The synergistic coordination of DNA 5-hydroxylmethylcytosine and RNA 5-methycytosine regulates human foetal development

Xiao Han
The First Affiliated Hospital of Zhengzhou University
Jia Guo
The First Affiliated Hospital of Zhengzhou University
Mengke Wang
Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation
Nan Zhang
The First Affiliated Hospital of Zhengzhou University
Jie Ren
Chinese Academy of Sciences https://orcid.org/0000-0002-6450-3163
Ying Yang
Beijing Institute of Genomics Chinese Academy of Sciences
Xu Chi
Beijing Institute of Genomics, Chinese Academy of Sciences
Yusheng Chen
Beijing Institute of Genomics, Chinese Academy of Sciences https://orcid.org/0000-0001-5292-1301
Huan Yao
Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation
Yongliang Zhao
Beijing Institute of Genomics, Chinese Academy of Sciences https://orcid.org/0000-0003-0121-1312
Yun-Gui Yang
Beijing Institute of Genomics, Chinese Academy of Sciences https://orcid.org/0000-0002-2821-8541
Yingpu Sun
First Affiliated Hospital of Zhengzhou University
Jiawei Xu (jiawxu@foxmail.com)
The First Affiliated Hospital of Zhengzhou University

Article

Keywords: Nano-hmC-Seal, RNA-BisSeq, human foetal organogenesis, DNA 5hmC, RNA m5C

Posted Date: May 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-478973/v1

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 and RNA bisulfite sequencing (RNA-BisSeq) were performed and the first multilayer landscapes of both DNA 5hmC and RNA m^5C epigenome, as well as the accompanying transcriptome were obtained, in the heart, kidney, liver and lung of the human foetuses at 13 – 28 weeks with 123 samples in total. DNA 5hmC plays a consistently leading role, while RNA m^5C acts as a dynamic downstream regulatory valve in regulating gene expression with a turning point of 18–19 weeks. Although, these modifications appear to coordinate in general, they also play different roles with regard to specific functions. Our integrated studies illustrate the epigenetic maps during different stages of human foetal organogenesis, which provide a foundation for understanding the in-depth epigenetic mechanisms for early development and birth defects.

Introduction

Epigenetic modifications, including DNA methylation[1–3], histone modifications[4–6], chromatin accessibility[7–10] and RNA modifications[11,12], not only act as heritable and stable indications for the specification of chromatin organization and structure, but also participating in regulation of transcriptional states during mammalian early development. The epigenomes of human embryos and the adult organs in both physiological and pathological conditions have been investigated extensively[3,6,11–15]; 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)[16]. This modification is widely distributed in the locations of promoters, enhancers and gene body, regulating target gene expression[17]. Now accumulating evidence showed 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 cell (ESCs)[18–22]. In fertilized oocytes, paternal 5mC is converted to 5hmC allowing drastic reprogramming of sperm DNA in mice[23]. 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 DNMT3A and DNMT1[24]. Although the regulatory power and plasticity of 5hmC have been revealed in human adults[20,25], 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[26–29]. The RNA m^6A profile in human adults[15] and mixed foetal samples from 18–25 weeks[12] has been

investigated, and the tissue-specific differences were found to be highly correlated with developmental processes[12]. Another prevalent modification, m^5C, is reported to regulate mRNA nuclear export[28,30], mRNA stability[31,32] and disease pathogenesis[33]. The conserved, organ-specific and dynamic features across mammalian transcriptomes of m^5C indicate its potential importance during mammalian development[28].

It has been hypothesizsed that the modifications on both DNA and RNA may coordinately work together in an intricate regulatory network. As histone H3 trimethylation at Lys36 (H3K36me3) and H3 trimethylation at Lys27 (H3K27me3) could guide or impede RNA m^6A deposition in mice and humans[34,35], this issue has been a subject of intense investigation in aim to elucidate the cross-talk between epigenetic factors in physiological and pathological conditions such as development and diseases. However, it is still elusive how the epigenome spatiotemporally shapes organ-specific development in a cooperative manner and further affects normal organ development in humans.

To address these questions, 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 fetuses, but also exhibit differential and synergistic roles during human foetal development.

Results

Genome-wide profiling of DNA 5hmC in human foetal organs. We collected samples of 4 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 (Fig. 1a and Supplementary Table 1). Nano-5hmC-Seal[36] was first conducted for DNA 5hmC profiling and displayed high reproducibility between replicates in different periods (Fig. S1a). To investigate genome-wide 5hmC features, we first examined the distributions of normalized reads of several development-related genes (Fig. 1b). The 5hmC enrichment of these genes prompted us to consider its organ-specific features in human foetuses at different stages.

To evaluate 5hmC global changes, we then illustrated 5hmC distributions along gene bodies (± 3 kb), which were similar to those of human adults[17,25] (Fig. 1c). Most of the 5hmC peaks were localized on vital functional elements, such as promoters (Fig. S1b). These results showed that the distribution of DNA 5hmC was conserved in human foetuses and adults, similar to other mammals[17,37]. Moreover, the distinct signal intensity of 5hmC among stages in each organ indicated that 5hmC may play different roles during organ development (Fig. 1c). For instance, in the foetal heart, the 5hmC signal was relatively more stable than that of other three organs probably because it was almost formed at 8 weeks and displayed fundamental biological functions (Fig. 1a, c). In the liver, the 5hmC signal was higher when haematopoiesis began, and even higher when the liver showed a strong haematopoietic ability (Fig. 1a,

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 differentially 5hmC-modified bins among four organs (ANOVA, fold change ≥ 1.5, *P* value < 0.05) (Fig. S1c), and the majority of these bins showed higher 5hmC levels than those of other organs (84.71–99.80%). Then adjacent bins were merged into one region called differentially hydroxymethylated regions (DhMRs) and 173,330–485,961 organ-specific DhMRs were obtained with around100–200 nt in length (Fig. S1d, S1e). No obvious preference of 5hmC modification was detected between gene body and intergenic regions, which was the same as for different chromosomes (Fig. S1f, g). Consistent with the annotation results of the 5hmC peaks in each organ, most of the organ-specific DhMRs were localized in functional elements (Fig. 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 4 types of organs. For example, in the liver, 5hmC enrichment in promoters was higher than that of other organs (Fig. 1d). Moreover, most of the organ-specific DhMRs were located near transcription start sites (TSSs), with prominent in the liver (Fig. 1e). These results indicated that mammalian-conserved 5hmC modifications are widespread but show significant organ-specific feature in human foetuses.

We further identified 10,658–20,419 organ-stage-specific DhMRs (ANOVA, fold change ≥ 1.5, *P* value < 0.05) and found that samples from the same organ clustered more closely, whereas that samples from the same stage were scattered on the periphery (Fig. 1f and Supplementary Table 2). The results indicated that 5hmC could serve as a potential marker specific for human organs with different developmental stages. Moreover, cluster analysis showed that both 5hmC signal intensity and the corresponding functions of the DhMRs were both organ- and stage-specific (Fig. 1g, 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 6 clusters according to the overall changes of 5hmC signal in different stages, considering that it contains all types of nonrepetitive trends for 6 clusters (Figs. 2a and S2a). Most of the DhMR-containing genes belong to at least 2 clusters, suggesting that the same gene may be regulated through multiple pathways in a 5hmC-dependent manner (Fig. S2b). Therefore, we separated these DhMRs from different clusters into 2 groups: a single-regulation group (gene sets with DhMRs only belonged to 1 cluster) and a mixed-regulation group (gene sets with DhMRs belonged to at least 2 clusters) (Fig. 2b). Then, TF prediction was performed on the DhMRs from both groups, and the top 15 TFs significantly related to development and three most commonly enriched TFs are shown (Fig. 2b). To explore how TFs are involved in 5hmC-dependent organ developmental processes, we performed RNA-Seq on samples used for the Nano-hmC-Seal data (Supplementary Table 3). There was a strong correlation between replicates (Fig. S2c). In both the single- and mixed-regulation groups, development-related TFs, including T-box transcription factor 20 (TBX20), paired box protein Pax-8 (PAX8), Krueppel-like factor 1 (KLF1), and transcription factor 21 (TCF21), were highly

enriched on DhMRs with organ-specific expression feature (Fig. 2b). Notably, a high 5hmC signal was 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[38]. The significant enrichment of this TF suggests its key role in the later stage of development, conferring full maturity of foetal heart. PAX8 was reported to be highly associated with the mesenchymal to epithelial transition involving metanephros morphogenesis[39]. 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 expression level of those TFs in foetuses was obviously higher than that in adults (Fig. S2d, e). 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)[40], some of which, however, could be used by their host genome as regulatory elements in certain organs during development[41–43]. 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 SINE elements show significant 5hmC enrichment on the DhMRs within cluster 3 and 4 in the foetal kidney (Figs. 2c and S2g–i). In particular, the binding sites of two TFs, namely CEBPB and TRPS1, are highly enriched in TSS-proximal Alu elements within these DhMRs (Figs. 2d and S2j), indicating that certain TFs preferentially bind to these Alu DhMRs, and their binding and *cis*-regulatory effect on target genes may be affected by differential 5hmC modifications of these Alu elements. The pattern of expression level changes of CEBPB target genes was consistent with its role as an activator in transcriptional regulation (Fig. 3e). Indeed, the expression levels of almost all CEBPB-targeted Alu-proximal genes (12/13) followed the 5hmC signals of the corresponding Alu DhMRs (Fig. 3e). 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 (Supplementary Table 4). 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 gene expression homeostasis across human organs. Most DhMRs were located around the TSSs in all foetal organs (Fig. 1e) and the sequential changes of 5hmC signal on DhMRs showed a strong association with RNA abundance during organ development (Fig. 2b, 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 (Fig. 3a), and most of the promoters contained 1–2 DhMRs (Fig. S3a). Importantly, their association was not set in stone during the

development of each organ, and higher correlation coefficients appeared at the later stages of almost all four organs (Fig. 3a). Then we identified genes with significant 5hmC alterations in promoters across adjacent stages (changes ≥ 25%, fold change ≥ 2, Student's t-test, *P* value < 1.03×10^{-6}) (Figs. 3b and S3b, c), and most of them were organ-specific promoters with obvious 5hmC changes (Fig. 3c). In particular, more than half of differentially expressed genes, such as those in the heart at 18–23 weeks (77.06%), liver at 18–23 weeks (74.09%) and 21–28 weeks (74.70%) (Figs. 3d and S3d), were subjected to 5hmC changes, and genes which were highly related to development were shown in Fig. 3d.

Additionally, we downloaded public DNase sequencing (DNase-Seq) data from the foetal heart, kidney and lung at 13–23 weeks[44] and found that there was a high correlation among the 5hmC signal, chromatin openness on promoters and gene expression level, for which genes could be detected from all three omics datasets (Fig. 3e). This finding suggests that 5hmC overlaps with open chromatin and is positively associated with transcriptional activity. However, we did not observe the above similar results between adjacent stages (Fig. 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 motifs, we performed pattern matching to reveal their one-to-one potential correspondence. To increase the reliability and specificity of the binding sequence, we used 30% of the single base proportion to obtain the optimal combinatorial sequences (Fig. S3f, g). Four kinds of random peaks/fragments were used to eliminate the effects of background noise (paired Student's t-test, *P* value < 0.05) (Fig. S3h). According to the results, the number of 5hmC-modified target genes is 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[45], with more target genes than others in foetal organs (Fig. S3i). We further calculated the correlation of the 5hmC signal and the gene expression level of target genes from different TF families (Fig. 3f). The regulatory relationship of DNA 5hmC and transcription remained almost stable in C2H2 ZF, bHLH, bZIP, *etc.*, whereas there were different regulatory roles in specific TF families, such as adaptor protein 2 (AP-2), early 2 factor (E2F), and homeodomain (Fig. 3f). All these findings indicate that 5hmC-modified promoters are highly associated with gene expression homeostasis during foetal organ development.

RNA m^5C acts as a downstream regulator during foetal organ development. In the setting of organ-specific DNA 5hmC in the regulation of upstream gene expression, we wondered how RNA m^5C drives organ-specific post-transcriptional regulatory networks, since it is one of the prevalent RNA post-transcriptional modifications associated with RNA stability[31,32]. 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 (Fig. S4a, b and Supplementary Table 5). Consistent with the findings in other vertebrates[28,29,31,32], the majority of m^5C sites were enriched near translation initiation sites (Fig. 4a). A total of 2,679–15,655 m^5C sites within 316–1,622 mRNAs were identified at each stage during organ development (Fig. 4b). Eliminating the effect of continuous Cs, most m^5C sites were strongly enriched in CG-rich and CDS regions (Fig. S4d, e). Considering that both the m^5C methylation level and the number of m^5C sites may affect the functions of corresponding mRNAs, we

calculated the total methylation level across different organs (Figs. 4c and S4c). Globally, m^5C levels dynamically change during the development of the kidney, liver and lung.

The global distributive features of m^5C in human foetal organs prompted us to explore the potential role of m^5C during development. Upon comparing RNA abundance changes between m^5C-modified and unmodified mRNAs across adjacent stages, we found that the m^5C-modified mRNAs were unstable before 18–19 weeks, whereas they became stabilized afterwards (P value < 5.40×10^{-7}), indicating that m^5C modification has a dual role in the regulation of organ development (Fig. 4d). This association was independent of coverage changes (Fig. S4f).

Additionally, consistent with that of DNA 5hmC, organ-specific m^5C-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 (Fig. S4g). Importantly, the organ-stage-specific m^5C-modified mRNAs were closely related to RNA abundance in each organ (Fig. 4e and Supplementary Table 6). These genes revealed a high level of enrichment for organ physiology-related processes. Thus, these results indicate that as a post-transcriptional regulator, RNA m^5C is indispensable during human foetal organ development.

DNA 5hmC and RNA m^5C synergistically regulate organ-specific developmental genes which are critical in maintaining normal foetal growth. We next proposed that DNA 5hmC and RNA m^5C 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 m^5C methylation level on their corresponding mRNAs, suggesting a dominant role of DNA 5hmC at the transcriptomic level during development (Fig. S5a). To further investigate the respective functions of DNA 5hmC and RNA m^5C, we divided the organ-specific modified genes into 4 groups: differentially expressed genes specifically regulated by both DNA 5hmC and RNA m^5C (group A), differentially expressed genes regulated only by DNA 5hmC (group B), differentially expressed genes regulated only by RNA m^5C (group C), and commonly expressed genes specifically regulated by DNA 5hmC and RNA m^5C (group D) (Figs. 5a, S5b, c and Supplementary Table 7). 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 (Fig. S5b). Specifically, in group A, there were close associations among DNA 5hmC signal intensity, RNA m^5C methylation level, and gene expression level (Fig. 5a).

Importantly, genes in all 3 groups (A, B and C) were involved in the process of organ formation and functional maturity (Fig. 5b and Supplementary Table 8). In the foetal heart, the genes from all three groups were significantly enriched in heart reconstruction, 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, m^5C modification is involved in atrial action potential and action potential depolarization, especially for SA and AV 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 m^5C was even more critical in cell differentiation of kidney, such as the regulation of mesenchymal epithelial cell transformation during metanephric morphogenesis, glomerular development, and renal vesicle morphogenesis. DNA 5hmC is the key driver of organic acid metabolic process in the liver. In the lung, both DNA 5hmC and RNA m^5C synergistically regulate respiratory gaseous exchange, and they separately modified different gene sets to modulate lung morphogenesis. These results suggest that DNA 5hmC and RNA m^5C generally play a synergistic role in the process of human foetal development, but in the foetal kidney and liver they exhibit differential and specific functions.

Interestingly, we found that 5hmC enrichment on promoters with m^5C modification on their corresponding mRNAs was much higher than that without m^5C (Fig. 5c), indicating a coordinated role of 5hmC and m^5C in foetal organs. TET3 was reported to be the only existed DNA 5mC oxidase after fertilization, mediating the oxidation of most paternal nuclear cytosines and a large-scale removal of 5mC[23,46]. 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 m^5C regulators. Apparently, the correlation coefficients of the expression levels of *TET3* and NOP2/Sun RNA methyltransferase 2 (*NSUN2*)/Aly/REF export factor (*ALYREF*) were high (0.7652 and 0.7664) (Fig. 5d). However, no significant relationship was found between *TET3* and Y-box binding protein 1 (*YBX1*). These results suggest that DNA 5hmC and RNA m^5C 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 developmental processes warrants further detailed investigation.

Discussion

The importance of epigenetic modifications in embryogenesis is well known, but the global function and co-regulation by DNA 5hmC and RNA m^5C 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 m^5C during human foetal development, and reveals their 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 m^5C dynamically modulate the changes in RNA abundance and are highly associated with the developmental functions of the corresponding organs. Moreover, this synergistic effect is more obvious at the later stages. The high correlation between *TET3* and *NSUN2*/*ALYREF* strongly suggests the potential cross-talk between DNA 5hmC and RNA m^5C. Collectively, the modifications in the layers of DNA and RNA perform their functions and serve synergistic roles during foetal organ development (Fig. 6).

A previous study showed 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[20], suggesting its positive regulatory roles in human organs. In foetuses, 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 (Fig. 3a, e). Chromatin accessibility of *cis*-regulatory sequences drives cell differentiation and zygotic genome activation (ZGA) by regulating gene expression in early mammalian embryos[7–9, 13]. Studies have shown that post-translational modifications, such as histone acetylation and ubiquitination, enhance chromatin accessibility by disrupting chromatin folding and globule formation[47,48]. 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[48–50]. 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.

RNA m^5C 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[31,32]. In human foetuses, the involvement of RNA m^5C was inferred to be divided into two periods with 18–19 weeks as the boundary. Compared with those in the former stages, m^5C-modified mRNAs were more stable than unmodified mRNAs after 18–19 weeks in all four organs. Generally, the fate of modified-transcripts is determined by their regulators: “writers”, “readers” and “erasers”[26]. In some cases, RNA methylation or its regulators play dual roles in cancer progression by affecting the stability of different target genes[33,51], such as the m^6A “writer” protein methyltransferase like 3 (METTL3) in colorectal cancer (CRC). We hypothesize that the differential roles of RNA m^5C may result from the abundance/binding affinity changes of m^5C regulators on different target genes to maintain normal development. In addition, the high correlation between the expression level of *TET3* and *NSUN2/ALYREF* suggests that DNA methylase and RNA m^5C methylase/binding proteins may synergistically regulate the early foetal organ development. A previous study demonstrated that H3K36me3 facilitates METTL3/METTL14/WTAP-dependent m^6A deposition at the 3'end of coding sequences[34]. We speculate that TET3 not only mediates 5hmC formation, but also might interact with NSUN2/ALYREF, inducing mRNAs to obtain m^5C modification and thus maintaining stability. However, the exact molecular mechanisms need to be further studied.

Taken together, this study not only uncovered dynamic DNA 5hmC and RNA m^5C 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 a foundation for understanding the in-depth epigenetic mechanisms for early human development.

Methods

Tissue collection and ethics approval of this study. All of the human foetal organs at different developmental stages were obtained from the donors voluntarily at the First Affiliated Hospital of Zhengzhou University with signed informed consent and the approval of the Ethics Committee (license no. 2020-KY-261). Foetuses were at 13 to 28 weeks in the second trimester of pregnancy. Foetal heart, liver, kidney, and lung were collected from selective induced labor fetus of donors with normal chromosome karyotypes and no family history of infectious diseases, hereditary diseases and smoking history. The foetal organs were isolated after a mechanical dissection, serially washed with medical saline for three times, stored in RNAlater Stabilization Reagent (AM7024, Invitrogen) and immediately frozen in liquid nitrogen. The details of this research, including the research procedure, the benefits and risks of donation, the confidentiality of the donor's information, the right of withdrawal, the generation and analysis of sequencing data, and the publication of results, were explained to the donors before they gave consent. All the experiments were also under the regulations of International Ethical Guidelines for Biomedical Research Involving Human Subjects and in compliance with the policies and laws of China.

Genomic DNA and total RNA preparation. A maximum amount of 20 mg of tissue ground in a 100:1 mixture of 600 µl Buffer RLT Plus and 6 µl 14.3M β-mercaptoethanol for genomic DNA and total RNA extraction by user operation manual of AllPrep DNA/RNA/miRNA Universal Kit (80224, Qiagen). For rRNA depletion, the purified total RNA was further treated by a 3:1 mixture of probe and total RNA followed by RNase H (EN0202, Thermo Fisher) treatment at 37°C. TURBO DNase (AM2238, Invitrogen) treatment was used to eliminate DNA contamination. Then, the rRNA-depleted RNAs were 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 (Thermo Fisher) and Qubit (Life Technologies).

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 (KK8541, KAPA). In brief, the high-quality, rRNA-depleted RNAs of foetal organs were fragmented for 6 min at 85℃ (about 300 nucleotides) by 2× Fragment, Prime and Elute Buffer in different periods. The DNA concentration and quality of constructed libraries were detected by Qubit (Life Technologies) and 2100 Bioanalyzer (Agilent) before sequencing.

Bisulfite conversion of RNA. RNA fragmentation and bisulfite conversion were consistent with previous description[28], but had some modifications. Briefly, in vitro transcribed mouse Dhfr mRNA, as a methylation conversion reference, was mixed with purified rRNA removal RNA at a ratio of 1:200. The mixture was fragmented into ~ 200 nucleotides by 10× RNA Fragmentation Reagent (AM8740, Thermo Fisher) for 22s at 88℃ which was terminated by 10× RNA stop solution. Then fragmented mRNAs were obtained by ethanol precipitation. After washing precipitation with 100% ethanol, the RNA precipitates were dissolved in 100 µl bisulfite solution that mixed 40% sodium bisulfite (243973,Sigma) and 600 µM hydroquinone (H9003Sigma) in a ratio of 100:1. The synthesis was incubated at 75 ℃ for 4 hours. Nanosep columns with 3K Omega membranes (OD003C35, Pall Corporation) 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 75μl of nuclease-free water and incubated at 75°C for 1 hour 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 (18064014, Invitrogen) and ACT random hexamers to form cDNA, KAPA RNA HyperPrep Kit (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 above 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[36]. The reactions were performed using TruePrep DNA Library Prep Kit V2 for Illumina (TD501, Vazyme) 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, 50ng gDNA and 5 μl TTE mix V50 at 55°C for 10 min. The DNA fragmentations were purified by 1X AMPure XP beads (A63882, Beckman) eluted by RNase-free water. The reaction of glucosylation occurred in synthesis of 50 mM HEPES buffer (pH 8.0), 25 mM $MgCl_2$, fragmented DNA, 100 μM N3-UDP-Glc, and 1 μM T4 beta-glucosyltransferase (βGT) (EO0831, Thermo) at 37°C for 1 hour. After glucosylation reaction, the reaction solution mixed with 2μL DBCO-PEG4-DBCO (20 mM stock in DMSO) and incubated at 37°C for 2 hr. The Post-reaction solution was desalted by Micro Bio-Spin P-30 Gel Columns (7326226,Bio–Rad) to obtain the purified DNA. 5 μL Dynabeads MyOne Streptavidin C1 (65002, Life Technologies) in 2× buffer (1× buffer: 5 mM Tris [pH 7.5], 0.5 mM EDTA, and 1 M NaCl) 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)[52] 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)[53]. Reads with quality score ≥ 20 were retained for the subsequent analysis. MACS2 (version 2.1.1)[54,55] were 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)[56]. analyzeRepeats.pl (HOMER, version 4.9.1)[56] was performed to do peak annotation and calculate the raw read counts (RPM) of gene bodies and promoters [defined as the regions 500bp from the transcription start sites (TSS)]. The average RPM of each gene from replicates was calculated for downstream analysis. The Pearson correlation coefficients of DNA 5hmC signal between biological replicates at each stage were from 0.7164 to 0.9537. The average profile of 5hmC in each sample was visualized by ngsplot (version 2.61)[57] with parameters: -G hg19 -R gene body -L 3000. (2) RNA-Seq analysis: Clean

reads were aligned to human genome (hg19) using HISAT2 (version 2.0.5)[58] 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)[59] 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. Pearson correlation coefficients of gene expression level between biological replicates at each stage were from 0.7385 to 0.9714. (3) RNA-BisSeq analysis: Clean reads were aligned to human genome (hg19) by meRanT align (meRanTK, version 1.2.0)[60] with parameters: -fmo -mmr 0.01. The samples with conversion ratio over 99% were used to perform m^5C-calling by meRanCall (meRanTK, version 1.2.0)[60] with parameters: -mBQ 20 -mr 0 -cr 0.99 -fdr 0.05. 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 m^5C sites were annotated by BEDTools' intersectBed (version 2.26.0)[61]. The overlap ratios of m^5C 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[41], the m^5C distribution among CDSs, 5'UTRs, 3'UTRs, intron, CG, CHG, and CHH (H = A, C, U) were calculated. The sequences that up- and down-stream 10 nt of m^5C sites were used to detect the sequence preference, and logo plots were generated with WebLogo (version 3)[62]. (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)[63] with the parameters: -ks = 50 -lpval = 2 -f = bam.

Differential 5hmC regions (DhMRs) identification. A non-overlapping 100bp window was applied to perform DhMRs identification across the samples from different organs. Briefly, the genome was first binned into consecutive 100bp windows, and the signal intensity of each window was calculated based on the normalized RPM. Then the average 5hmC signal of replicates were calculated. ANOVA (Analysis of Variance) was used to compute the magnitude of the difference for each bin in R (version 3.4.1) (https://www.r-project.org/). The *P* values were calculated using Benjamini-Hochberg method. The bins with fold change > 1.5 and *P* value < 0.05 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 or ChMRs. annotatePeaks.pl (HOMER, version 4.9.1)[56] was used to do DhMRs and ChMRs annotation. Principal component analysis (PCA) was performed by factoextra (R package). The distance between TSS and DhMRs was calculated by BEDTools bedtools closest (version 2.26.0)[61]. GREAT (version 4.0.4)[64] was used to do annotation for DhMRs.

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) (http://mev.tm4.org). When the number of clusters was greater than 6, similar 5hmC trends appeared. Thus, the DhMRs was divided into 6 clusters finally. HOMER (version 4.9.1)[56] was used to performed TF prediction in DhMRs of each

clusters with the following command: findMotifsGenome.pl input.bed hg19 -size given -len 6 -norevopp -cache 1000. Genes with TF-potentially-binding DhMRs only from one clusters are considered as single-regulation groups, whereas genes with TF-potentially-binding DhMRs from two or more clusters are considered as mixed-regulation groups.

Differential hypomethylated TEs analysis. Each of the six DhMR clusters of the four organs was intersected with RepeatMasker (version 4.0.5) (http://repeatmasker.org) using BEDTools' intersectBed (version 2.26.0)[61]. 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 towards 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 (Fig. S2f). 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)[65] human meta cluster interval file, keeping only the intersections that covers over 50% of the Alu elements. The results were further filtered to remove binding sites of TFs that were scarcely expressed at all 4 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 intersection of either TF binding regions with these Alu DhMRs. This density was then divided by either the density of the intersection 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 intersection 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 for which the fold change of 5hmC signal between samples from two adjacent stages exhibited a greater than 25% were defined as significant different promoters (Student's t-test with P value < 0.05). Thus, the promoters were separated into two group sets: promoters with decreased 5hmC signal and promoters with increased 5hmC signal. To explore the relationship between 5hmC signals in promoters and gene expression, the gene expression levels of corresponding group sets were compared and the P values were calculated using two-sided Wilcoxon and Mann-Whitney test. The genes with fold change over 1.2 were defined as differential expressed genes in 5hmC-decreased and -increased groups.

Corresponding 5hmC peak discovery based on known TF motifs. To further explore how different TFs regulate gene expression via binding to 5hmC-enriched promoters, TFs-binded 5hmC peaks were identified based on the known motifs of TFs by pattern matching. The motifs of TFs were downloaded from HumanTFs database[45] (http://humantfs.ccbr.utoronto.ca/) which contains ~ 1600 TFs' information (~ 5000 possible motifs). Different TFs could bind to the sequence with the same motif and one TF could recognize the sequence 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 sequence, 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 kind 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 that of the average of 5hmC peaks; 4) random genome fragments in distal region in the same length of that of the average of 5hmC peaks. The fragments were randomly selected for 20 times per background. The motif sequence remained if the number of actual peaks divided by the number of fragments in the background > 2 for at least one random fragments group. The *P* values were calculated using Paired Student t-test. The remaining motifs that could be detected based on all the four kinds of backgrounds were considered credible.

Dynamic changes of RNA m^5C. All expressed genes were divided into two groups based on with or without m^5C modification at adjacent time points. To further evaluate the effects of m^5C dynamic changes on mRNA abundance change, mRNAs with m^5C modification at last stage but without m^5C modification at next stage were defined as mRNAs with m^5C loss; mRNAs without m^5C modification at last stage but with m^5C modification at next stage were defined as mRNAs with m^5C 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-specific m^5C-modified genes were identified according to the DhMRs methods.

Association analysis with different organ-specific groups. To analyze the relationship between DNA 5hmC, RNA m^5C and transcriptome, the 5hmC signal of promoter, m^5C methylation level and gene expression level (RPKM) were subjected to $\log_2$ 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-specific 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 4 groups: differentially expressed genes that DNA 5hmC and RNA m^5C specifically regulated, differentially expressed genes that were only specifically regulated by DNA 5hmC, differentially expressed genes that were only specifically regulated by RNA m^5C, and commonly expressed genes that DNA 5hmC and RNA m^5C specifically regulated.

Gene ontology analysis and visualization. DAVID (version 6.8)[66] was used to perform Gene Ontology (GO) analysis. GO terms with *P* value < 0.05 were considered as statistically significant. IGVTools (version 2.3.8)[67] was used for visualization.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw and processed Nano-hmC-Seal, RNA-BisSeq and RNA-seq data have been deposited in the Genome Sequence Archive[68] in BIG Data Center[69], Beijing Institute of Genomics (BIG)/ China National Center for Bioinformation, Chinese Academy of Sciences, with accession number GSA: HRA000705 (project PRJCA004624). Public Nano-hmC-Seal and RNA-Seq data from adult organs in this study were downloaded from GSE144530. Public DNase-Seq data from 13–23 weeks foetal organs were downloaded from GSE18927.

Code availability

Software and code used to analyze these data are listed in the Nature Research Reporting Summary.

References

1. Zhou, F. et al. Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature* **572**, 660–664 (2019).
2. Zhu, P. et al. Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat. Genet.* **50**, 12–19 (2018).
3. Zhao, C. et al. A DNA methylation state transition model reveals the programmed epigenetic heterogeneity in human pre-implantation embryos. *Genome Biol.* **21**, 277 (2020).
4. Xia, W. et al. Resetting histone modifications during human parental-to-zygotic transition. *Science* **365**, 353–360 (2019).
5. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
6. Yan, L. et al. Epigenomic landscape of human fetal brain, heart, and liver. *J. Biol. Chem.* **291**, 4386–4398 (2016).
7. Wu, J. et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018).
8. Guo, H. et al. DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Res.* **27**, 165–183 (2017).
9. Gao, L. et al. Chromatin accessibility landscape in human early embryos and its association with evolution. *Cell* **173**, 248–259 (2018).

10. Li, L. et al. Single-cell multi-omics sequencing of human early embryos. *Nat. Cell Biol.* **20**, 847–858 (2018).
11. Zhang, H. et al. Dynamic landscape and evolution of m^6A methylation in human. *Nucleic Acids Res.* **48**, 6251–6264 (2020).
12. Xiao, S. et al. The RNA N^6-methyladenosine modification landscape of human fetal tissues. *Nat. Cell Biol.* **21**, 651–661 (2019).
13. Wang, Y. et al. Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nat. Commun.* **12**, 1247 (2021).
14. Domcke, S. et al. A human cell atlas of fetal chromatin accessibility. *Science* **370** (2020).
15. Liu, J'e. et al. Landscape and regulation of m^6A and m^6Am methylome across human and mouse tissues. *Mol. Cell* **77**, 426–440 (2019).
16. Melamed, P., Yosefzon, Y., David, C., Tsukerman, A. & Pnueli, L. Tet Enzymes, variants, and differential effects on function. *Front. Cell Dev. Biol.* **6**, 22 (2018).
17. Hahn, M.A., Szabo, P.E. & Pfeifer, G.P. 5-Hydroxymethylcytosine: a stable or transient DNA modification? *Genomics* **104**, 314–323 (2014).
18. Koh, K.P. et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**, 200–213 (2011).
19. Wossidlo, M. et al. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* **2**, 241 (2011).
20. Li, X., Liu, Y., Salz, T., Hansen, K.D. & Feinberg, A. Whole-genome analysis of the methylome and hydroxymethylome in normal and malignant lung and liver. *Genome Res.* **26**, 1730–1741 (2016).
21. Tian, X. et al. Circulating tumor DNA 5-hydroxymethylcytosine as a novel diagnostic biomarker for esophageal cancer. *Cell Res.* **28**, 597–600 (2018).
22. Zhang, J. et al. 5-Hydroxymethylome in circulating cell-free DNA as a potential biomarker for non-small-cell lung cancer. *Genomics Proteomics Bioinformatics* **16**, 187–199 (2018).
23. Gu, T.P. et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
24. Amouroux, R. et al. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18**, 225–233 (2016).
25. Cui, X.L. et al. A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat. Commun.* **11**, 6161 (2020).
26. Yang, Y., Hsu, P.J., Chen, Y.S. & Yang, Y.G. Dynamic transcriptomic m^6A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Res.* **28**, 616–624 (2018).
27. Amort, T. et al. Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biol.* **18**, 1 (2017).

28. Yang, X. et al. 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m^5C reader. *Cell Res.* **27**, 606–625 (2017).
29. Chen, Y.S., Yang, W.L., Zhao, Y.L. & Yang, Y.G. Dynamic transcriptomic m^5C and its regulatory role in RNA processing. *Wiley Interdiscip. Rev. RNA*, e1639 (2021).
30. Pfaff, C. et al. ALY RNA-binding proteins are required for nucleocytosolic mRNA transport and modulate plant growth and development. *Plant Physiol.* **177**, 226–240 (2018).
31. Yang, Y. et al. RNA 5-methylcytosine facilitates the maternal-to-zygotic transition by preventing maternal mRNA decay. *Mol. Cell* **75**, 1188–1202 (2019).
32. Chen, X. et al. 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nat. Cell Biol.* **21**, 978–990 (2019).
33. Han, X., Wang, M., Zhao, Y.L., Yang, Y. & Yang, Y.G. RNA methylations in human cancers. *Semin. Cancer Biol.* S1044-579X(20)30241–8 (2020).
34. Huang, H. et al. Histone H3 trimethylation at lysine 36 guides m^6A RNA modification co-transcriptionally. *Nature* **567**, 414–419 (2019).
35. Wu, C. et al. Interplay of m^6A and H3K27 trimethylation restrains inflammation during bacterial infesction. *Sci. Adv.* **6**, eaba0647 (2020).
36. Han, D. et al. A highly sensitive and robust method for genome-wide 5hmC profiling of rare cell populations. *Mol. Cell* **63**, 711–719 (2016).
37. Gan, H. et al. Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. *Nat. Commun.* **4**, 1995 (2013).
38. Kirk, E.P. 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.* **81**, 280–291 (2007).
39. Poleev, A. et al. PAX8, a human paired box gene: isolation and expression in developing thyroid, kidney and Wilms' tumors. *Development* **116**, 611–623 (1992).
40. Yoder, J.A., Walsh, C.P. & Bestor, T.H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
41. Coluccio, A. et al. 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* **11**, 7 (2018).
42. Percharde, M. et al. A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell* **174**, 391–405 (2018).
43. Chen, L.L. & Yang, L. ALUternative regulation for gene expression. *Trends Cell Biol.* **27**, 480–490 (2017).
44. Bernstein, B.E. et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
45. Lambert, S.A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).

46. Iqbal, K., Jin, S.G., Pfeifer, G.P. & Szabo, P.E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Nat.l Acad. Sci. U. S. A.* **108**, 3642–3647 (2011).
47. Tolsma, T.O. & Hansen, J.C. Post-translational modifications and chromatin dynamics. *Essays Biochem.* **63**, 89–96 (2019).
48. 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* **126**, 83–91 (2017).
49. 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* **151**, 1417–1430 (2012).
50. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
51. Haruehanroengra, P., Zheng, Y.Y., Zhou, Y., Huang, Y. & Sheng, J. RNA modifications and cancer. *RNA Biol.* **17**, 1560–1575 (2020).
52. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
53. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
54. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X.S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).
55. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
56. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
57. 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* **15**, 284 (2014).
58. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
59. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
60. Rieder, D., Amort, T., Kugler, E., Lusser, A. & Trajanoski, Z. meRanTK: methylated RNA analysis ToolKit. *Bioinformatics* **32**, 782–785 (2016).
61. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
62. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
63. Kumar, V. et al. Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.* **31**, 612–622 (2010).

64. McLean, C.Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
65. Yevshin, I., Sharipov, R., Kolmykov, S., Kondrakhin, Y. & Kolpakov, F. GTRD: a database on gene transcription regulation-2019 update. *Nucleic Acids Res.* **47**, D100-D105 (2019).
66. Dennis, G., Jr. et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, P3 (2003).
67. Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* **14**, 178–192 (2013).
68. Wang, Y. et al. GSA: Genome Sequence Archive. *Genomics Proteomics Bioinformatics* **15**, 14–18 (2017).
69. BIG Data Center Members. The BIG Data Center: from deposition to integration to translation. *Nucleic Acids Res.* **45**, D18-D24 (2017).

Declarations

Acknowledgements

This work was supported by National Key R&D Program of China (2019YFA0110900), National Natural Science Foundation of China (31870817), Scientific and Technological Innovation Talent Project of Universities of Henan Province (20HASTIT045), and Shanghai Municipal Science and Technology Major Project (2017SHZDZX01).

Author contributions

J.X., Y.S. and Y.G.Y. conceived and designed the project. J.G. and N.Z collected foetal tissues and performed Nano-hmC-Seal, RNA-BisSeq and RNA-Seq library construction. X.H., M.W. and X.C. performed bioinformatics analysis and data interpretation with the help from J.R., Y.C. and Y.Y.. H.Y. assisted with the high-throughput sequencing. X.H., J.G., M.W. and J.R. wrote the manuscript with the help of all authors. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Figures

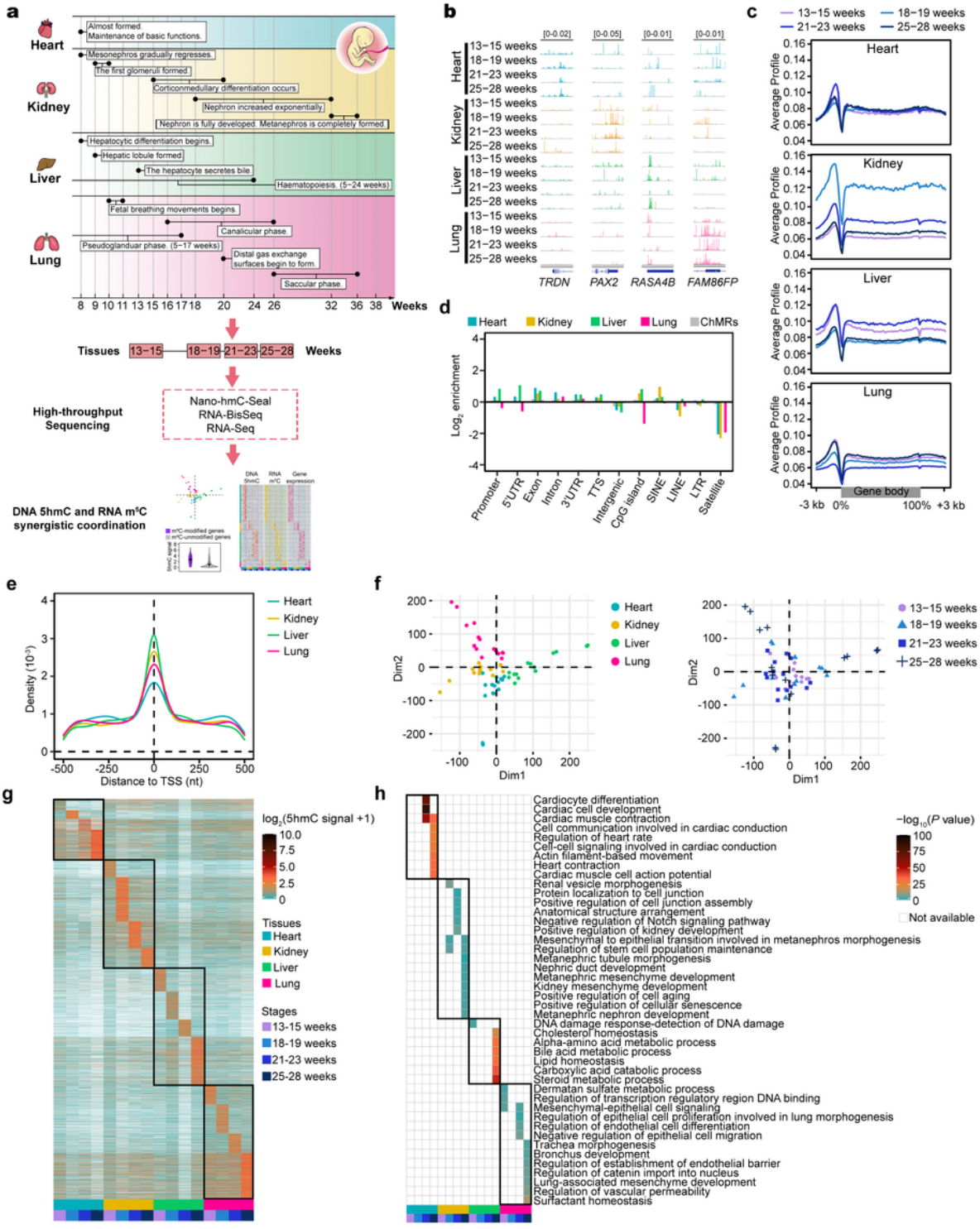

Figure 1

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, Representative Integrative Genomics Viewer (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. f, Principal component analysis (PCA) of 5hmC signals in organ-stage-specific DhMRs from different organ samples. g, Heatmap showing the 5hmC signals of organ-stage-specific DhMRs. h, GREAT analysis results for organ-stage-specific DhMRs. The color represents the significance of each biological process.

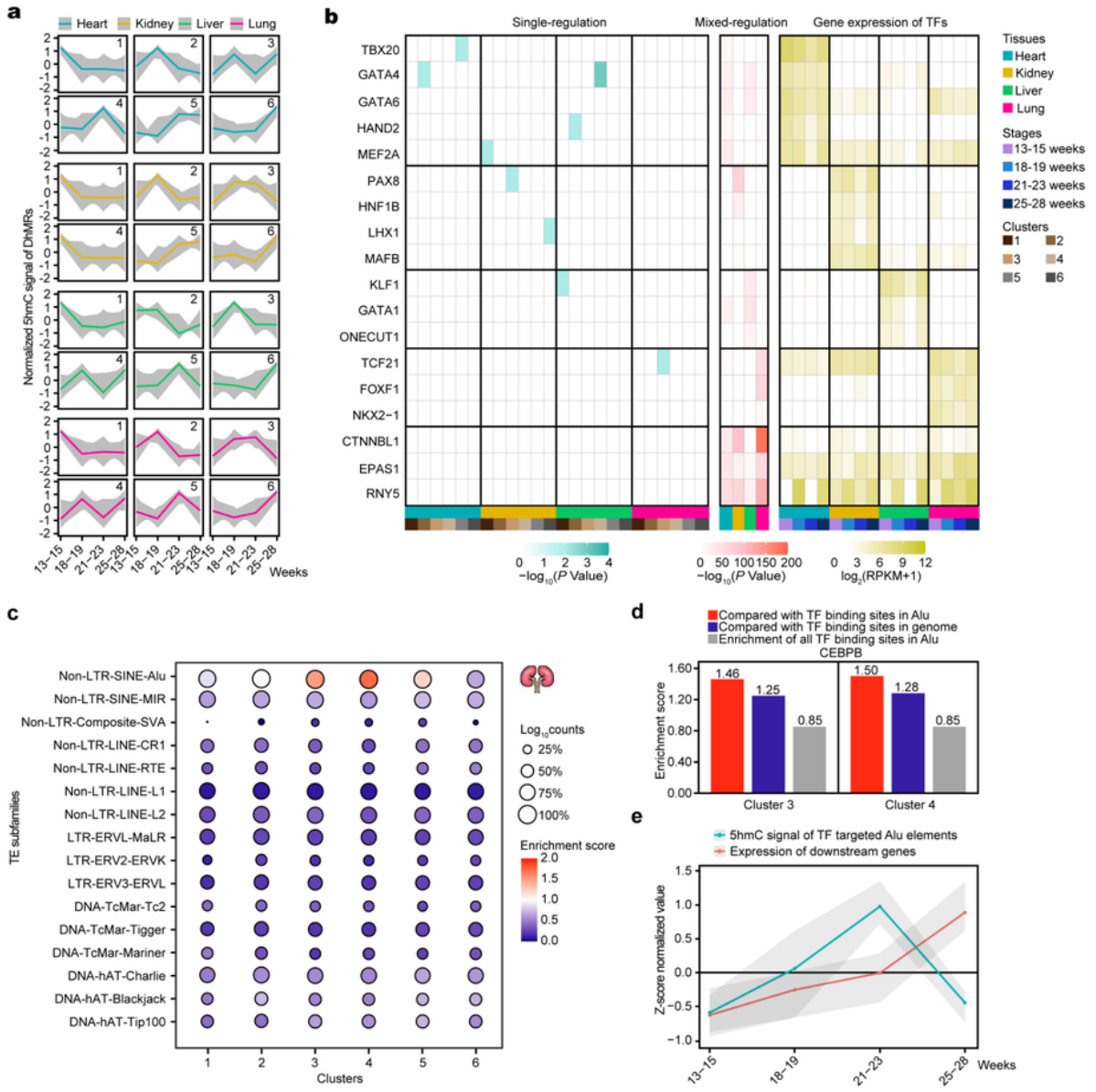

Figure 2

DNA 5hmC dynamically shapes foetal organ development by transcription factors (TFs) as well as transposable elements (TEs). a, K-means clustering analysis of organ-specific DhMRs from heart, kidney, liver and lung. Each organ was separated into 6 clusters. b, Heatmap showing the TF motifs identified by DhMRs. The TF-binding significance of single- and mis-regulation gene groups, and the expression levels of each TF are shown from left to right. c, Enrichment of specific TE families in organ-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 signal of CEBPB-targeted Alu elements as in d and the expression level of downstream genes during human foetal kidney development

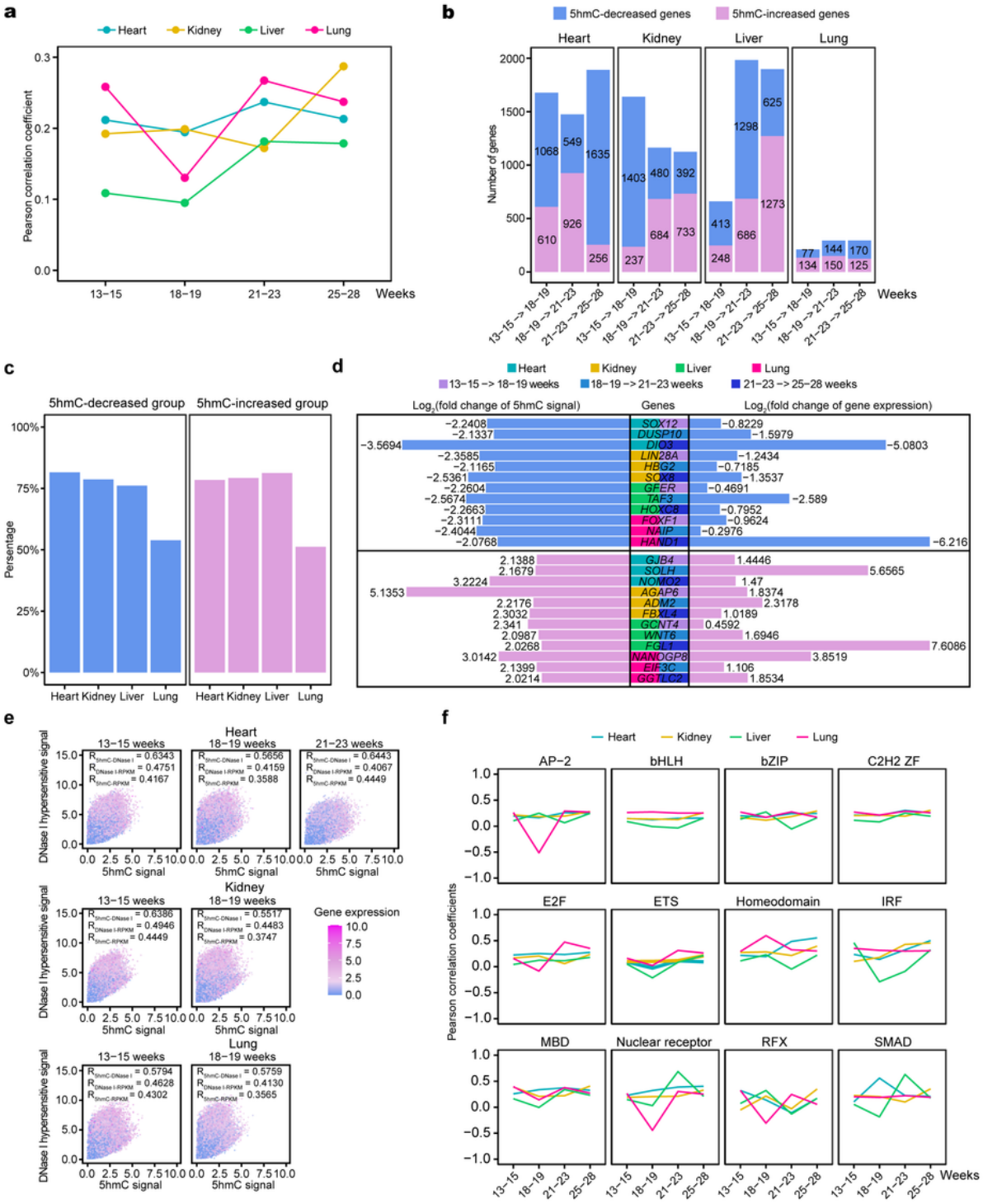

Figure 3

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 5hmC significant changes in each organ. d, Bar plots showing the fold changes (log2) of 5hmC signal and gene

expression level of 18 organ-specific genes from 5hmC decrease and increase 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 coefficient of different TF family groups from heart, kidney, liver and lung. Pearson's correlation coefficients were calculated based on 5hmC signals and gene expression level in corresponding gene sets.

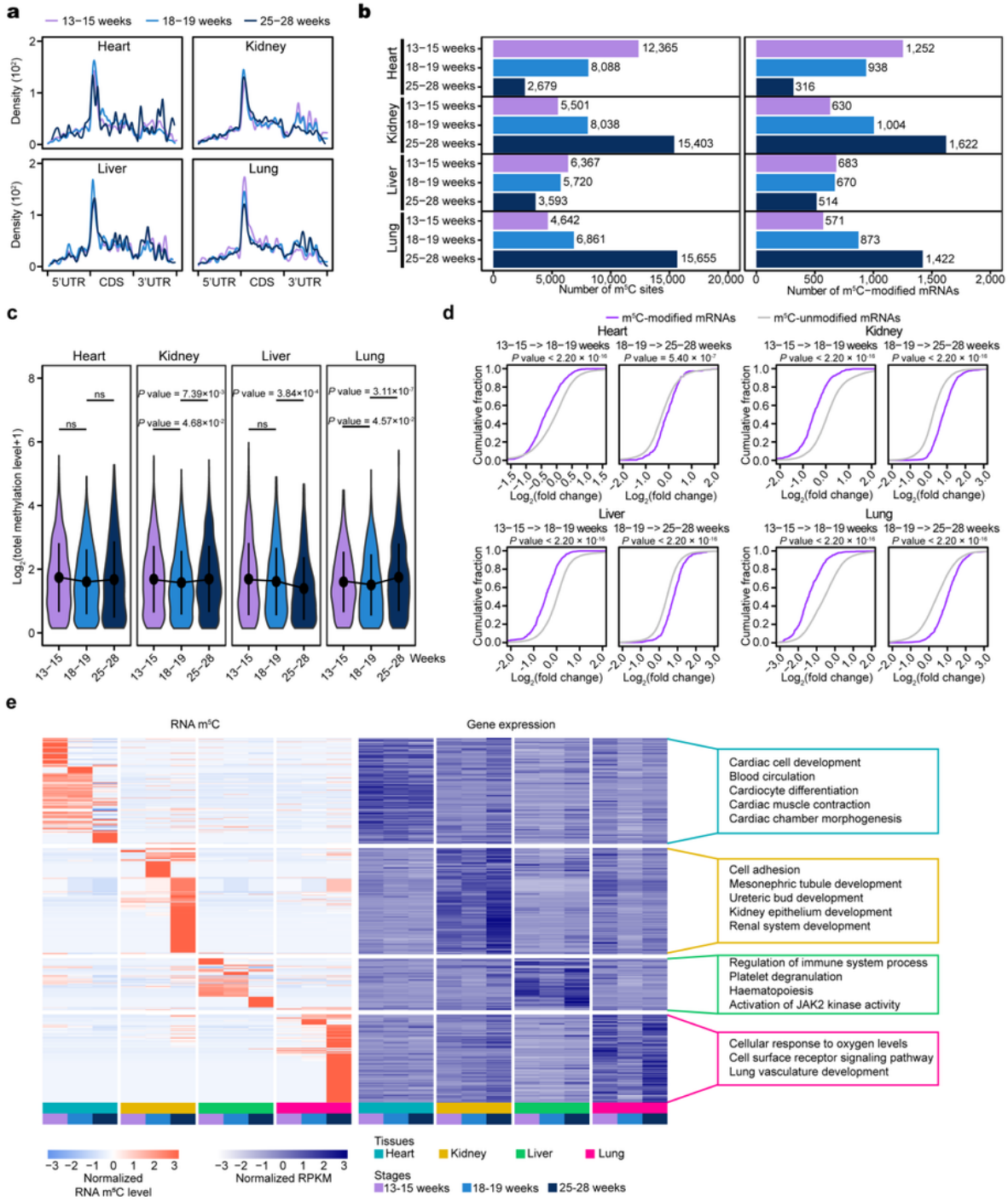

Figure 4

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 show the numbers of m5C sites and m5C-modified mRNAs. c, Overall total RNA m5C methylation level across different developmental stages in heart, kidney, liver and lung. d, Cumulative distribution displaying the expression level change of m5C-modified and unmodified mRNAs from RNA-Seq data comparing samples from adjacent stages. The P values in c and d were calculated using two-sided Wilcoxon and Mann-Whitney tests. e, Heatmap showing the strong positive correlation between the m5C methylation level of organ-stage-specific genes and the expression level of corresponding genes during foetal organ development. GO biological processes for each organ-stage-specific gene set are listed on the right.

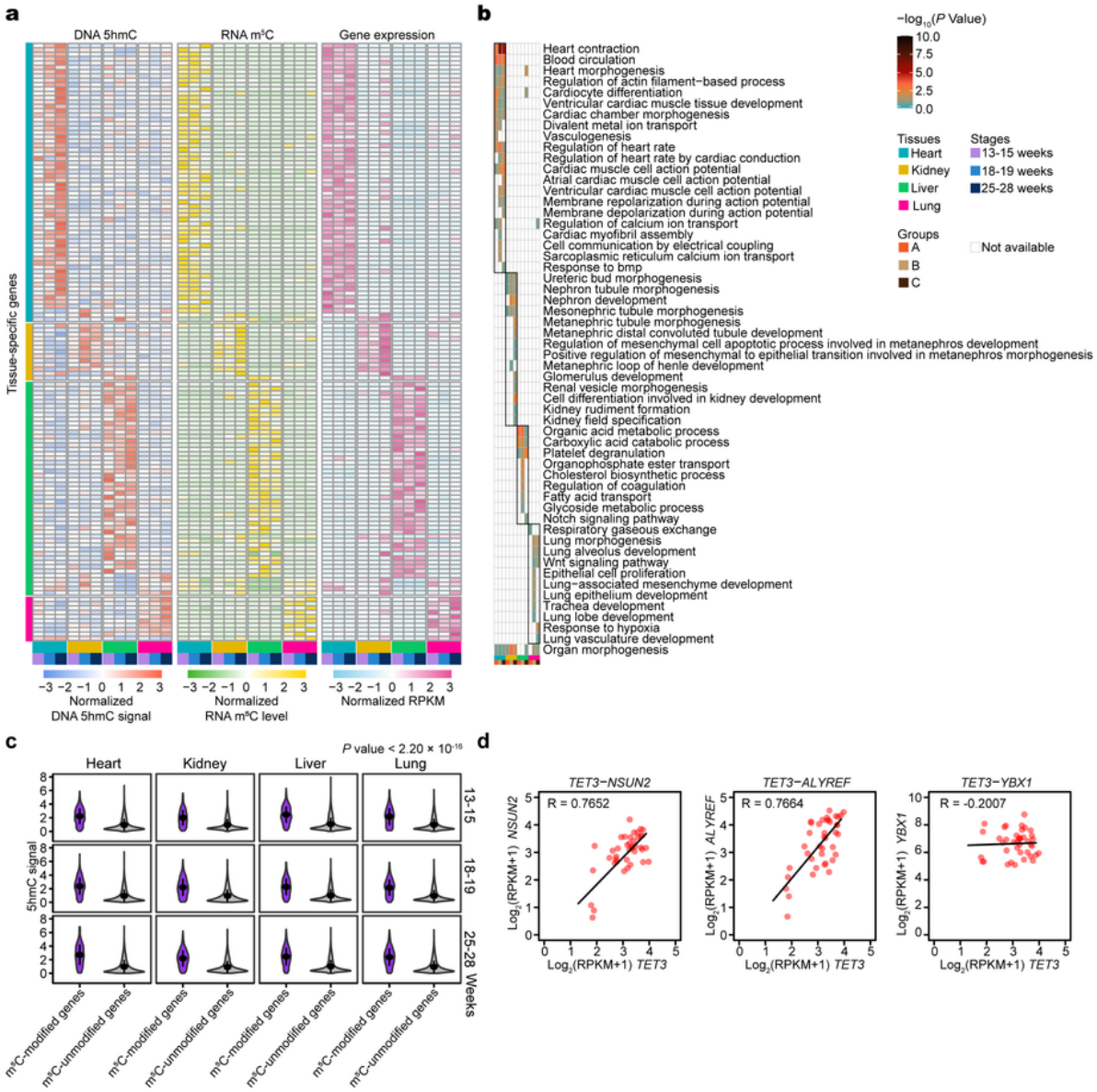

Figure 5

DNA 5hmC and RNA m5C consistently regulate foetal organ development. a, Heatmaps showing the dynamics of DNA 5hmC signal on promoters (left), m5C methylation level (middle), and expression level (right) of DNA5hmC-RNA m5C co-regulated specific-expressed gene (group A). DNA 5hmC signal, m5C methylation level and gene expression level 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 signal on promoters of corresponding genes with or without m5C modification. 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.

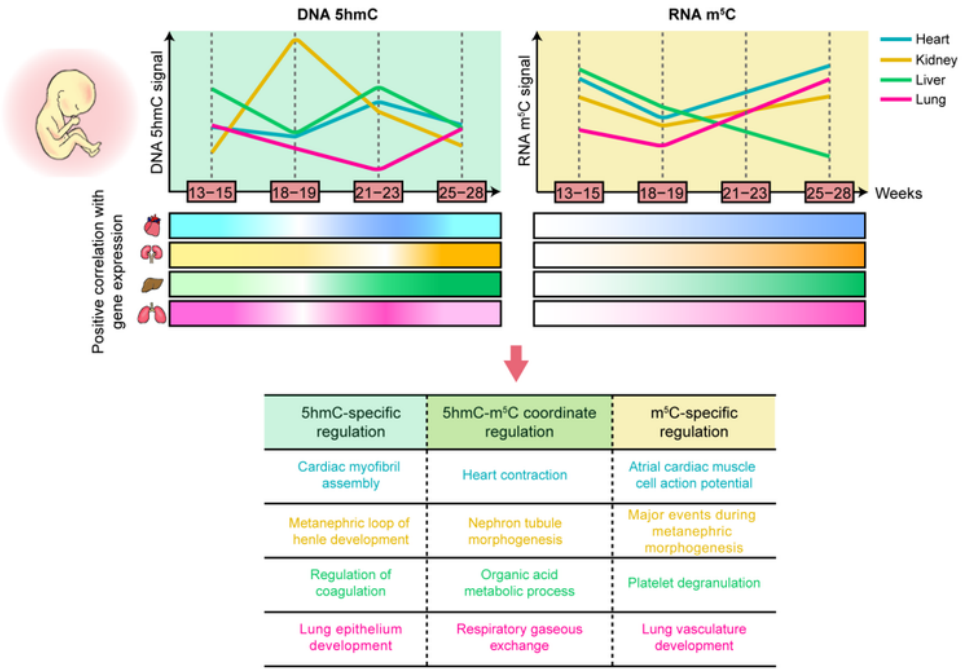

Figure 6

Schematic model of the dynamic regulation of DNA 5hmC and RNA m5C during foetal organ development. In human foetuses, both DNA 5hmC and RNA m5C subject to dynamic changes during the development of heart, kidney, liver and lung. The correlation changes between DNA 5hmC signal and gene expression level are different in distinct organs, whereas the correlation changes between m5C methylation level and gene expression level are consistent in four organs. Both DNA 5hmC and RNA m5C coordinated regulate foetal organ development and the modified gene sets are highly associated with corresponding developmental processes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 20210429Hanetal.SupplementaryMaterials.docx
- 20210429SupplementaryTables.xlsx