic DNA and total RNA extraction

by user operation manual of AllPrep DNA/RNA/miRNA Universal Kit (Catalog No. 80224, Qiagen, Hilden, Germany). For rRNA depletion, the purified total RNA was further treated by a 3:1 mixture of probe and total RNA followed by RNase H (Catalog No. EN0202, ThermoFisher Scientific, Waltham, MA) treatment at 37 °C. TURBO DNase (Catalog No. AM2238, ThermoFisher Scientific) treatment was used to eliminate DNA contamination. Then, the rRNA-depleted RNA was purified by ethanol precipitation and used for construction of the RNA-Seq and RNA-BisSeq libraries. The concentration and quality of purified genomic DNA and rRNA-depleted RNA were detected by NanoDrop One UV-Vis Spectrophotometer (ThermoFisher Scientific) and Qubit 4 Fluorometer (ThermoFisher Scientific).

RNA-Seq library generation and sequencing

The RNA-Seq libraries of human foetal organs were constructed by using user manual of KAPA RNA HyperPrep Kit (Catalog No. KK8541, KAPA, Cape Town, ZA). In brief, the high-quality, rRNA-depleted RNA samples of foetal organs were fragmented for 6 min at 85 °C (about 300 nt) by 2× Fragment, Prime and Elute Buffer in different periods. The DNA concentration and quality of constructed libraries were detected by Qubit 4 Fluorometer and Agilent 2100 Bioanalyzer before sequencing.

Bisulfite conversion of RNA

RNA fragmentation and bisulfite conversion were consistent with previous description [21], but had some modifications. Briefly, *in vitro* transcribed mouse Dhfr mRNA, as a methylation conversion reference, was mixed with purified rRNA-depleted RNA at a ratio of 1:200. The mixture was fragmented into ~ 200 nt by 10× RNA Fragmentation Reagent (Catalog No. AM8740, ThermoFisher Scientific) for 22 s at 88 °C which was terminated by 10× RNA stop solution. Then fragmented RNA was obtained by ethanol precipitation. After washing precipitation with 80% ethanol, the RNA precipitates were dissolved in 100 µl bisulfite solution that mixed 40% sodium bisulfite (Catalog No. 243973, Sigma, Saint Louis, MO) and 600 µM hydroquinone (Catalog No. H9003, Sigma) in a ratio of 100:1. The synthesis was incubated at 75 °C for 4 h. Nanosep columns with 3K Omega membranes (Catalog No. OD003C35, Pall Corporation, Fajardo, Puerto Rico) were used to desalt the reaction mixture with centrifugation. The RNA pellet was washed with 1 M Tris-HCl (pH 9.0) followed by centrifugation for five times. Finally, the RNA was re-suspended in 75 µl of nuclease-free water and incubated at 75 °C for 1 h with equal volume of 1 M Tris-HCl (pH 9.0) for desulfonation. After ethanol precipitation, the bisulfite-converted RNA was dissolved in RNase-free water and prepared for library construction of RNA-BisSeq. After reverse transcription with SuperScript II Reverse Transcriptase (Catalog No. 18064014, ThermoFisher Scientific) and ACT random hexamers to form cDNA, KAPA RNA HyperPrep Kit (Catalog No. KK8541, KAPA) was used to perform the subsequent procedures as described in the manufacturer's instructions. Paired-end sequencing was performed on the Illumina HiSeq2500 instrument with 125 bp read length.

Nano-hmC-Seal in human tissues

Genomic DNA was prepared from the aforementioned steps of genomic DNA and total RNA preparation. The input DNA ranged from 5 ng to 50 ng. Nano-hmC-Seal library construction was similar to the description of previous study [22]. The reactions were performed using TruePrep DNA Library Prep Kit V2 for Illumina (Catalog No. TD501, Vazyme, Nanjing, China) according to the manufacture's instruction with some modifications. In brief, the genomic DNA was fragmented in 50 µl mixture containing 10 µl 5× TTBL, 50 ng gDNA, and 5 µl TTE mix V50 at 55 °C for 10 min. The DNA fragments were purified by 1× AMPure XP beads (Catalog No. A63882, Beckman Coulter, Indianapolis, IN) and then eluted by RNase-free water. The reaction of glucosylation occurred in synthesis of 50 mM HEPES buffer (pH 8.0), 25 mM MgCl₂, fragmented DNA, 100 µM N₃-UDP-Glc, and 1 µM T4 beta-glucosyltransferase (βGT; Catalog No. EO0831, ThermoFisher Scientific) at 37 °C for 1 h. After glucosylation reaction, the reaction solution mixed with 2 µl DBCO-PEG4-DBCO (20 mM stock in DMSO) and incubated at 37 °C for 2 h. The post-reaction solution was desalted by Micro Bio-Spin P-30 Gel Columns (Catalog No. 7326226, Bio-Rad, Hercules, CA) to obtain the purified DNA. 5 µl Dynabeads MyOne Streptavidin C1 (Catalog No. 65002, ThermoFisher Scientific) in 2× buffer [1× buffer: 5 mM Tris (pH 7.5), 0.5 mM EDTA, and 1 M NaCl] was incubated with purified DNA at room temperature for 15 min with gentle rotation and washed six times with 1× buffer by user operation manual. The PCR amplification with magnetic beads purified by 1× AMPure XP beads and prepared for sequencing on HiSeq instrument.

High-throughput sequencing data pre-processing and analysis

Nano-hmC-Seal, RNA-Seq, and RNA-BisSeq were carried out on Illumina HiSeq 2500 platform with paired-end 125 bp read length. Trimmomatic (version 0.33) [67] was used to trim off adaptor sequences, and reads < 35 nt in length were filtered out. To eliminate gender effects, all the reads aligned to chromosome X or Y were removed. (1) Nano-hmC-Seal analysis. Clean reads were mapped to the human genome (hg19) using Bowtie 2 (version 2.2.9) [68]. Reads with quality score ≥ 20 were retained for the subsequent analysis. MACS2 (version 2.1.1) [69,70] was used for the peak calling with the parameters: -c -f BAM --nomodel -gsize = hs --keep-dup all -n -B -p 0.05. Peaks from replicates were merged using mergePeaks (HOMER, version 4.9.1) [71]. analyzeRepeats.pl (HOMER, version 4.9.1) [71] was performed to do peak annotation and calculate the raw read counts (RPM) of gene bodies and promoters [defined as the regions 500 bp from the TSSs]. The average RPM of each gene from replicates was calculated for downstream analysis. The average profile of 5hmC in each sample was visualized by ngs.plot (version 2.61) [72] with parameters: -G hg19 -R gene body -L 3000. (2) RNA-Seq analysis. Clean reads were aligned to the human genome (hg19) using HISAT2 (version 2.0.5) [73] with default parameters. The number of uniquely mapped reads (quality score ≥ 20) mapped to each gene was counted using the featureCounts (version 1.6.2) [74] with parameters: -p -t exon -g gene_id -s 2.

Reads Per Kilobase per Million mapped reads (RPKM) was computed as the number of reads which map per kilobase of exon model per million mapped reads for each gene. (3) RNA-BisSeq analysis. Clean reads were aligned to the human genome (hg19) by meRanT align (meRanTK, version 1.2.0) [75] with parameters: -fmo -mmr 0.01. Under a high-stringency condition, the equal conversion rate with internal reference RNA sequence confirms that the unmethylated Cs in the transcripts are fully converted. Considering that the Bis-Seq data with a conversion rate greater than 99% is normally used for further analysis [21,76], those samples fitted the criteria were used to perform m⁵C-calling by meRanCall (meRanTK, version 1.2.0) [75] with parameters: -mBQ 20 -mr 0 -cr 0.99 -fdr 0.05 (Table S6). The methylation level was calculated according to the following formula: $i/(i + j)$. “i” represents the number of reads showing methylation at each site. “j” represents the number of reads without methylation. Only sites with coverage depth ≥ 30 , methylated cytosine depth ≥ 5 , and methylation level ≥ 0.1 were used to do further analysis. The credible m⁵C sites were annotated by BEDTools’ intersectBed (version 2.26.0) [77]. The overlap ratios of m⁵C sites and mRNAs between biological replicates at each stage were over 91.94% and 87.92%, respectively, and the Pearson correlation coefficients of methylation level were over 0.88. According to our previous method [32], the m⁵C distribution among CDSs, 5′-untranslated regions (5′UTRs), 3′UTRs, intron, CG, CHG, and CHH (H = A, C, U) were calculated. The sequences 10 nt up- and down-stream of m⁵C sites were used to detect the sequence preference, and logo plots were generated with WebLogo (version 3) [78]. To compare the overall changes of RNA m⁵C among different stages in four organs, total methylation level of each mRNA (the cumulative value of the methylation level of all the m⁵C sites per gene) was calculated according to the previous study [37]. And (4) the mapped reads of public DNase-Seq data from foetal heart, kidney, and lung were downloaded and used to do peak calling by DFilter (version 1.6) [79] with the parameters: -ks = 50 -l pval = 2 -f = bam. Public nano-hmC-Seal and RNA-Seq data from adult organs in this study were downloaded from GSE144530. Public DNase-Seq data from foetal organs at 13–23 weeks were downloaded from GSE18927.

DhMR identification

A non-overlapping 100 bp window was applied to perform DhMR identification across the samples from different organs. Briefly, the genome was first binned into consecutive 100 bp windows, and the signal intensity of each window was calculated based on the normalized RPM. Then the average 5hmC signal of replicates was calculated. ANOVA was used to compute the magnitude of the difference for each bin in R (version 3.4.1) (<https://www.r-project.org/>). Differentially 5hmC-modified bins were distinguished from each specific organ and other three organs. The *P* values were calculated using Benjamini-Hochberg method. The bins with fold change > 1.5 and $P < 0.01$ were considered statistically significant. The 5% least significant bins were considered as common bins (control). The significantly differential bins and common bins were merged respectively according to their positions on the chromosome and were considered as DhMRs and ChMRs, respectively. annotatePeaks.pl (HOMER, version

4.9.1) [71] was used to do DhMR and ChMR annotation. Principal component analysis (PCA) was performed by factoextra (R package). The distance between TSSs and DhMRs was calculated by BEDTools bedtools closest (version 2.26.0) [77]. GREAT (version 4.0.4) [80] was used to do annotation for DhMRs. There were 376,775 (average number of DhMRs of the four organs) genome fragments randomly extracted in length of the average length of DhMRs (159 nt) for 20 times, which served as the background when calculated the distance between DhMRs and TSSs.

Identification of clusters with temporal changes

The clusters with sequential changes were obtained by K-means with the method Pearson correlation (MEV, version 4.9.0) (<https://mev.tm4.org>). When the number of clusters was greater than 6, similar 5hmC trends appeared. Thus DhMRs were classified to 2–9 clusters, and according to the changes of 5hmC signals of all DhMRs, it contains almost all types of nonrepetitive trends for 6 clusters. Therefore, the DhMRs were divided into 6 clusters finally. HOMER (version 4.9.1) [71] was used to performed TF prediction in DhMRs of each clusters with the following command: findMotifsGenome.pl input.bed hg19 -size given -len 6 -novevopp -cache 1000. Genes with TF-potentially-binding DhMRs only from one cluster are considered as single-regulation groups, whereas genes with TF-potentially-binding DhMRs from two or more clusters are considered as mixed-regulation groups.

Differentially hypomethylated TE analysis

Each of the six DhMR clusters of the four organs was intersected with RepeatMasker (version 4.0.5) (<https://repeat-masker.org>) using BEDTools’ intersectBed (version 2.26.0) [77]. The enrichment scores were calculated by the ratio of the density of a TE family in a cluster divided by the density of this TE family in total genome. Since the enrichment score may be biased toward small TE families although their intersections with DhMRs may be rare, we plotted the enrichment scores with the number of intersections. Indeed, with the increase of the intersection counts, the enrichment score first drops then gradually goes up (Figure S21). To avoid the biases of the size of TE families, we set up an arbitrary cutoff of 200 for the counts of intersections. For the identification of TFs whose binding sites are enriched in organ- and stage-specifically 5hmC-modified Alu elements, the intersections of Cluster 3 and Cluster 4 DhMRs of foetal kidney with RepeatMasker were first filtered for Alu elements, then intersected with Gene Transcription Regulatory Database (GTRD, version 19.04) [81] human meta cluster interval file, keeping only the intersections that cover over 50% of the Alu elements. The results were further filtered to remove binding sites of TFs that were scarcely expressed at all four stages in foetal kidney (RPKM < 0.1). In order to eliminate ambiguity in the correlation between TF-binding sites, differential 5hmC modifications, and downstream gene expression, Alu elements that were bound by only one TF were kept for further analysis. The resulting Alu elements were then filtered by their proximity to downstream genes, keeping those within 1000 bp upstream of a gene with proper orientation and observable expression

(RPKM > 0.1) in at least one of the four stages. To obtain the enrichment scores of binding sites for CEBPB and TRPS1 with these Alu DhMRs, we first calculated the density of the intersections of either TF-binding regions with these Alu DhMRs. This density was then divided by either the density of the intersections of these TF-binding sites with all Alu elements in the human genome, or by the density of these TF-binding sites in the whole genome. The enrichment scores of global TF-binding sites in all Alu elements were calculated by the ratio of the density of the intersections of corresponding TF-binding sites with all Alu elements in genome divided by the density of corresponding TF-binding sites in the whole genome.

Dynamic changes of DNA 5hmC between adjacent stages

The promoters with the fold change of 5hmC signals between samples from two adjacent stages greater than or equal to 25% were defined as significant different promoters (two-sided Student's *t*-test with $P < 0.05$). Thus, the promoters were separated into two groups: promoters with decreased 5hmC signals and promoters with increased 5hmC signals. To explore the relationship between 5hmC signals in promoters and gene expression, the gene expression levels of corresponding groups were compared and the P values were calculated using two-sided Wilcoxon and Mann–Whitney test. The genes with fold change > 1.2 were defined as differentially expressed genes in 5hmC-decreased and -increased groups.

Corresponding 5hmC peak discovery based on known TF-binding motifs

To further explore how different TFs regulate gene expression via binding to 5hmC-enriched promoters, TF-bound 5hmC peaks were identified based on the known motifs of TFs by pattern matching. The motifs of TFs were downloaded from HumanTFs database [36] (<https://humantfs.ccb.utoronto.ca/>) which contains ~ 1600 TFs (~ 5000 possible motifs). Different TFs could bind to the sequences with the same motif and one TF could recognize the sequences with different motifs. In addition, there are different kinds of combinatorial sequences for each motif. Thus, we exhausted all the possible combinatorial sequences of 10 nt motifs for each TF. To increase the reliability of the sequences, 0.1–0.9 gradients of threshold for single base proportion were set to obtain the optimal combinatorial sequences. The single base proportion represents the probability of a specific base (A, T, C, and G) at a specific position. 5hmC peaks on promoters were subjected to cyclic matching to those combinatorial sequences of motifs. Four kinds of backgrounds were used to eliminate the effects of background noise: (1) random 5hmC peaks; (2) random 5hmC peaks in distal region; (3) random genome fragments in the same length of the average length of 5hmC peaks; (4) random genome fragments in distal region in the same length of the average length of 5hmC peaks. The fragments were randomly selected for 20 times per background. The motif sequence remained if the ratio of the number of actual peaks divided by the number of fragments in the background was > 2 for at least one random fragment group. The P values were calculated using two-sided paired Student's *t*-test. The remaining motifs that could be detected based on all the four kinds of backgrounds were considered credible.

Dynamic changes of RNA m⁵C

All expressed genes were divided into two groups based on with or without m⁵C modification at adjacent time points. To further evaluate the effects of m⁵C dynamic changes on mRNA abundance change, mRNAs with m⁵C modification at last stage but without m⁵C modification at next stage were defined as mRNAs with m⁵C loss; mRNAs without m⁵C modification at last stage but with m⁵C modification at next stage were defined as mRNAs with m⁵C gain. The mRNA abundance changes of these two groups were compared. The P values were calculated using two-sided Wilcoxon and Mann–Whitney test. The organ-specifically m⁵C-modified genes were identified according to the DhMR methods.

Association analysis with different organ-specific groups

To analyze the relationship among DNA 5hmC, RNA m⁵C, and transcriptome, the 5hmC signal of promoter, m⁵C methylation level, and gene expression level (RPKM) were subjected to log₂ transformation and deviation standardization $[(x - \min) / (\max - \min)]$ (x , input specific value; min, the minimum value of one specific omics data; max, the maximum value of one specific omics data). The organ-specifically expressed genes were identified using ANOVA in R (version 3.4.1). To further explore how epigenetic marks differentially regulate foetal organ development, organ-specific genes were separated into four groups: differentially expressed genes that were specifically regulated by both DNA 5hmC and RNA m⁵C, differentially expressed genes that were only specifically regulated by DNA 5hmC, differentially expressed genes that were only specifically regulated by RNA m⁵C, and commonly expressed genes that were specifically regulated by DNA 5hmC and RNA m⁵C.

Gene Ontology analysis and visualization

DAVID (version 6.8) [82] was used to perform Gene Ontology (GO) analysis. GO terms with $P < 0.05$ were considered as statistically significant. IGVTools (version 2.3.8) [83] was used for visualization.

Statistical analysis

ANOVA was used to compute the magnitude of the difference for each 5hmC-modified bin, the organ-specifically m⁵C-modified genes, and the organ-specifically expressed genes, respectively. The P values were calculated using Benjamini–Hochberg method for this part. The significance of dynamic changes of DNA 5hmC between adjacent stages and corresponding 5hmC peak discovery were calculated using two-sided paired Student's *t*-test. The P values of the DNA 5hmC signal, RNA m⁵C level, and mRNA abundance changes were calculated using two-sided Wilcoxon and Mann–Whitney test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

Ethical statement

All human foetal organs at different stages of development were obtained voluntarily in the First Affiliated Hospital of

Zhengzhou University, China with informed consent signed and approved by the Ethics Committee (License No. 2020-KY-261).

Data availability

The raw and processed nano-hmC-Seal, RNA-BisSeq, and RNA-Seq data reported in this study have been deposited in the Genome Sequence Archive for Human [84] at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformatics (GSA-Human: HRA000705 with BioProject: PRJCA004624), which are publicly accessible at <https://ngdc.cncb.ac.cn/gsa-human/>.

Competing interests

The authors have declared no competing interests.

CRedit authorship contribution statement

Xiao Han: Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Jia Guo:** Investigation, Resources, Writing – original draft. **Mengke Wang:** Formal analysis, Writing – original draft, Visualization. **Nan Zhang:** Investigation, Resources. **Jie Ren:** Visualization, Writing – original draft. **Ying Yang:** Visualization. **Xu Chi:** Software, Formal analysis, Data curation. **Yusheng Chen:** Visualization. **Huan Yao:** Resources. **Yong-Liang Zhao:** Writing – original draft. **Yun-Gui Yang:** Conceptualization, Supervision, Funding acquisition. **Yingpu Sun:** Conceptualization, Supervision, Funding acquisition. **Jiawei Xu:** Conceptualization, Supervision, Project administration, Funding acquisition. All authors have read and approved the final manuscript.

Acknowledgments

This work was supported by the National Key R&D Program of China (Grant Nos. 2019YFA0110900, 2019YFA0802202, 2019YFA0802200 and 2020YFA0803401), the National Natural Science Foundation of China (Grant Nos. 31870817 and 32170819), the Scientific and Technological Innovation Talent Project of Universities of Henan Province, China (Grant No. 20HASTIT045), the Shanghai Municipal Science and Technology Major Project, China (Grant No. 2017SHZDZX01), and the China Postdoctoral Science Foundation (Grant No. 2021M692927).

Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gpb.2022.05.005>.

ORCID

ORCID 0000-0001-9262-7254 (Xiao Han)
ORCID 0000-0003-0359-9680 (Jia Guo)

ORCID 0000-0001-5361-0986 (Mengke Wang)
ORCID 0000-0001-9477-9315 (Nan Zhang)
ORCID 0000-0002-6450-3163 (Jie Ren)
ORCID 0000-0002-8104-5985 (Ying Yang)
ORCID 0000-0002-6434-6882 (Xu Chi)
ORCID 0000-0001-5292-1301 (Yusheng Chen)
ORCID 0000-0003-4531-1550 (Huan Yao)
ORCID 0000-0003-0121-1312 (Yong-Liang Zhao)
ORCID 0000-0002-2821-8541 (Yun-Gui Yang)
ORCID 0000-0003-3783-6509 (Yingpu Sun)
ORCID 0000-0001-5177-5664 (Jiawei Xu)

References

- [1] He X, Memczak S, Qu J, Belmonte JCI, Liu GH. Single-cell omics in ageing: a young and growing field. *Nat Metab* 2020;2:293–302.
- [2] Zhao C, Zhang N, Zhang Y, Tuersunjiang N, Gao S, Liu W, et al. A DNA methylation state transition model reveals the programmed epigenetic heterogeneity in human pre-implantation embryos. *Genome Biol* 2020;21:277.
- [3] Yan L, Guo H, Hu B, Li R, Yong J, Zhao Y, et al. Epigenomic landscape of human fetal brain, heart, and liver. *J Biol Chem* 2016;291:4386–98.
- [4] Zhang H, Shi X, Huang T, Zhao X, Chen W, Gu N, et al. Dynamic landscape and evolution of m6A methylation in human. *Nucleic Acids Res* 2020;48:6251–64.
- [5] Xiao S, Cao S, Huang Q, Xia L, Deng M, Yang M, et al. The RNA N⁶-methyladenosine modification landscape of human fetal tissues. *Nat Cell Biol* 2019;21:651–61.
- [6] Wang Y, Yuan P, Yan Z, Yang M, Huo Y, Nie Y, et al. Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nat Commun* 2021;12:1247.
- [7] Domecke S, Hill AJ, Daza RM, Cao J, O'Day DR, Pliner HA, et al. A human cell atlas of fetal chromatin accessibility. *Science* 2020;370:eaba7612.
- [8] Je L, Li K, Cai J, Zhang M, Zhang X, Xiong X, et al. Landscape and regulation of m⁶A and m⁶Am methylome across human and mouse tissues. *Mol Cell* 2019;77:426–40.e6.
- [9] Melamed P, Yosefzon Y, David C, Tsukerman A, Pnueli L. Tet enzymes, variants, and differential effects on function. *Front Cell Dev Biol* 2018;6:22.
- [10] Hahn MA, Szabo PE, Pfeifer GP. 5-Hydroxymethylcytosine: a stable or transient DNA modification? *Genomics* 2014;104:314–23.
- [11] Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* 2011;8:200–13.
- [12] Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, et al. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* 2011;2:241.
- [13] Li X, Liu Y, Salz T, Hansen KD, Feinberg A. Whole-genome analysis of the methylome and hydroxymethylome in normal and malignant lung and liver. *Genome Res* 2016;26:1730–41.
- [14] Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 2011;477:606–10.
- [15] Amouroux R, Nashun B, Shirane K, Nakagawa S, Hill PW, D'Souza Z, et al. *De novo* DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat Cell Biol* 2016;18:225–33.
- [16] 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;11:6161.

- [17] Yang Y, Hsu PJ, Chen YS, Yang YG. Dynamic transcriptomic m⁶A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Res* 2018;28:616–24.
- [18] Chen YS, Yang WL, Zhao YL, Yang YG. Dynamic transcriptomic m⁵C and its regulatory role in RNA processing. *Wiley Interdiscip Rev RNA* 2021;12:e1639.
- [19] Xiong X, Hou L, Park YP, Molinie B, Consortium GT, Gregory RI, et al. Genetic drivers of m⁶A methylation in human brain, lung, heart and muscle. *Nat Genet* 2021;53:1156–65.
- [20] Han X, Wang M, Zhao YL, Yang Y, Yang YG. RNA methylations in human cancers. *Semin Cancer Biol* 2020;75:97–115.
- [21] Yang X, Yang Y, Sun BF, Chen YS, Xu JW, Lai WY, et al. 5-methylcytosine promotes mRNA export – NSUN2 as the methyltransferase and ALYREF as an m⁵C reader. *Cell Res* 2017;27:606–25.
- [22] Han D, Lu X, Shih AH, Nie J, You Q, Xu MM, et al. A highly sensitive and robust method for genome-wide 5hmC profiling of rare cell populations. *Mol Cell* 2016;63:711–9.
- [23] He B, Zhang C, Zhang X, Fan Y, Zeng H, Liu J, et al. Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nat Commun* 2021;12:4249.
- [24] Gan H, Wen L, Liao S, Lin X, Ma T, Liu J, et al. Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. *Nat Commun* 2013;4:1995.
- [25] Arab K, Karaulanov E, Musheev M, Trnka P, Schafer A, Grummt I, et al. GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat Genet* 2019;51:217–23.
- [26] Yoshida H, Lareau CA, Ramirez RN, Rose SA, Maier B, Wroblewska A, et al. The *cis*-regulatory atlas of the mouse immune system. *Cell* 2019;176:897–912.
- [27] Verma N, Pan H, Dore LC, Shukla A, Li QV, Pelham-Webb B, et al. TET proteins safeguard bivalent promoters from *de novo* methylation in human embryonic stem cells. *Nat Genet* 2018;50:83–95.
- [28] Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 2010;466:1129–33.
- [29] Kirk EP, Sunde M, Costa MW, Rankin SA, Wolstein O, Castro ML, et al. Mutations in cardiac T-box factor gene *TBX20* are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *Am J Hum Genet* 2007;81:280–91.
- [30] Poleev A, Fickenscher H, Mundlos S, Winterpacht A, Zabel B, Fidler A, et al. *PAX8*, a human paired box gene: isolation and expression in developing thyroid, kidney and Wilms' tumors. *Development* 1992;116:611–23.
- [31] Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 1997;13:335–40.
- [32] Coluccio A, Ecco G, Duc J, Offner S, Turelli P, Trono D. Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naive embryonic stem cells. *Epigenetics Chromatin* 2018;11:7.
- [33] Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, et al. A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell* 2018;174:391–405.
- [34] Chen LL, Yang L. ALU alternative regulation for gene expression. *Trends Cell Biol* 2017;27:480–90.
- [35] Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol* 2010;28:1045–8.
- [36] Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The human transcription factors. *Cell* 2018;172:650–65.
- [37] Yang Y, Wang L, Han X, Yang WL, Zhang M, Ma HL, et al. RNA 5-methylcytosine facilitates the maternal-to-zygotic transition by preventing maternal mRNA decay. *Mol Cell* 2019;75:1188–202.
- [38] Chen X, Li A, Sun BF, Yang Y, Han YN, Yuan X, et al. 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nat Cell Biol* 2019;21:978–90.
- [39] Iqbal K, Jin SG, Pfeifer GP, Szabo PE. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A* 2011;108:3642–7.
- [40] Wu J, Xu J, Liu B, Yao G, Wang P, Lin Z, et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* 2018;557:256–60.
- [41] Gao L, Wu K, Liu Z, Yao X, Yuan S, Tao W, et al. Chromatin accessibility landscape in human early embryos and its association with evolution. *Cell* 2018;173:248–59.
- [42] Tolsma TO, Hansen JC. Post-translational modifications and chromatin dynamics. *Essays Biochem* 2019;63:89–96.
- [43] Roque A, Ponte I, Suau P. Post-translational modifications of the intrinsically disordered terminal domains of histone H1: effects on secondary structure and chromatin dynamics. *Chromosoma* 2017;126:83–91.
- [44] Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 2012;151:1417–30.
- [45] Jacques PE, Jeyakani J, Bourque G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* 2013;9:e1003504.
- [46] Senft AD, Macfarlan TS. Transposable elements shape the evolution of mammalian development. *Nat Rev Genet* 2021;22:691–711.
- [47] Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, et al. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* 2013;45:836–41.
- [48] Carlsson AC, Nowak C, Lind L, Ostgren CJ, Nystrom FH, Sundstrom J, et al. Growth differentiation factor 15 (GDF-15) is a potential biomarker of both diabetic kidney disease and future cardiovascular events in cohorts of individuals with type 2 diabetes: a proteomics approach. *Ups J Med Sci* 2020;125:37–43.
- [49] Schmidt C, Christ B, Maden M, Brand-Saberi B, Patel K. Regulation of *Epha4* expression in paraxial and lateral plate mesoderm by ectoderm-derived signals. *Dev Dyn* 2001;220:377–86.
- [50] Arvaniti E, Moulos P, Vakkrou A, Chatziantoniou C, Chadjichristos C, Kavvadas P, et al. Whole-transcriptome analysis of UUO mouse model of renal fibrosis reveals new molecular players in kidney diseases. *Sci Rep* 2016;6:26235.
- [51] He Z, Zhang J, Huang H, Yuan C, Zhu C, Magdalou J, et al. Glucocorticoid-activation system mediated glucocorticoid-insulin-like growth factor 1 (GC-IGF1) axis programming alteration of adrenal dysfunction induced by prenatal caffeine exposure. *Toxicol Lett* 2019;302:7–17.
- [52] Skvortsova K, Iovino N, Bogdanovic O. Functions and mechanisms of epigenetic inheritance in animals. *Nat Rev Mol Cell Biol* 2018;19:774–90.
- [53] Shen H, Ontiveros RJ, Owens MC, Liu MY, Ghanty U, Kohli RM, et al. TET-mediated 5-methylcytosine oxidation in tRNA promotes translation. *J Biol Chem* 2021;296:100087.
- [54] Ren W, Fan H, Grimm SA, Guo Y, Kim JJ, Yin J, et al. Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. *Proc Natl Acad Sci U S A* 2020;117:18439–47.
- [55] Li Y, Xia L, Tan K, Ye X, Zuo Z, Li M, et al. N⁶-methyladenosine co-transcriptionally directs the demethylation of histone H3K9me2. *Nat Genet* 2020;52:870–7.
- [56] Wang Y, Li Y, Yue M, Wang J, Kumar S, Wechsler-Reya RJ, et al. N⁶-methyladenosine RNA modification regulates embryonic

- neural stem cell self-renewal through histone modifications. *Nat Neurosci* 2018;21:195–206.
- [57] Huang H, Weng H, Zhou K, Wu T, Zhao BS, Sun M, et al. Histone H3 trimethylation at lysine 36 guides m⁶A RNA modification co-transcriptionally. *Nature* 2019;567:414–9.
- [58] Wu C, Chen W, He J, Jin S, Liu Y, Yi Y, et al. Interplay of m⁶A and H3K27 trimethylation restrains inflammation during bacterial infection. *Sci Adv* 2020;6:eaba0647.
- [59] Zhou L, Tian S, Qin G. RNA methylomes reveal the m⁶A-mediated regulation of DNA demethylase gene *SIDML2* in tomato fruit ripening. *Genome Biol* 2019;20:156.
- [60] Wen L, Tang F. Single-cell sequencing in stem cell biology. *Genome Biol* 2016;17:71.
- [61] Li L, Guo F, Gao Y, Ren Y, Yuan P, Yan L, et al. Single-cell multi-omics sequencing of human early embryos. *Nat Cell Biol* 2018;20:847–58.
- [62] Carter B, Zhao K. The epigenetic basis of cellular heterogeneity. *Nat Rev Genet* 2021;22:235–50.
- [63] Ren X, Kang B, Zhang Z. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biol* 2018;19:211.
- [64] Widschwendter M, Jones A, Evans I, Reisel D, Dillner J, Sundstrom K, et al. Epigenome-based cancer risk prediction: rationale, opportunities and challenges. *Nat Rev Clin Oncol* 2018;15:292–309.
- [65] Jeschke J, Collignon E, Fuks F. Portraits of TET-mediated DNA hydroxymethylation in cancer. *Curr Opin Genet Dev* 2016;36:16–26.
- [66] Lagos C, Carvajal P, Castro I, Jara D, Gonzalez S, Aguilera S, et al. Association of high 5-hydroxymethylcytosine levels with Ten Eleven Translocation 2 overexpression and inflammation in Sjögren's syndrome patients. *Clin Immunol* 2018;196:85–96.
- [67] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20.
- [68] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
- [69] Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 2012;7:1728–40.
- [70] Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R137.
- [71] Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–89.
- [72] Shen L, Shao N, Liu X, Nestler E. ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 2014;15:284.
- [73] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015;12:357–60.
- [74] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30.
- [75] Rieder D, Amort T, Kugler E, Lusser A, Trajanoski Z. meRanTK: methylated RNA analysis ToolKit. *Bioinformatics* 2016;32:782–5.
- [76] Huang T, Chen W, Liu J, Gu N, Zhang R. Genome-wide identification of mRNA 5-methylcytosine in mammals. *Nat Struct Mol Biol* 2019;26:380–8.
- [77] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2.
- [78] Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* 2004;14:1188–90.
- [79] Kumar V, Muratani M, Rayan NA, Kraus P, Lufkin T, Ng HH, et al. Uniform, optimal signal processing of mapped deep-sequencing data. *Nat Biotechnol* 2013;31:615–22.
- [80] McLean CY, Bristol D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat Biotechnol* 2010;28:495–501.
- [81] Yevshin I, Sharipov R, Kolmykov S, Kondrakhin Y, Kolpakov F. GTRD: a database on gene transcription regulation-2019 update. *Nucleic Acids Res* 2019;47:D100–5.
- [82] Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 2003;4:P3.
- [83] Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;14:178–92.
- [84] Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, et al. The Genome Sequence Archive Family: toward explosive data growth and diverse data types. *Genomics Proteomics Bioinformatics* 2021;19:578–83.