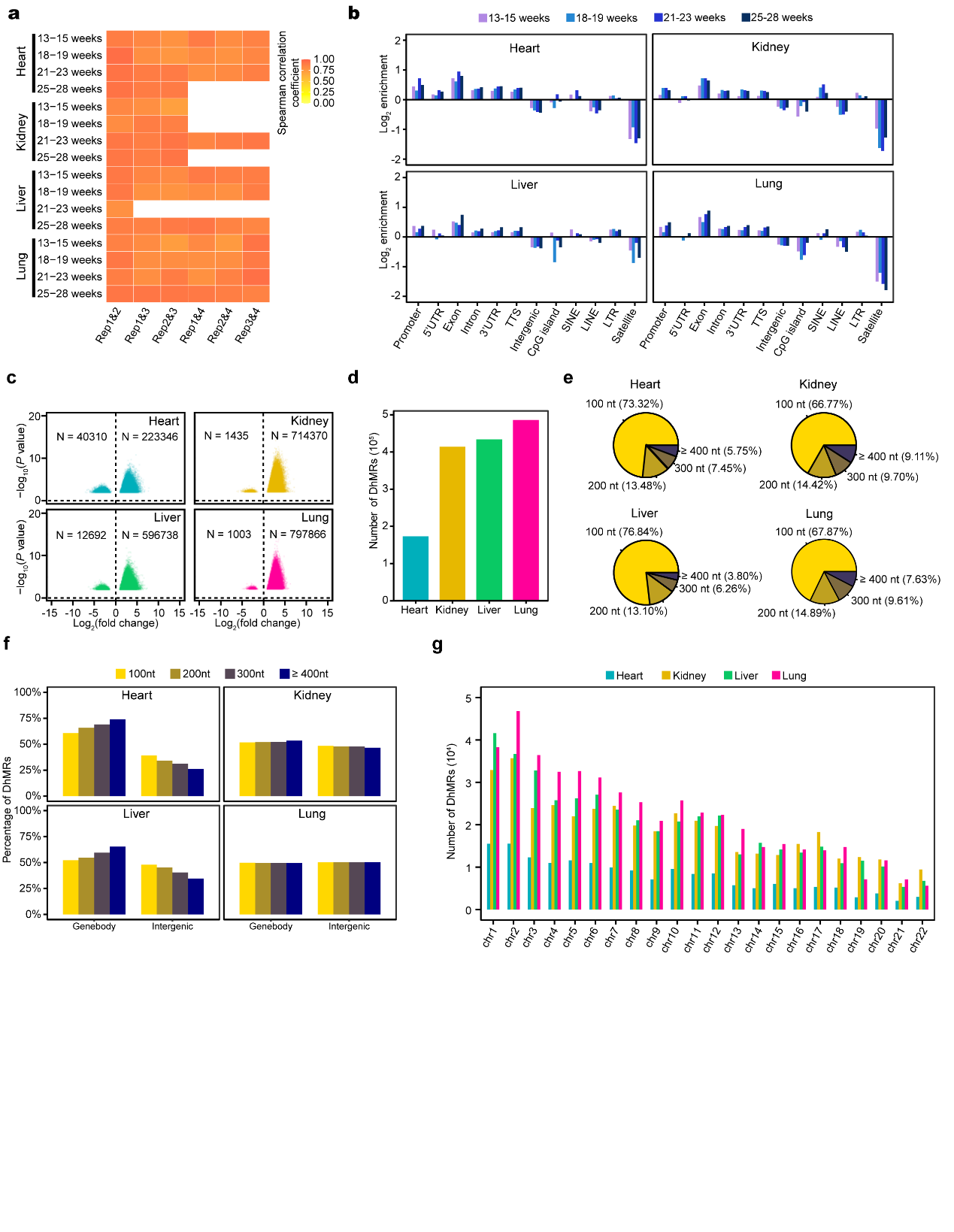
**Supplementary Information**

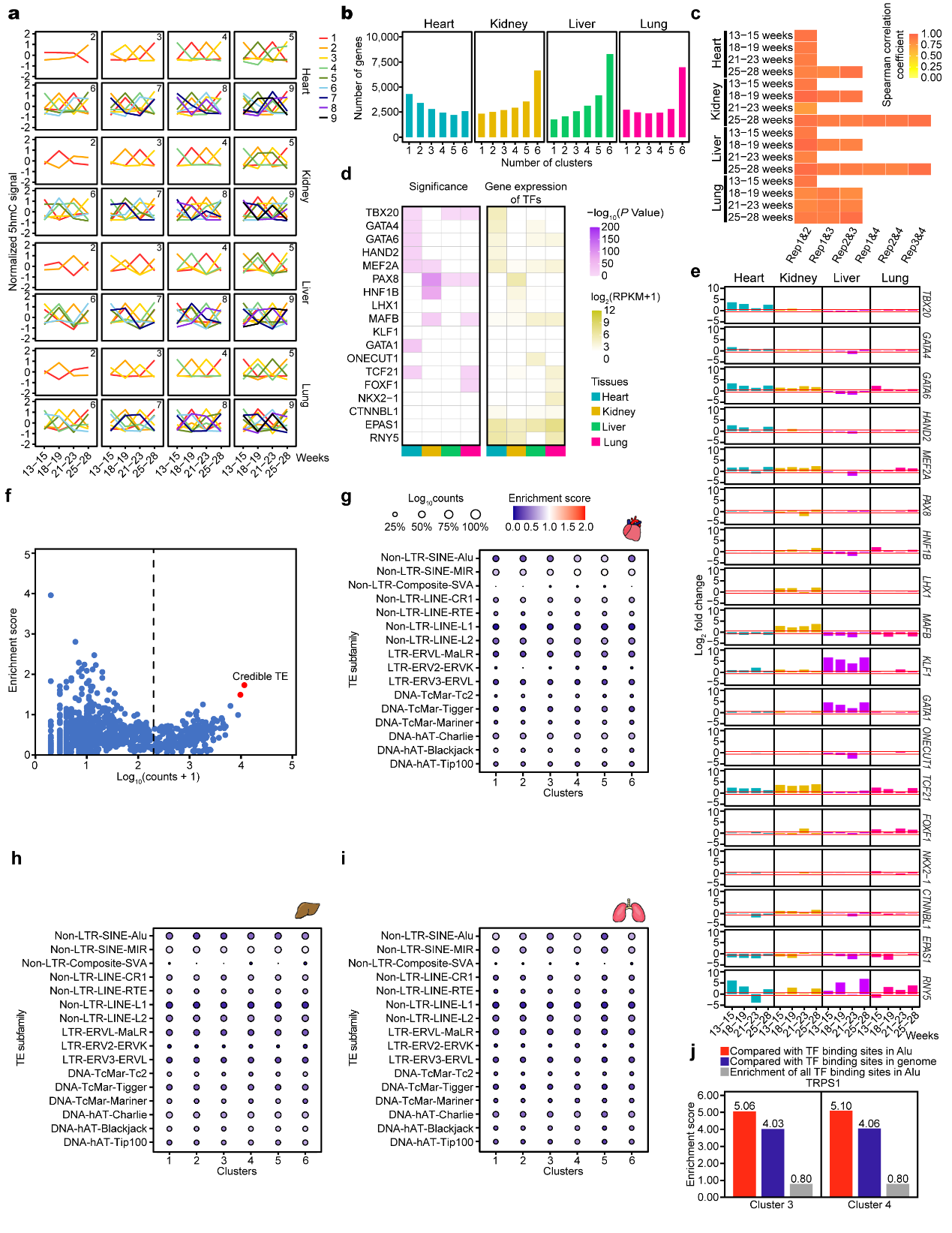
**The** **synergistic coordination of DNA 5-hydroxylmethylcytosine and RNA 5-methycytosine regulates human foetal development**

Xiao Han, Jia Guo, Mengke Wang, Nan Zhang, Jie Ren, Ying Yang, Xu Chi, Yusheng Chen, Huan Yao, Yong-Liang Zhao, Yun-Gui Yang, Yingpu Sun and Jiawei Xu

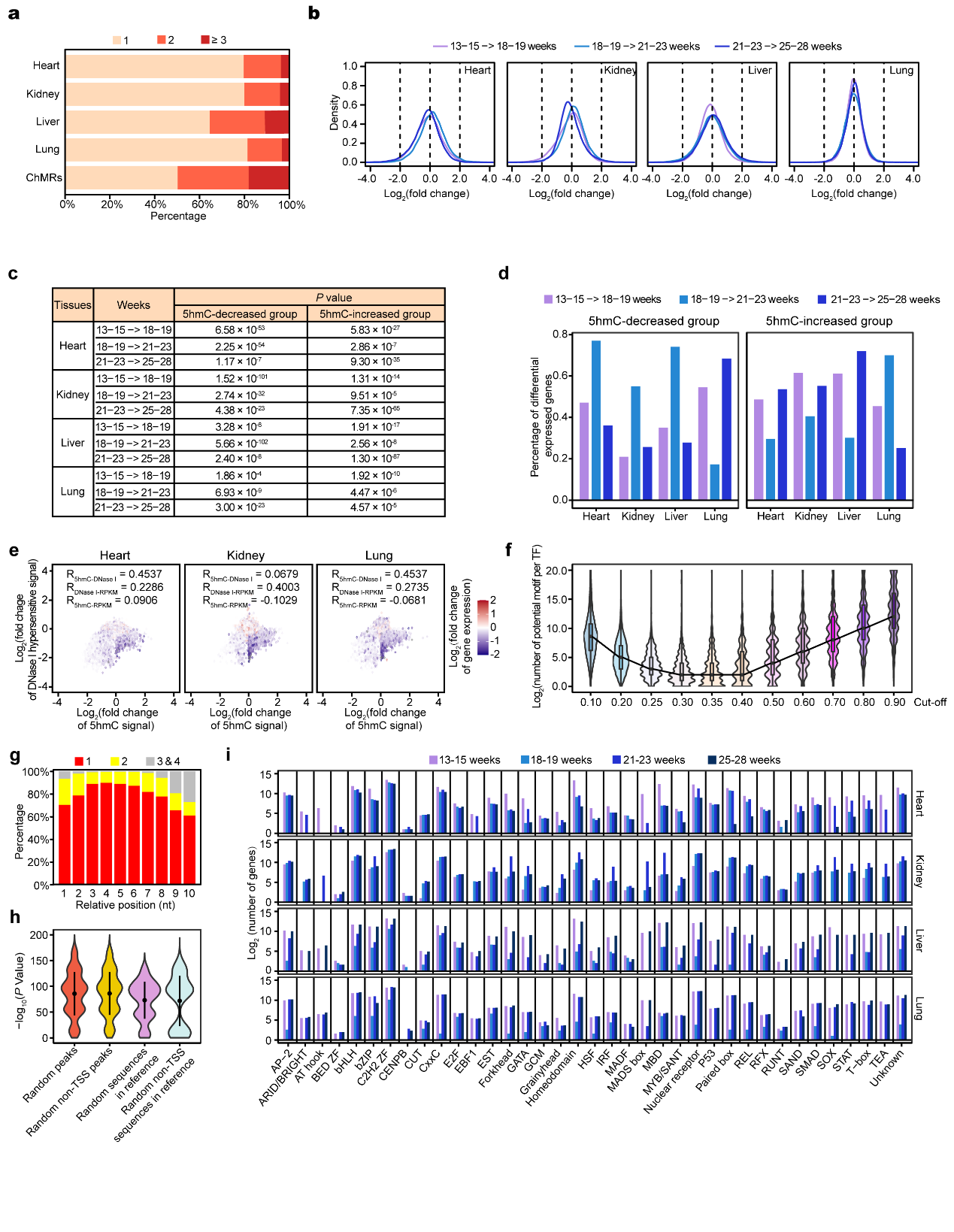
**Supplementary Figures and Legends**



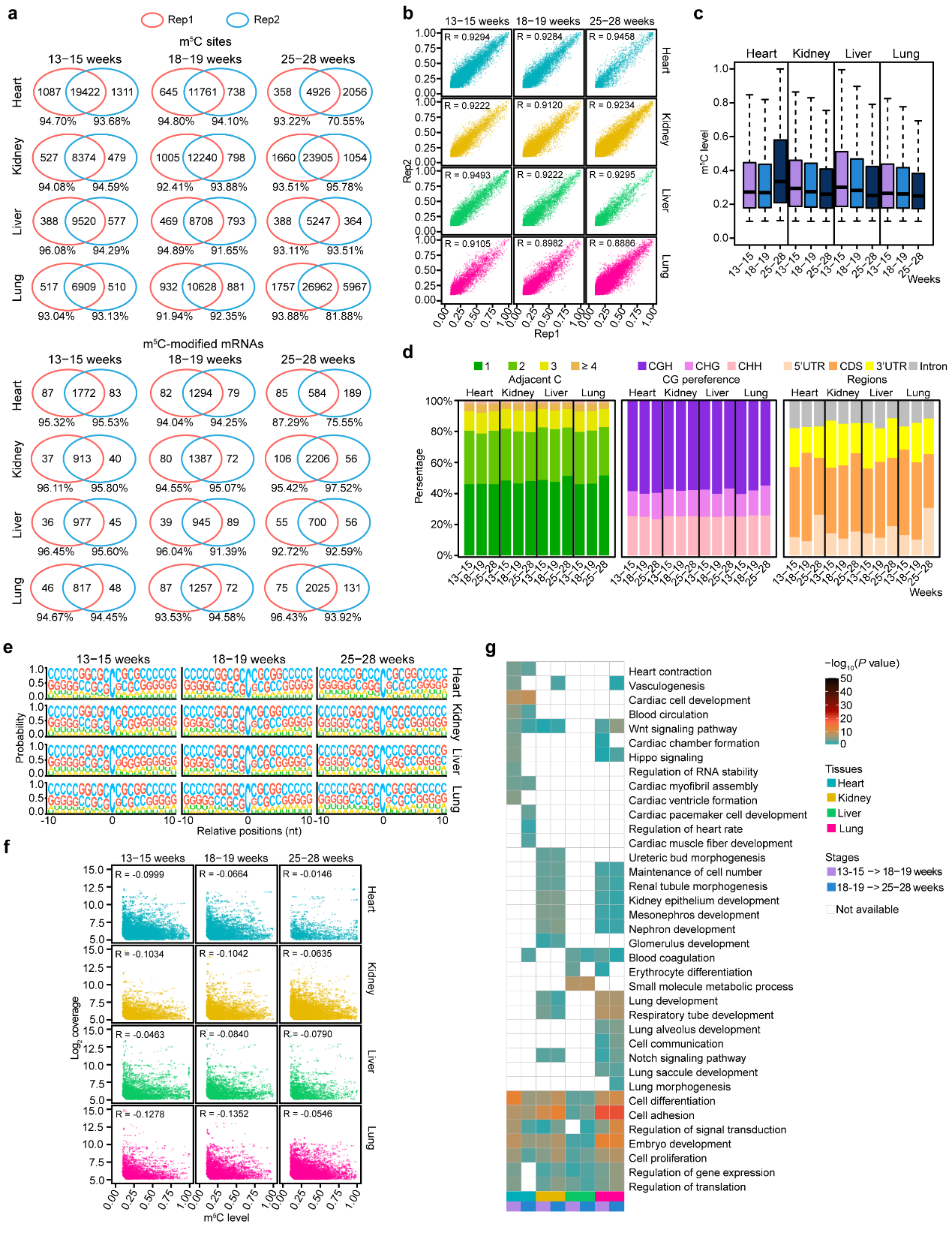
**Supplementary Fig. 1** **Distributions of differentially hydroxymethylated regions (DhMRs) in different foetal organs. a**, Heatmap showing the correlation of replicates from Nano-hmC-Seal data. **b**, Normalized enrichment score of 5hmC peaks across distinct genomic regions relative to that expected in different organs, with positive values indicating enriched more than expected. **c**, Volcano plots displaying the fold change (log2) of DhMRs calculating by ANOVA in each organ. The number of DhMRs are shown on the top. **d**, Bar plots showing the number of DhMRs in each organs. **e**, The proportion of DhMRs in length with 100 nt, 200 nt, 300 nt and over 400 nt. **f**, Bar plots showing the percentage of DhMRs with different length in gene body and intergenic regions. **e**, The numbers of DhMRs in different chromosomes.



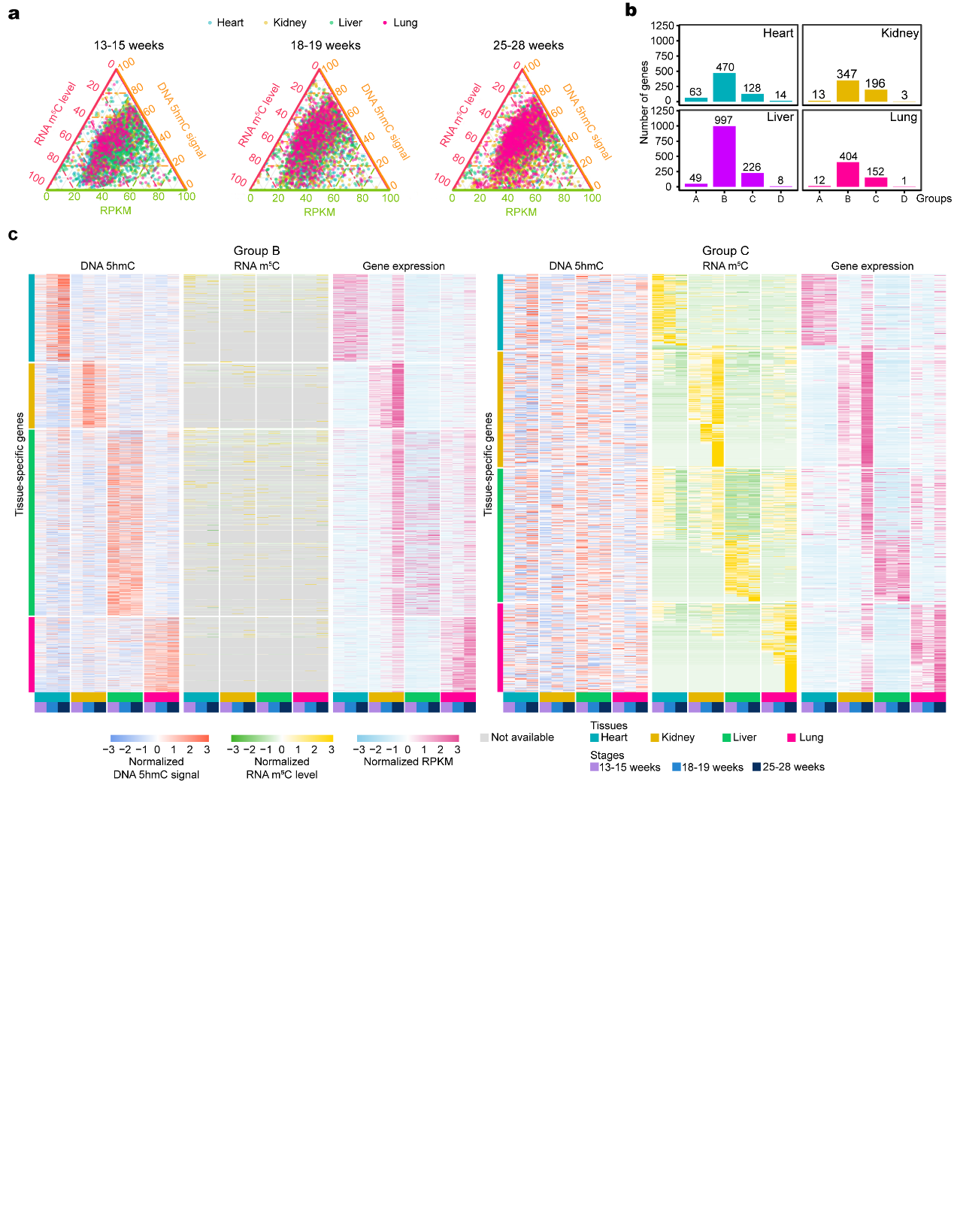
**Supplementary Fig. 2** **Transcription factors (TFs) dynamically regulate foetal development through DhMR recognition. a**, K-means clustering analysis of organ-specific DhMRs from heart, kidney, liver and lung. Each organ was separated into 2 to 9 clusters. **b**, Bar plots showing the number of genes that containing different number of clusters. **c**, Heatmap showing the correlation of replicates from RNA-Seq data. **d**, Heatmap showing the TF motifs identified by DhMRs in human adults. The TF-binding significance and the expression levels of each TF are shown from left to right. Public Nano-hmC-Seal data were from GSE144530. **e**, Bar plots showing the fold change (log2) of 18 TFs between foetal and adult organs. **f**, The scatter plot of enrichment score of each TE families and their number of intersected regions. The high enrichment scores at low count region (log10 (counts+1) < 2.3) are most possibly introduced by small TE families and random intersections. The two credible higher enrichment score ( over 1.5) are Alu elements from cluster 3 and cluster 4 in kidney (red). **g−i**, Enrichment of TE families within six identified organ specific clusters in foetal heart (**g**), liver (**h**) and lung (**i**). **j**, Bar plots showing the enrichment score of TRPS1 binding sites in Alu elements in cluster 3 and cluster 4 DhMRs compared with their overall enrichment in all Alu elements (red) and their genome-wide distribution (blue). The overall enrichment of all TRPS1 binding sites in all Alu elements compared with their genome-wide distribution is also shown (gray).



**Supplementary Fig. 3** **DNA 5hmC organ-specifically regulates gene expression during foetal organ development.** **a**, The percentage of the number of DhMRs (N = 1, N = 2 and N ≥ 3) in promoters. **b**, The density of fold change (log2) of 5hmC signal between adjacent stages on promoters. **c**, The *P* values of the significant differential 5hmC-modified genes, which calculated by Student’s t-test. **d**, The percentage of differential expressed genes (fold change ≥ 1.2) in 5hmC decreased and increased gene groups. **e**, Scatter plots showing the relationship of the fold change (log2) of 5hmC signal (x axis), DNase I hypersensitive signal (y axis) and gene expression (color). Pearson’s correlation coefficients are indicated on the top. **f**. The number of potential motifs per TF under different thresholds. The cut-offs are the ratio of a signal base on each position of 10 nt motifs (N = 0.1, 0.2, 0.3, 0.35, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9). **g.** The base possibility on each position of 10 nt motifs. **h**, The *P* values of the significance of peak enrichment compared to 4 types of backgrounds. **i**, The number of 5hmC-modified genes (log2) that each TF family recognized.



**Supplementary Fig. 4** **The features of RNA m5C in human foetal organs. a.** Venn diagrams showing the overlap of methylated sites and mRNAs between the two replicates. The overlap percentages of m5C sites and mRNAs between the biological replicates at each stage were over 91.94% and 87.29%, respectively. **b**, Scatter plots illustrating the methylation levels of the two replicates at different time points. The Pearson correlation coefficients (R) and *P* values are shown. The correlations were high (~ 0.8886) for each stage sample. **c**, Boxplots showing the overall distributions of mRNA m5C levels across different stages in foetal organs. **d**, The proportions of number of m5C sites per mRNA (N = 1, 2, 3 and ≥ 4) (left), the normalized proportions of mRNA m5C sites identified in each sequence context: CG, CHG or CHH, where H = A, C, or U, and transcriptome-wide distribution of mRNA m5C sites. The m5C numbers in CG, CHG or CHH were normalized to their individual context proportion within the transcriptome. **e**, Sequence frequency logo for the sequences proximal to mRNA m5C sites. **f**, Association between the methylation level (x axis) and the coverage (y axis). Each dot represents an individual m5C site. No significant correlation was found. **g**, GO biological processes for organ-specific m5C-modified mRNAs. The color represents the significance of each biological process.



**Supplementary Fig. 5** **DNA 5hmC and RNA m5C perform their own functions during foetal organ development.** **a**, Ternary plots showing the 5hmC signal on promoters, total m5C methylation level and expression level of each gene from different organs. **b**, Bar plots showing the number of DNA 5hmC-RNA m5C co-regulated specific-expressed genes (group A), DNA 5hmC regulated specific-expressed genes (group B), RNA m5C regulated specific expressed genes (group C) and DNA 5hmC-RNA m5C co-regulated common expressed genes (group D). **c**, Heatmaps showing the dynamics of DNA 5hmC signal on promoters, m5C methylation level, and expression level of the genes in group B (left) and group C (right). DNA 5hmC signal, m5C methylation level and gene expression level were normalized by z-score.

**Supplementary Tables:**

**Supplementary Table 1. Summary of the Nano-hmC-Seal sample information.** The 5hmC signals of each sample were obtained by Nano-hmC-Seal.

**Supplementary Table 2. Organ-stage-specific DhMRs from Nano-hmC-Seal data.** Organ-stage-specific DhMRs were identified by ANOVA (Analysis of Variance). 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 significantly differential bins were merged respectively according to their positions on the chromosome and were considered as DhMRs.

**Supplementary Table 3. Summary of the RNA-Seq sample information.** The transcriptome information of each samples was obtained by RNA-Seq.

**Supplementary Table 4. CEBPB-targeted Alu-proximal genes in foetal kidney.** Alu elements were 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.

**Supplementary Table 5. Summary of the RNA-BisSeq sample information.** The RNA m5C methylome of each sample were obtained by RNA-BisSeq.

**Supplementary Table 6. Organ-stage-specific genes from RNA-BisSeq data.** Organ-stage-specific genes were identified by ANOVA (Analysis of Variance). The *P* values were calculated using Benjamini-Hochberg method. The genes with fold change > 1.5 and *P* value < 0.05 were considered statistically significant.

**Supplementary Table 7. Gene list of different coordinate regulation groups.** Differentially expressed Organ-specific genes groups were DNA 5hmC and RNA m5C specifically regulated genes (group A), genes that were only specifically regulated by DNA 5hmC (group B), genes that were only specifically regulated by RNA m5C (group A).

**Supplementary Table 8. GO biological process categories of different coordinate regulation groups.** DAVID (version 6.8) was used to perform Gene Ontology (GO) analysis. GO terms with *P* value < 0.05 were considered as statistically significant.