Supplementary Methods and Figures

Supplementary Methods

Genomic features

Genomic regions of LADs, UCSC promoters, and CGIs were downloaded from the UCSC Genome Browser (https://genome.ucsc.edu). CGI shores (< 2 kb flanking CGIs) and CGI shelves (<2 kb flanking outward from the CGI shores) were defined based on the genomic positions of CGIs. FANTOM5 enhancers were downloaded from the FANTOM5 data server (https://fantom.gsc.riken.jp/5/data/). TADs from IMR-90 and HMEC cell lines were downloaded from the ENCODE Project (The ENCODE Project Consortium 2012) (https://www.encodeproject.org).

Validation of analytical pipeline using hg38

To validate the discovery of this study using a newer genome assembly, we reprocessed 4 WGBS samples (GSM3030954, GSM3030955, GSM3030956, GSM3030957) from adipose tissue. Raw FASTQ files were aligned to hg38 (https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/) with the same analytical pipeline. The resulting MHBs were converted to the corresponding hg19 coordinates by liftOver (Kent et al. 2002).

Summary of methods for region set enrichment analysis

There are four main enrichment techniques utilized:

(1) Locus Overlap Analysis (LOLA) to test enrichment of a query region set (e.g., tissue MHBs) against a database of regions (e.g., ChIP-seq peaks across tissues). The R package LOLA (Sheffield and Bock 2016) (version 1.22.0) was used to test the enrichment of a query region set in a database of reference region sets. More details on LOLA are available at http://code.databio.org/LOLA/.

(2) An in-house method to control for mean methylation and region size when comparing MHB and UMR/LMR enrichment. The computeCpgCov function in mHapSuite was utilized to evaluate enrichment of a query region set in open chromatin regions while controlling for region size and mean methylation levels. This method is documented at https://github.com/yoyoong/mHapSuite.

(3) GREAT analysis to perform gene set enrichment in genomic regions. More details on GREAT are available at http://great.stanford.edu/public/html/ and https://jokergoo.github.io/rGREAT/.

(4) For the enrichment of genomic features in tissue MHBs, a permutation-based method was utilized to evaluate the enrichment of a reference set within a query set.

Tools	Description	Reporting	Figure	Panel
LOLA	Test enrichment of a query region set in a database of reference region sets.	FDR	Fig. 1	F
			Fig. 2	C, E, F, H
			Fig. 3	B, E
			Fig. 5	B
			Fig. S2	D
			Fig. S5	C
			Fig. S6	
			Fig. S7	
			Fig. S8	
			Fig. S9	A, D
			Fig. S10	
In-house	Evaluate enrichment of a query region set in open chromatin regions while controlling for region size and mean methylation levels	% of CpGs	Fig. 2	A, B, G
			Fig. 3	D
			Fig. S4	A
			Fig. S5	A
			Fig. S9	C, F
GREAT Tool	Assess the enrichment of tissue-specific genes in a region set	Fold-enrichment and FDR	Fig. 1	H
			Fig. 4	E, F
			Fig. 5	J
			Fig. S13	C
Permutation -base	Evaluate the enrichment of a reference set in a query set	Fold-enrichment	Fig. S3	A

Supplementary Figures

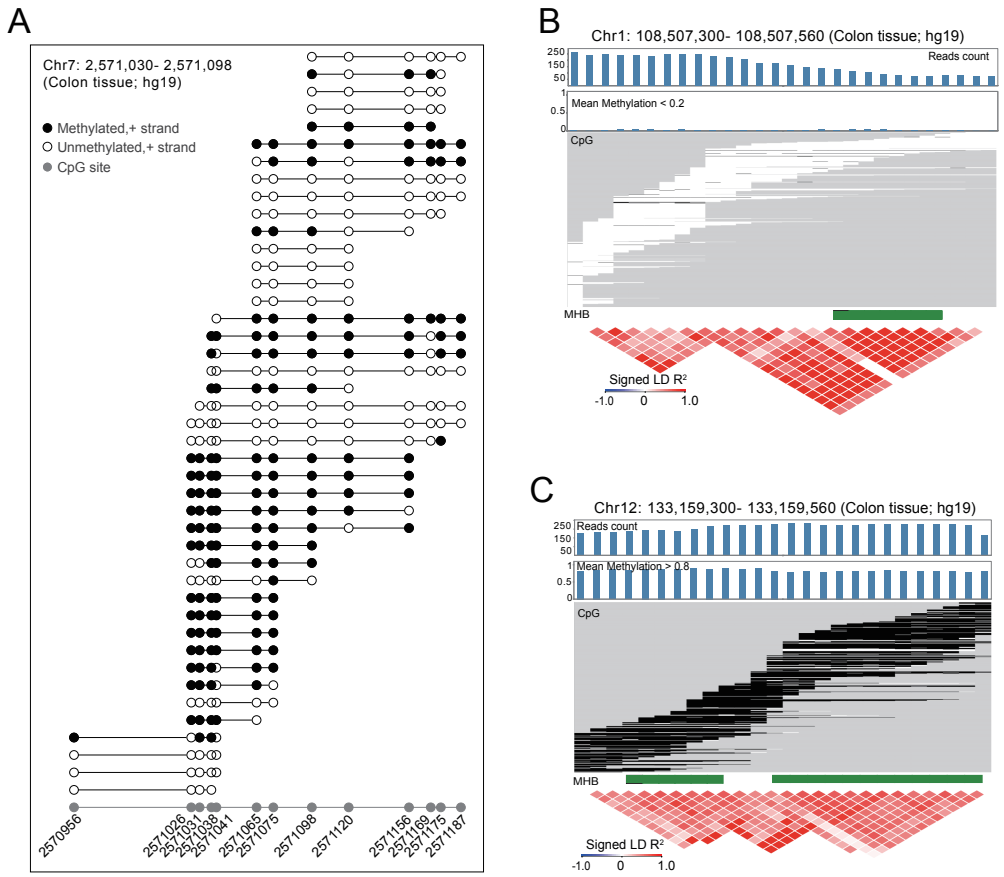

Fig. S1. Fragment-level DNA methylation patterns of MHBs. (**A**) A Tanghulu plot showing a region with MHB. (**B - C**) Two example MHB loci with low (< 0.2) and high (> 0.8) methylation, respectively. The upper part shows the coverage and mean CpG methylation of each CpG site, the middle part displays the DNA methylation status of the individual fragment, in which black and white represent methylated and unmethylated CpG sites, respectively. The bottom part shows the heatmap of signed linkage disequilibrium R^2 score between pairs of covered CpG sites.

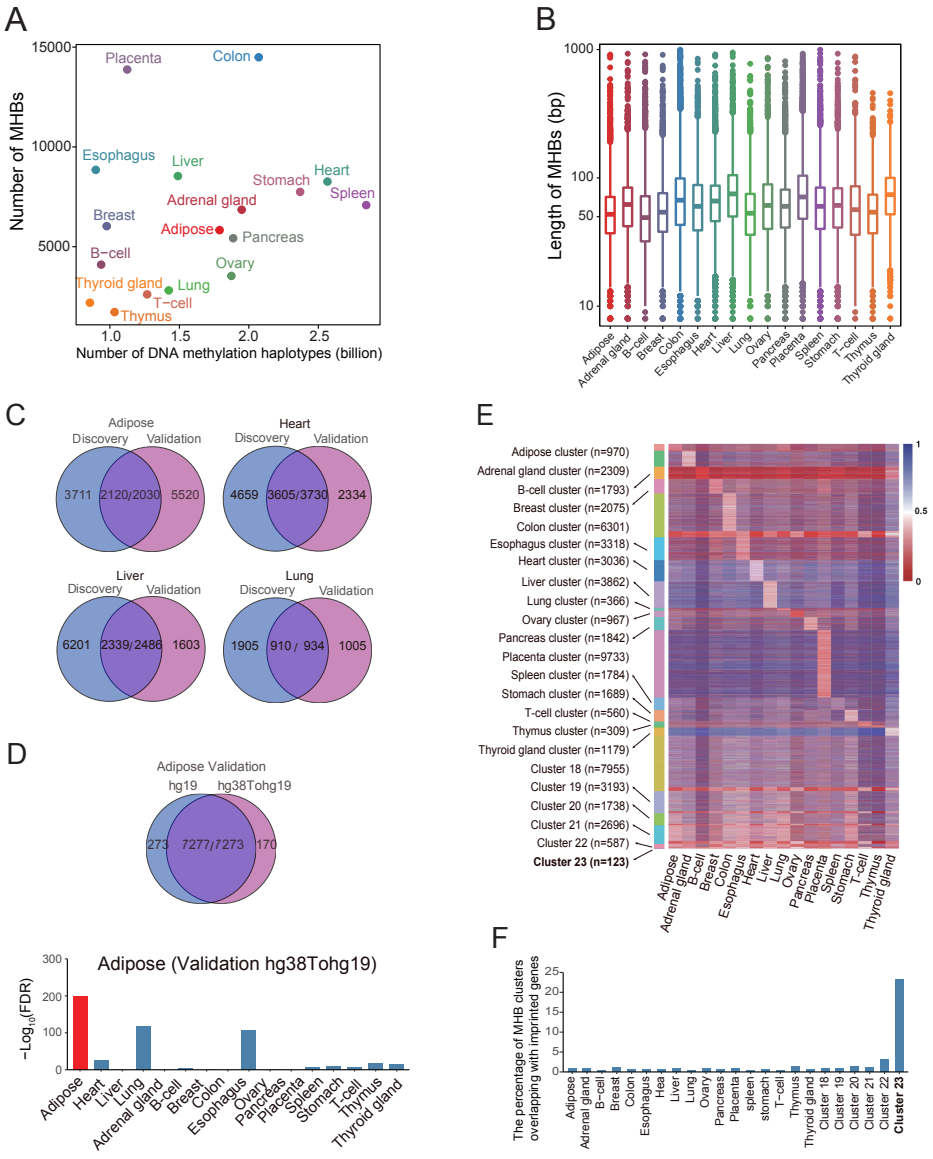

Fig. S2. Identification of MHBs in human normal tissues. (**A**) A scatter plot showing the relationship between number of MHBs and number of DNA methylation haplotypes. (**B**) Box plots showing the distribution of MHB lengths. (**C**) Pie charts illustrating the count of shared MHBs within the discovery and validation cohorts. (**D**) Validation of MHBs in adipose tissue using the genome assembly hg38. Adipose validation samples were reprocessed using hg38 and the resulting MHBs were converted to hg19 for comparison (hg38Tohg19). Upper panel shows high concordance (>96% shared MHBs) between hg19 and hg38 results. Lower panel shows enrichment of adipose MHBs from the validation set in the discovery set. Enrichment was determined by the R package LOLA, using the union of MHBs from all tissue types as the background. To preclude computing the logarithm of zero, FDR values of zero were converted to 1 x 10^{-300} prior to logarithmic transformation. (**E**) Mean methylation levels of MHB clusters. (**F**) The percentage of MHBs annotated to imprinted genes for each MHB cluster.

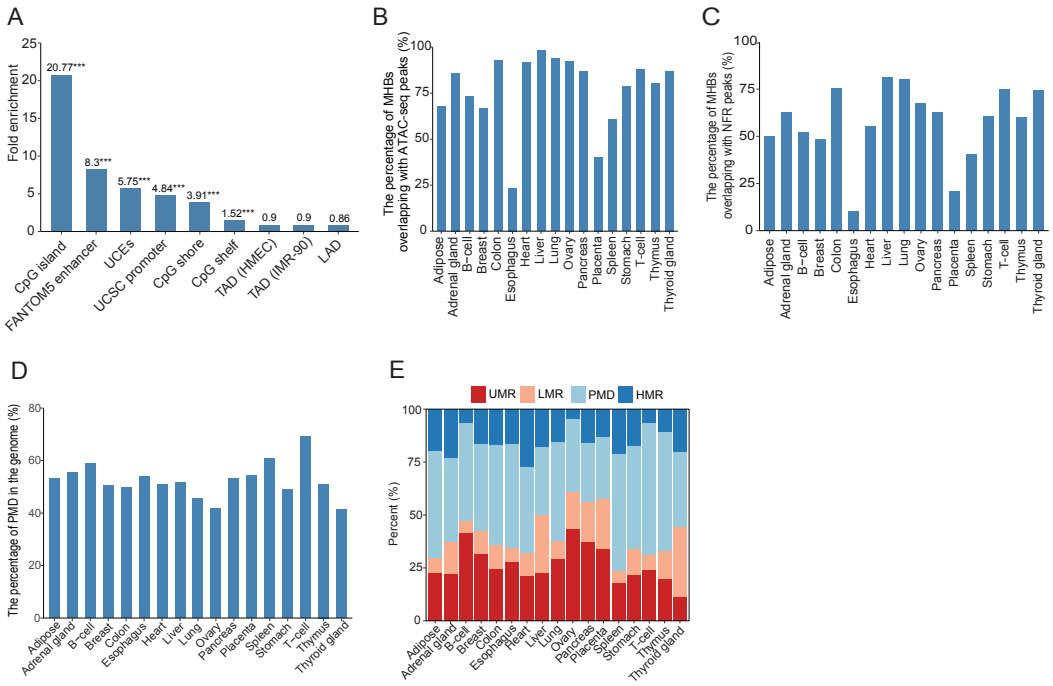

Fig. **S3**. **Distribution of MHBs in open chromatin regions.** (**A**) Enrichment of MHBs within genomic features. Enrichment scores were calculated as the ratio of observed to random overlapping counts between MHBs and genomic regions. Statistical significance was determined via a permutation test (1000 iterations). ***, $P < 0.001$. (**B**) Overlapping of tissue MHBs with open chromatin regions in matched tissue types. (**C**) Overlapping of tissue MHBs with NFR peaks in matched tissue types. (**D**) The percentage of PMDs in 17 normal tissue types. (**E**) Annotation of MHBs to DNA methylation states, including UMR, LMR, PMD, and HMR.

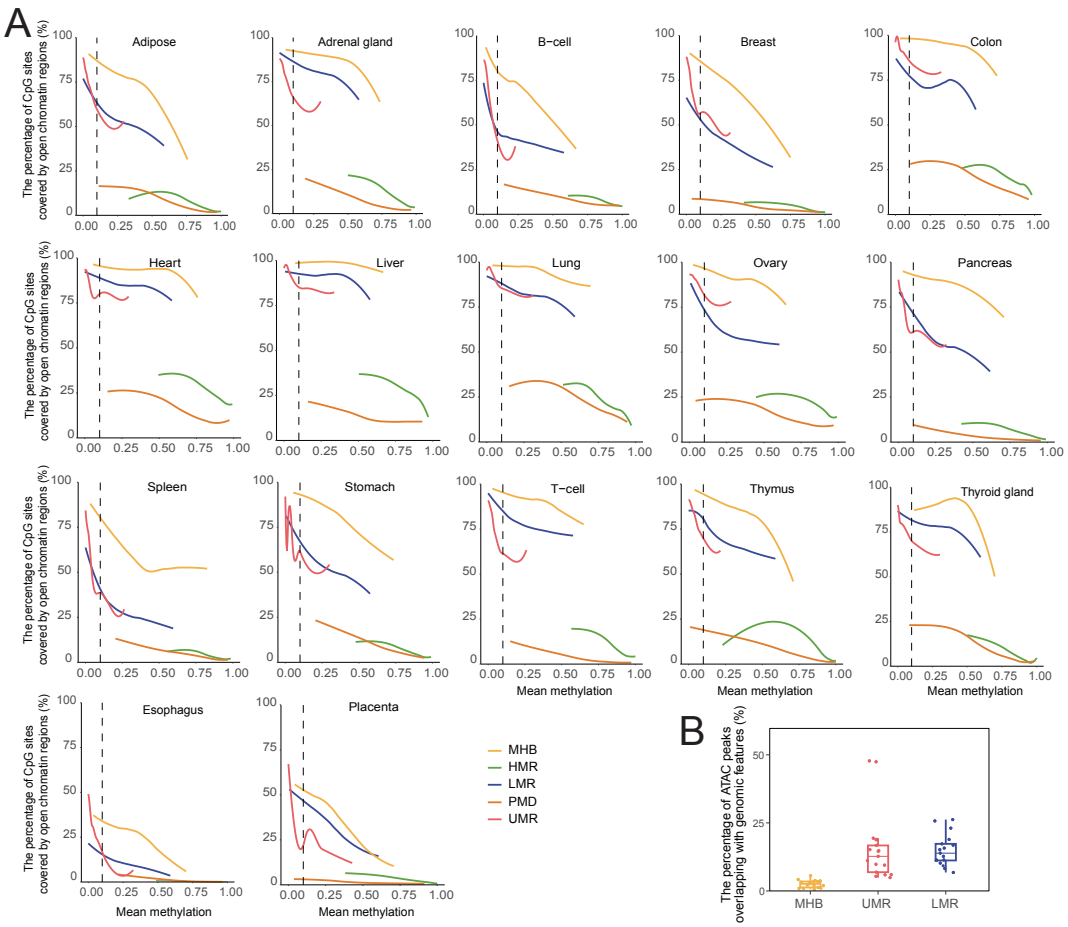

Fig. S4. Enrichment of MHBs in open chromatin regions. (**A**) Enrichment of MHBs in open chromatin regions while control for mean methylation levels and region size. The enrichment score is computed as the percentage of CpG sites covered by open chromatin regions. (**B**) Mapping open chromatin regions to MHBs, UMRs, and LMRs.

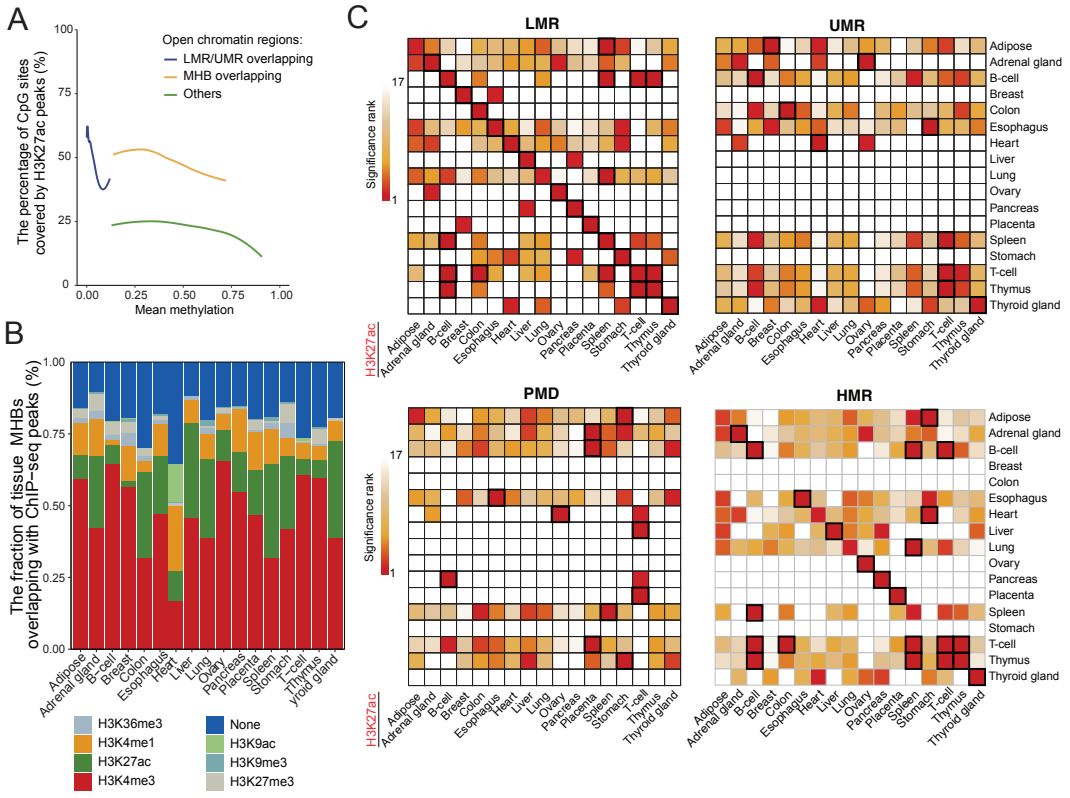

Fig. S5. Enrichment of MHBs in histone modification regions. (**A**) Enrichment of H3K27ac histone modification in open chromatin regions. Open chromatin regions were categorized based on overlap with UMRs/LMRs, MHBs, or neither. The enrichment score is calculated as the percentage of CpG sites covered by H3K27ac ChIP-seq peaks. (**B**) Annotation of MHBs to regions characterized by various histone modifications. (**C**) Enrichment of LMRs, UMRs, PMDs, and HMRs in H3K27ac peaks. The enrichment analysis was conducted using the R package LOLA with an FDR threshold of 5%. The resulting *P*-values for significant enrichments were ranked across all tissue types, with the highest enrichment highlighted in a black frame. Non-significant results (FDR > 5%) are displayed in white.

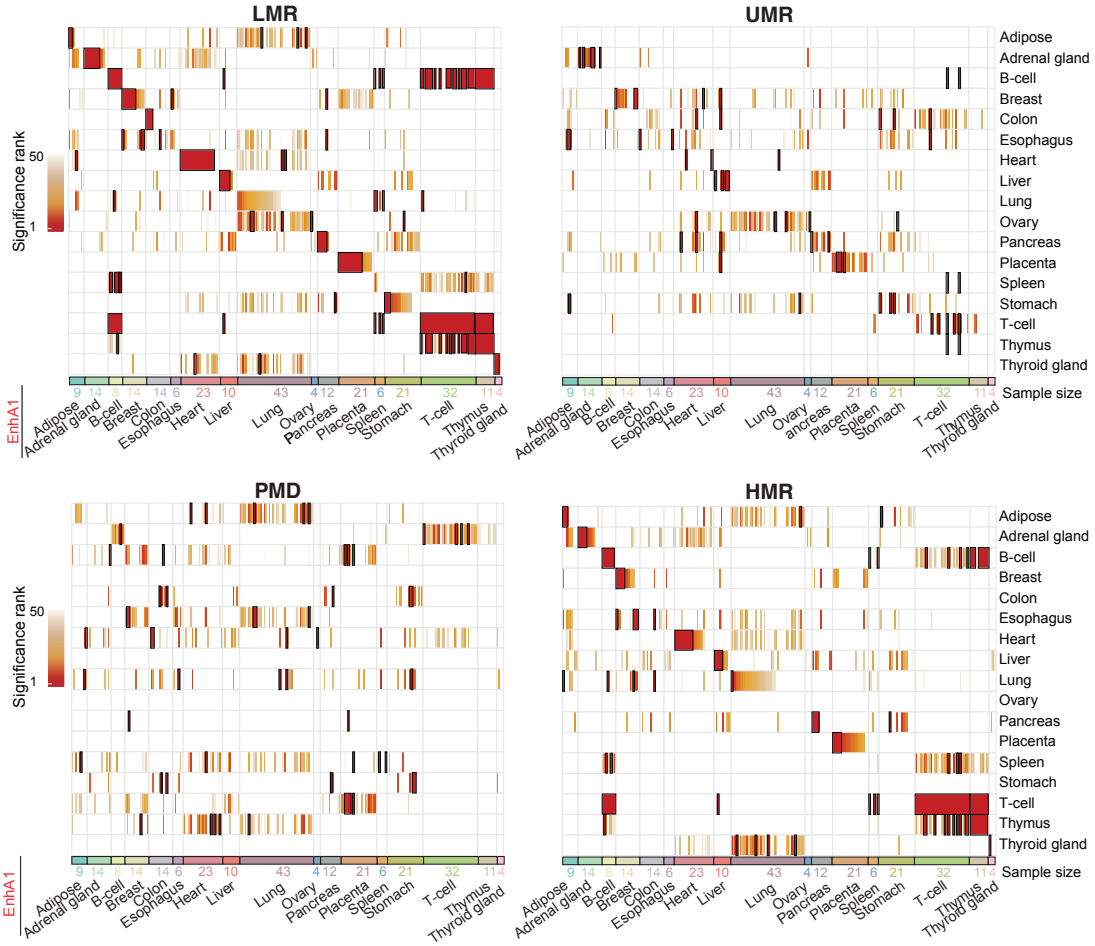

Fig. S6. Enrichment of LMRs, UMRs, PMDs, and HMRs in EnhA1 enhancers. Enrichment analysis was conducted using the R package LOLA. The resulting *P*-values for significant enrichment were ranked across all tissue types (FDR < 5%). The top five tissues with the highest enrichment are outlined in black. Non-significant results (FDR > 5%) are displayed in white.

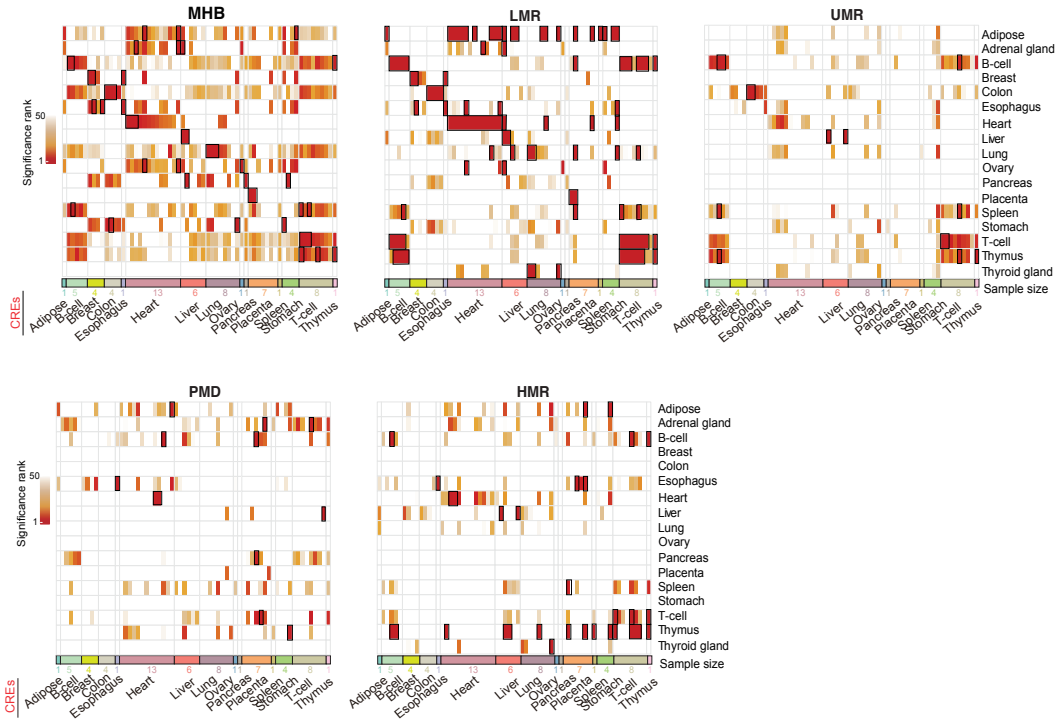

Fig. S7. Enrichment of MHBs, LMRs, UMRs, PMDs, and HMRs in *cis*-regulatory elements (CREs). Enrichment analysis was conducted using the R package LOLA. The resulting *P*-values for significant enrichment were ranked across all tissue types (FDR < 5%). The top three tissues with the highest enrichment are highlighted with black rectangles. Non-significant results (FDR > 5%) are presented in white.

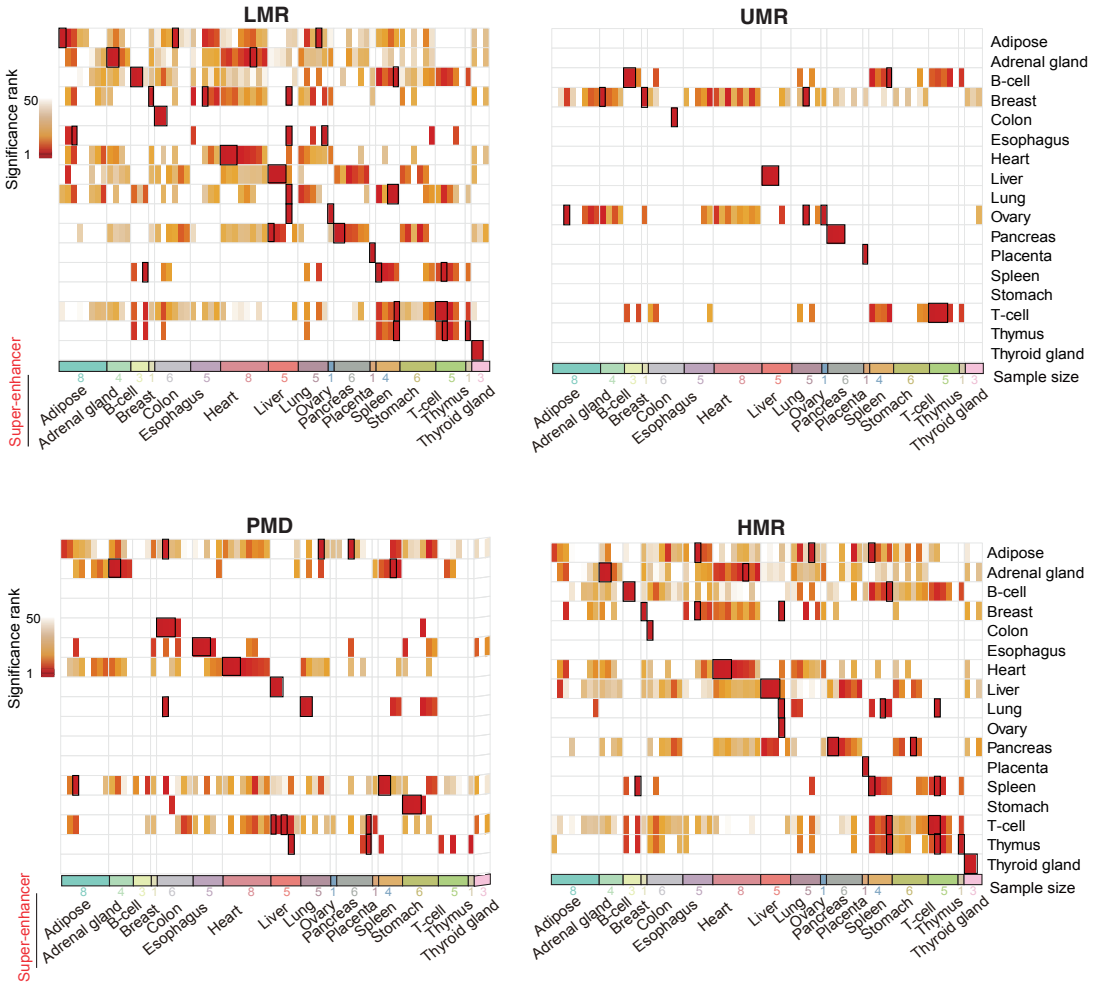

Fig. S8. Enrichment of MHBs, LMRs, UMRs, PMDs, and HMRs in super-enhancers. Enrichment analysis was conducted using the R package LOLA. The resulting *P*-values for significant enrichment were ranked across all tissue types (FDR < 5%). The top three tissues with the highest enrichment are highlighted in a black frame. Non-significant results (FDR > 5%) are presented in white.

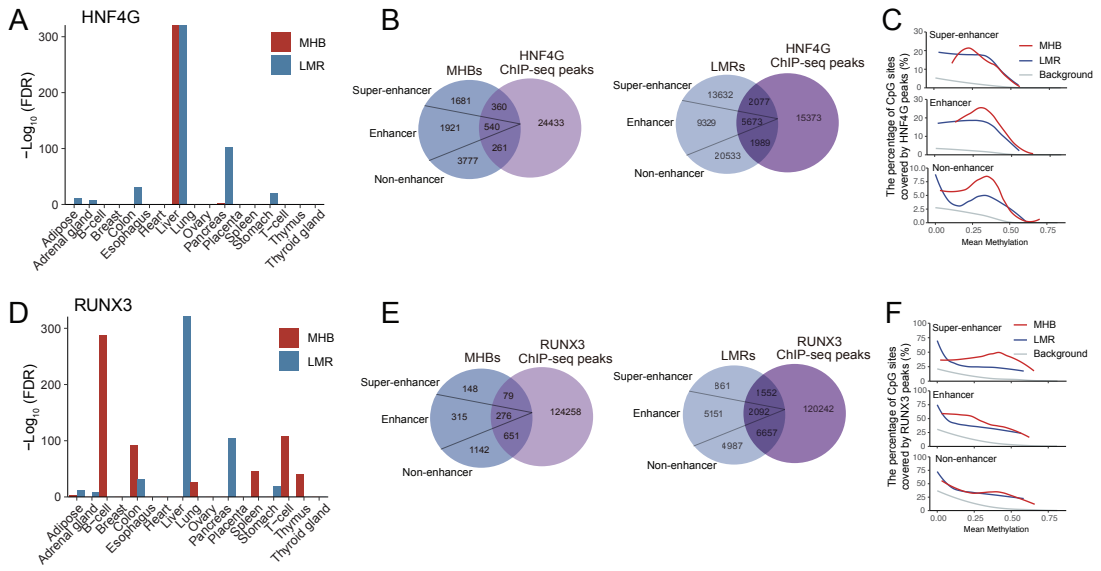

Fig. S9. Enrichment of MHBs and LMRs in regions of ChIP-seq peaks. (**A**) Enrichment of MHBs and LMRs in liver HNF4G peaks. Enrichment test was performed using R package LOLA. (**B**) Venn diagrams showing the overlap between MHBs or LMRs and HNF4G ChIP-seq peaks in liver. Liver MHBs and LMRs were categorized into three groups based on overlap with super-enhancers and enhancers (EnhA1). (**C**) Enrichment of MHBs and LMRs in liver HNF4G peaks. Enrichment scores were defined as percentage of CpG sites covered by HNF4G peaks. Enrichment analysis was conducted using the computeCpgCov function in mHapSuite. (**D**) Enrichment of T-cell MHBs and T-cell LMRs in B-cell RUNX3 peaks. (**E**) Venn diagrams showing the overlap between MHBs or LMRs and RUNX3 ChIP-seq peaks in T-cells. T-cell MHBs and LMRs were categorized into three groups based on overlap with super-enhancers and regular enhancers (EnhA1). (**F**) Enrichment of T-cell MHBs and T-cell LMRs in B-cell RUNX3 ChIP-seq peaks. Enrichment scores were defined as the percentage of CpG sites covered by RUNX3 ChIP-seq peaks. Enrichment analysis was conducted using the computeCpgCov function in mHapSuite.

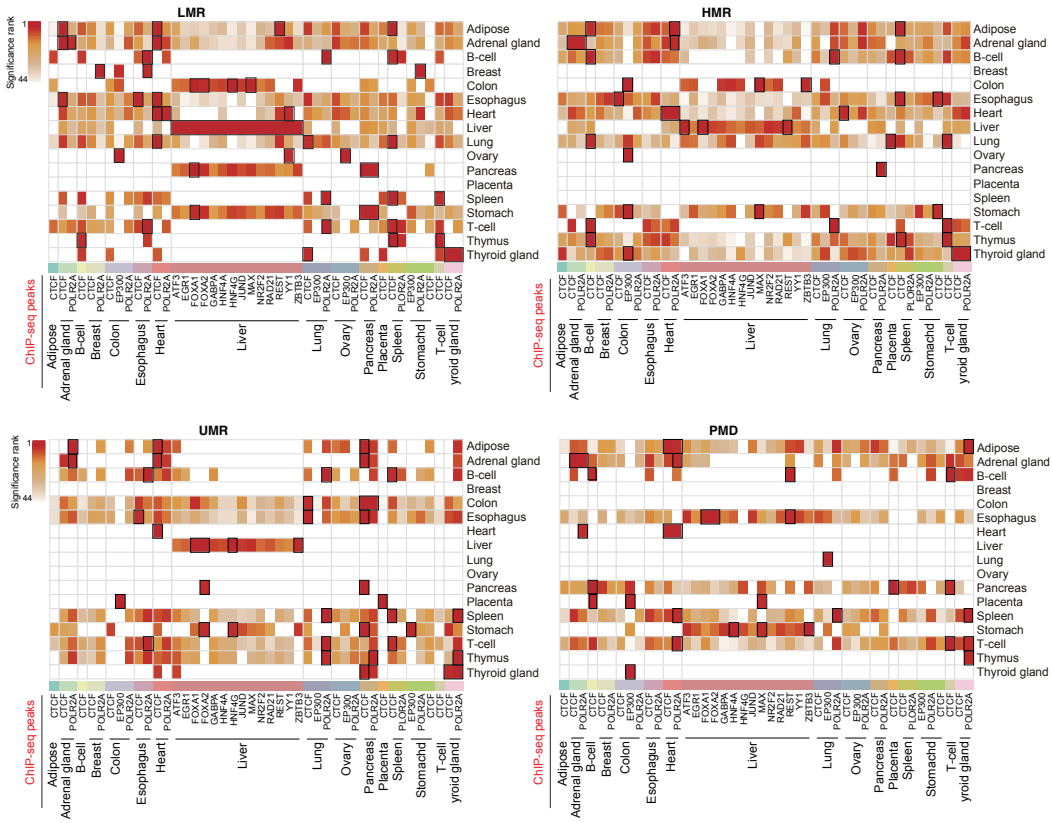

Fig. S10. Enrichment of ChIP-seq peaks in UMRs, LMRs, PMDs and HMRs. Enrichment tests were conducted using the R package LOLA, with the union of MHBs as the background for testing. Significant enrichments (FDR < 5%) were ranked across all tissue types, and the top three are highlighted by black rectangles.

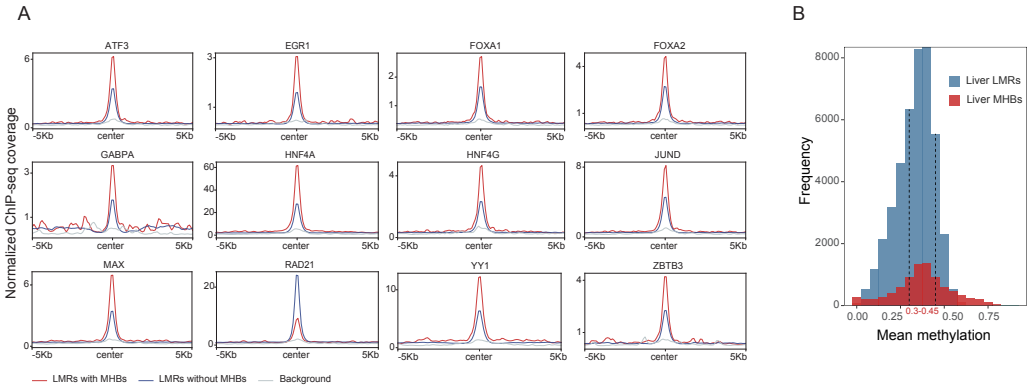

Fig. **S11**. **ChIP-seq signal around center of LMRs with MHBs in liver tissue**. (**A**) ChIP-seq signals of transcription factors within liver LMRs, categorized by the presence (red) or absence (blue) of MHBs. All regions were included. (**B**) Histograms showing distribution of mean methylation levels in MHBs and LMRs.

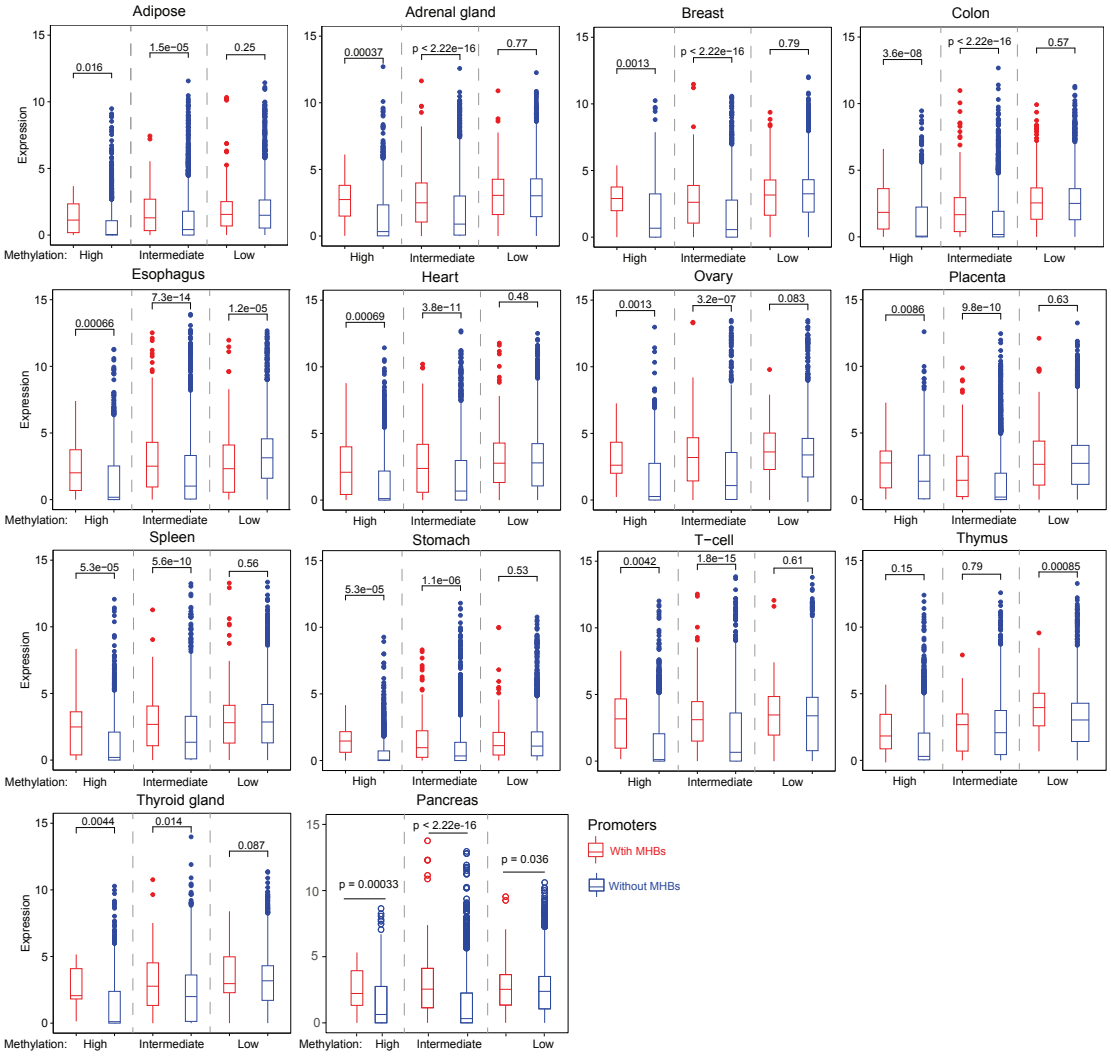

Fig. S12. MHBs are associated with gene expression. Promoters were categorized into low- (<0.2), intermediate- (0.2-0.8), and high-methylation (>0.8) groups. Expression was compared between promoters with and without MHBs within each group. Statistical significance was assessed by two-sided t-test.

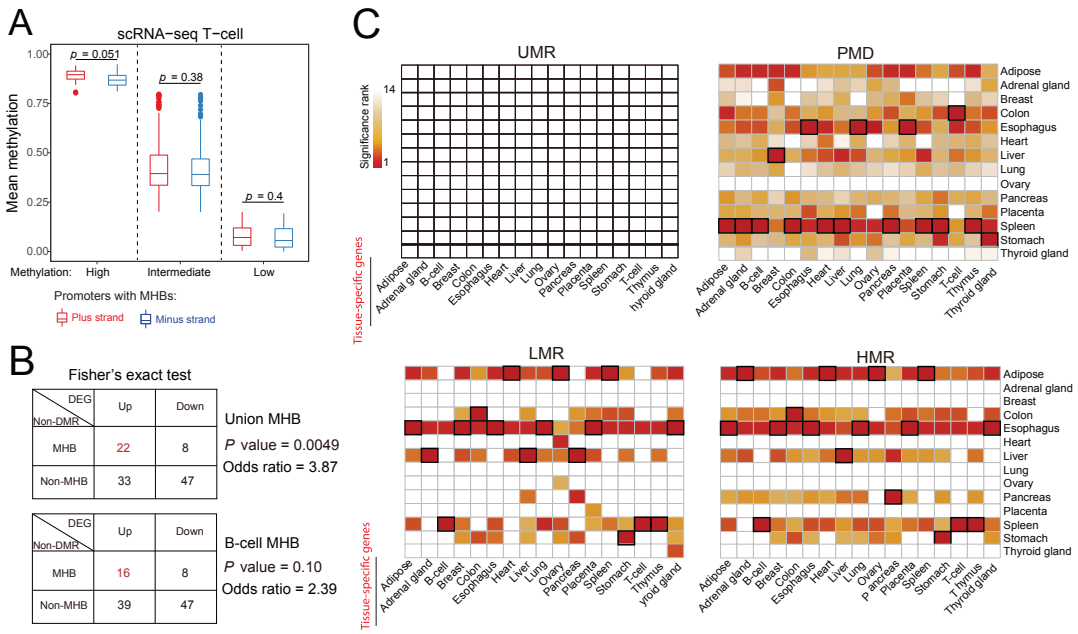

Fig. S13. Enrichment of DNA methylation states in tissue-specific genes. (**A**) Distribution of mean methylation in MHBs, stratified by forward and reverse strands. (**B**) Association of MHBs and differential gene expression. DMRs in T cells and B cells were identified using Metilene (FDR < 5%, Δ beta > 0.1). Among non-DMR genes with almost identical mean methylation in promoters (Δbeta < 0.01), 110 genes were differentially expressed. The 2x2 contingency tables categorizing genes by differential expression and MHB status in T-cells, B-cells, or union of two sets. Statistical significance was evaluated using Fisher's exact test. (**C**) Enrichment analysis was conducted using the R package rGREAT. Adjusted *P*-values were ranked across all tissue types, with the top-ranked tissue of highest significance highlighted by black rectangles.

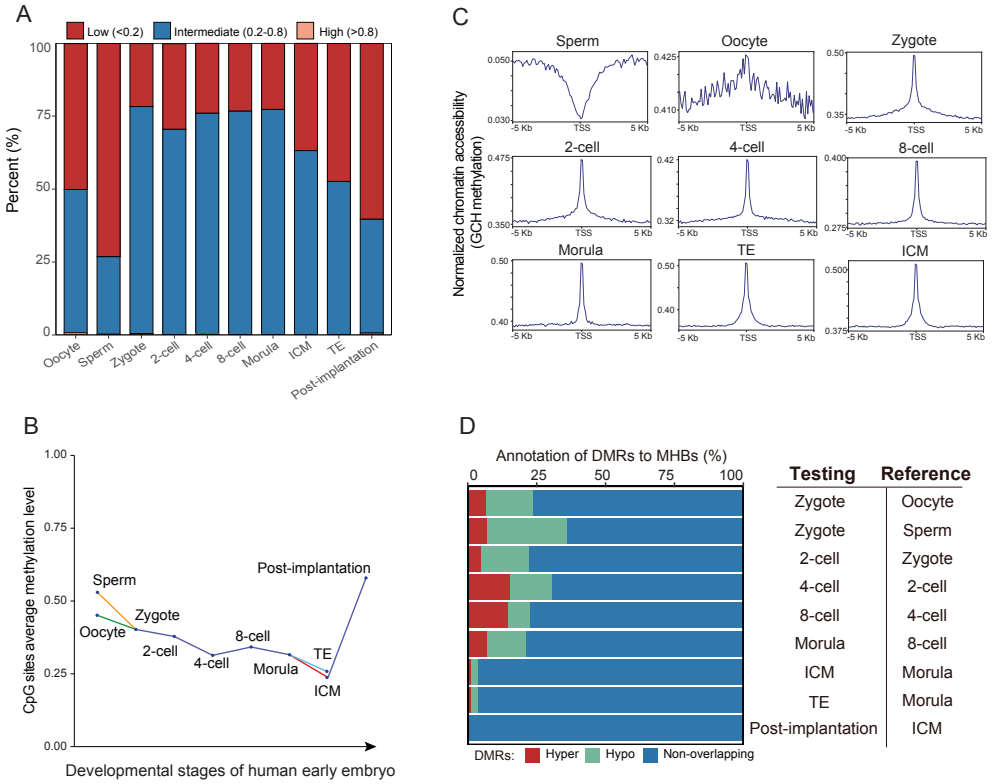

Fig. S14. Identification of MHBs in human early embryos. (**A**) Distribution of mean methylation levels in MHBs. (**B**) DNA Methylation landscape during human embryonic development. (**C**) Normalized chromatin accessibility signal around the TSSs (±5 kb). (**D**) Annotation of DMRs to MHBs in the matched stages.