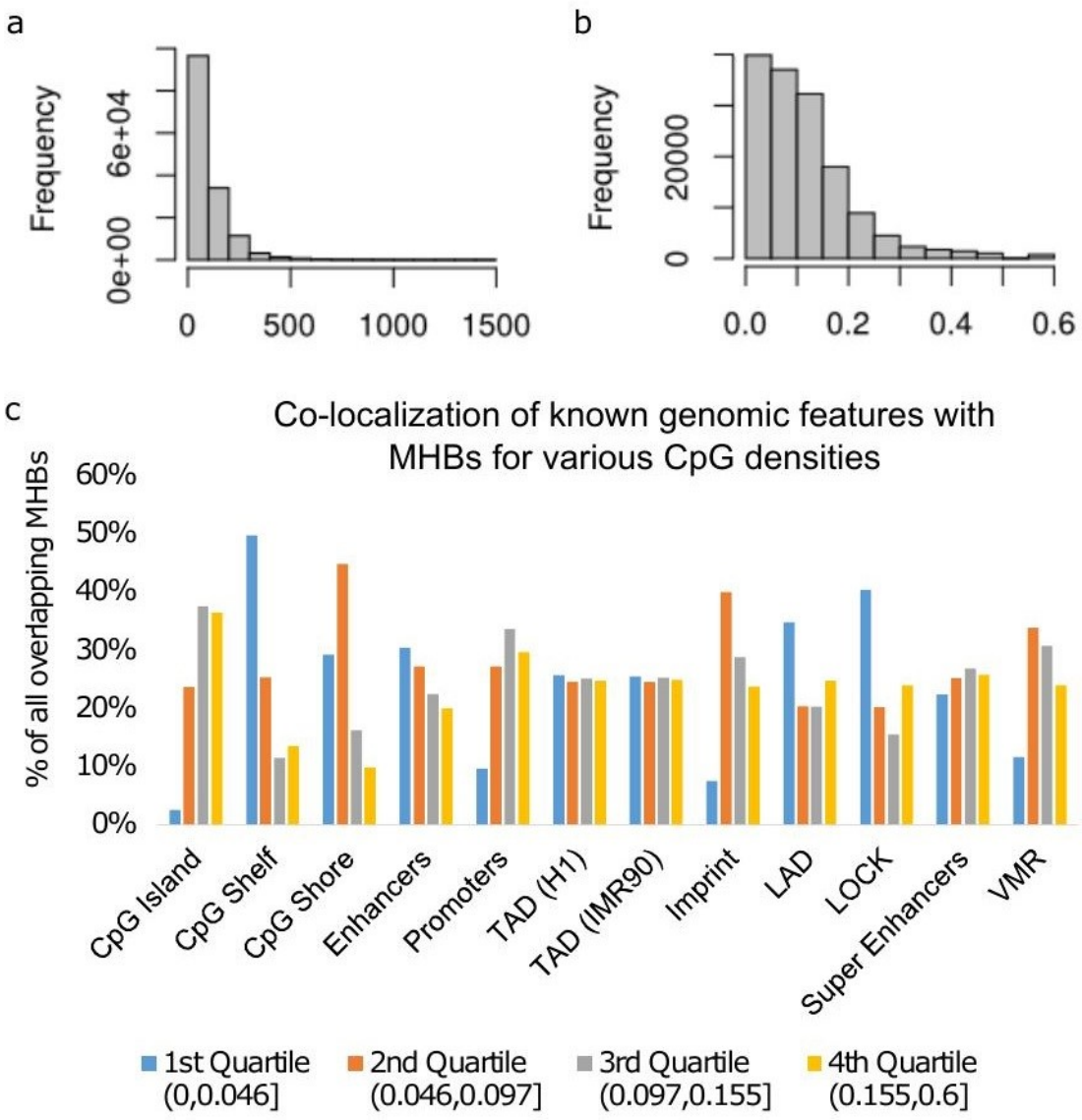
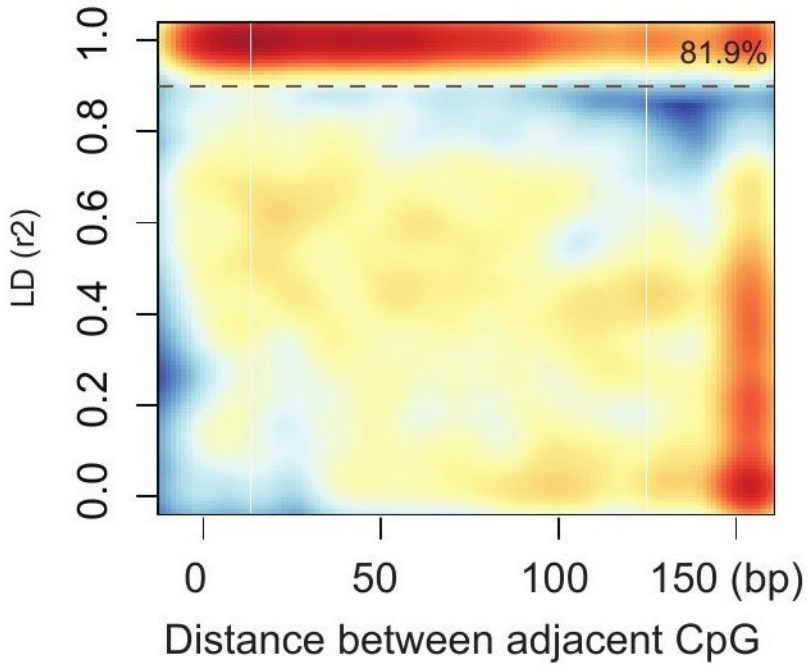


Supplementary Figure 1

Characteristics of MHBs in human genome

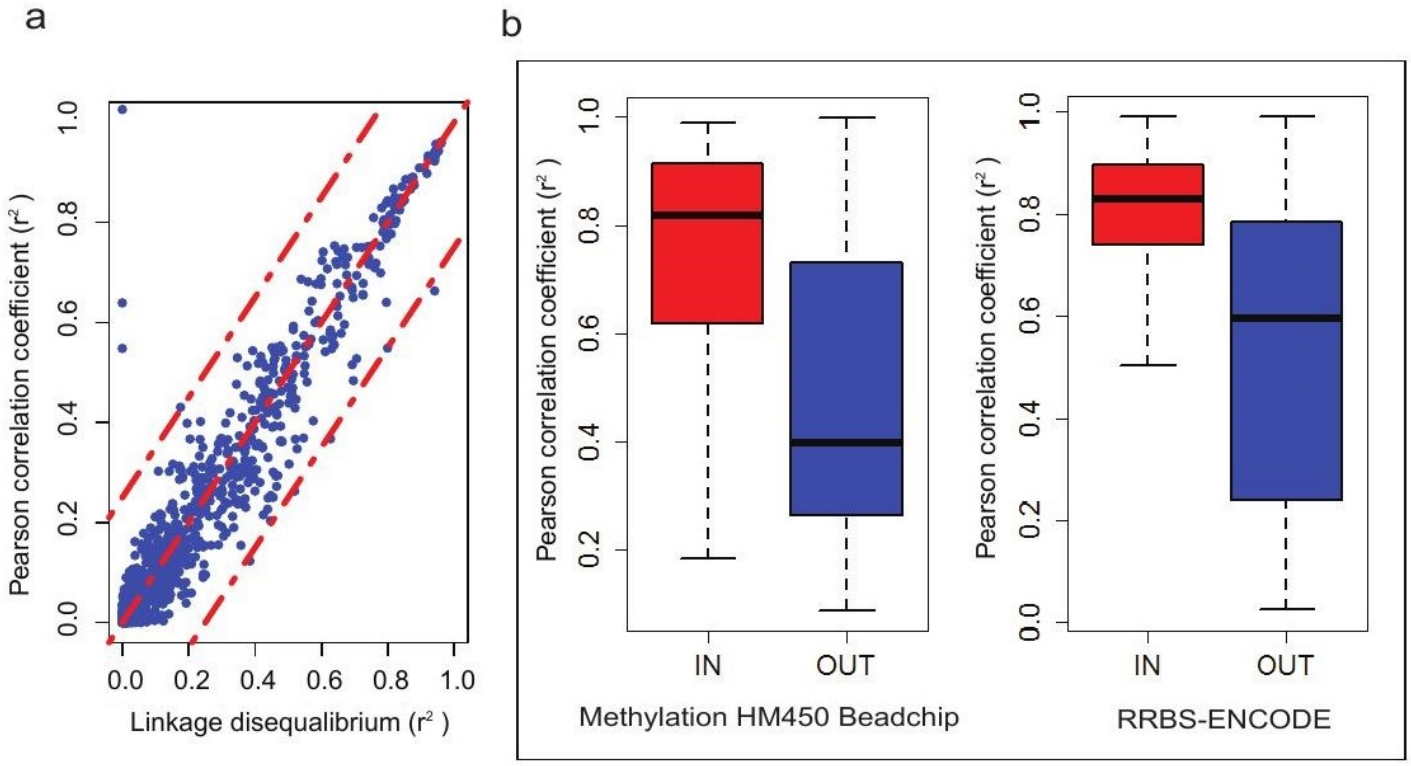
(a) Distribution of MHB sizes. (b) Distribution of MHBs CpG densities (CpGs/bp). (c) Co-localization of known genomic features broken down by CpG density. We split all MHBs into quartiles where each quartile is as follows: (0, 0.046], (0.046, 0.097], (0.097, 0.155], (0.155, 6]. Note that closed brackets are inclusive. The 1st quartile (MHBs with the lowest CpG densities) are mostly in CGI shelf or shore, and are enriched for LAD, LOCK and enhancers.



Supplementary Figure 2

Loss of CpG linkage disequilibrium replicated in two additional kidney cancer samples.

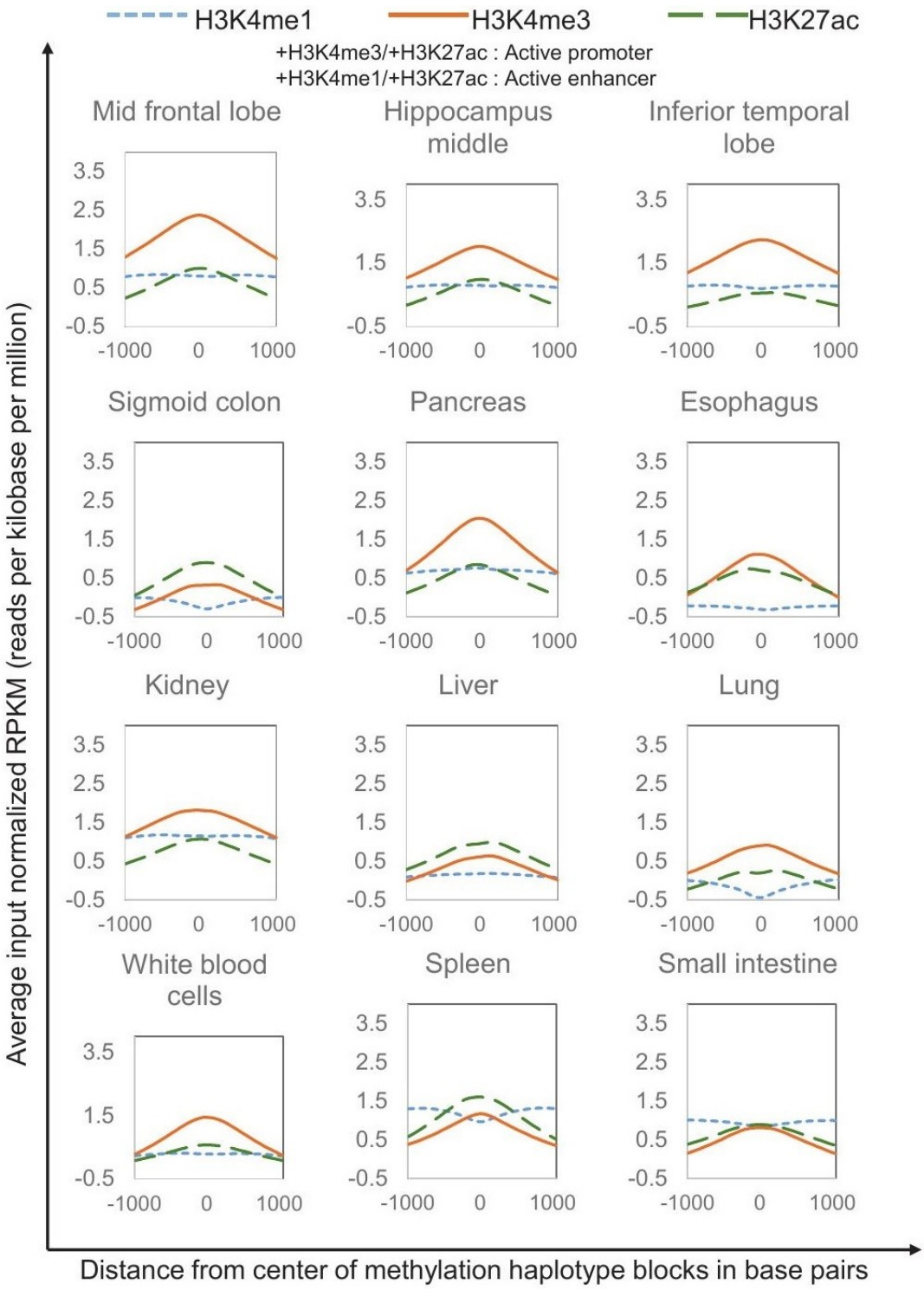
Two kidney cancer WGBS data were downloaded from NCBI GEO (GSE63183), and processed with the same computational procedures.



Supplementary Figure 3

Validation of MHBs with TCGA Methylation HM450K beadchip and ENCODE RRBS data

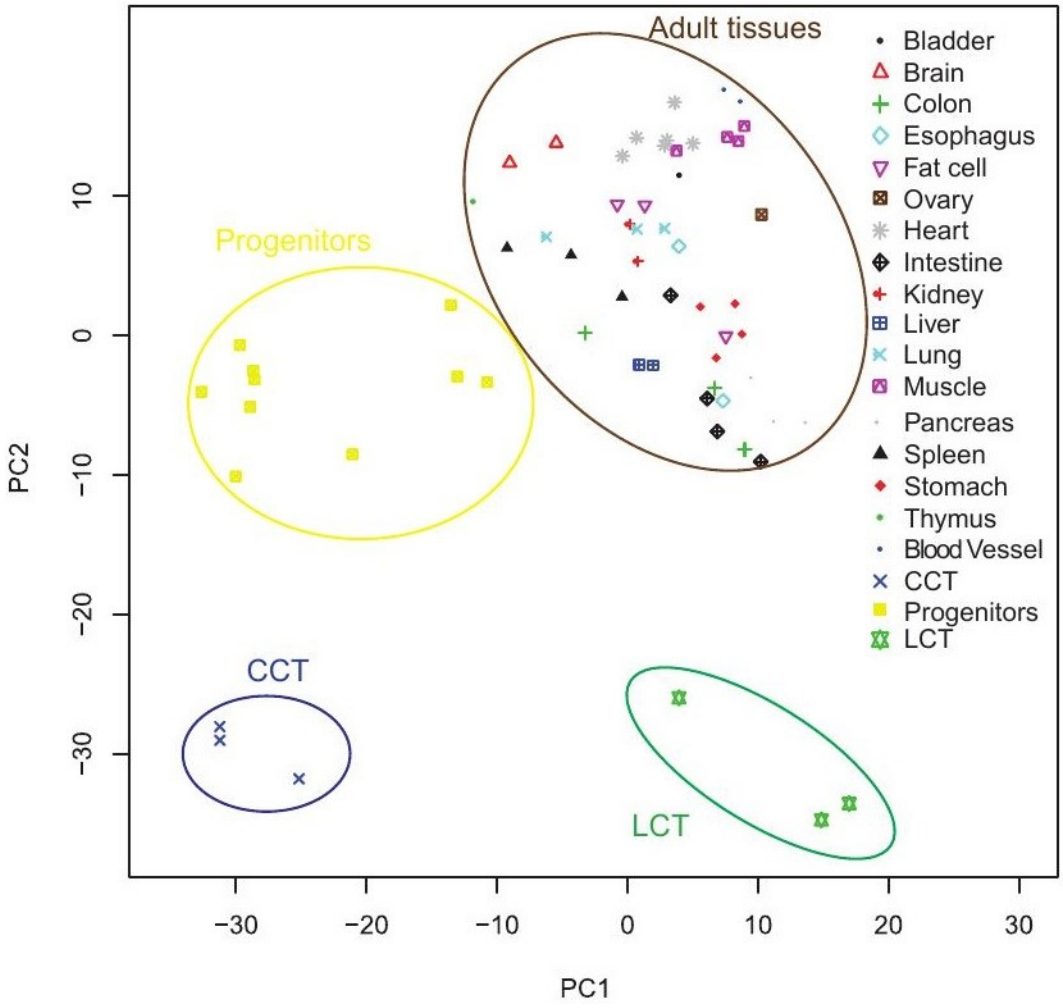
(a) Squared Pearson correlation coefficient r^2 versus LD r^2. (b) The Pearson correlation coefficient for CpGs in RRBS and HM450K data were significantly higher in regions overlapping with MHBs compared with the CpGs without overlapping with MHBs. IN denotes RRBS or HM450K CpGs within MHBs. OUT denotes RRBS or HM450K regions beyond MHBs.



Supplementary Figure 4

Profiles of H3K27ac, H3K4me3 and H3K4me1 over methylation haplotype blocks for 12 human adult tissue types.

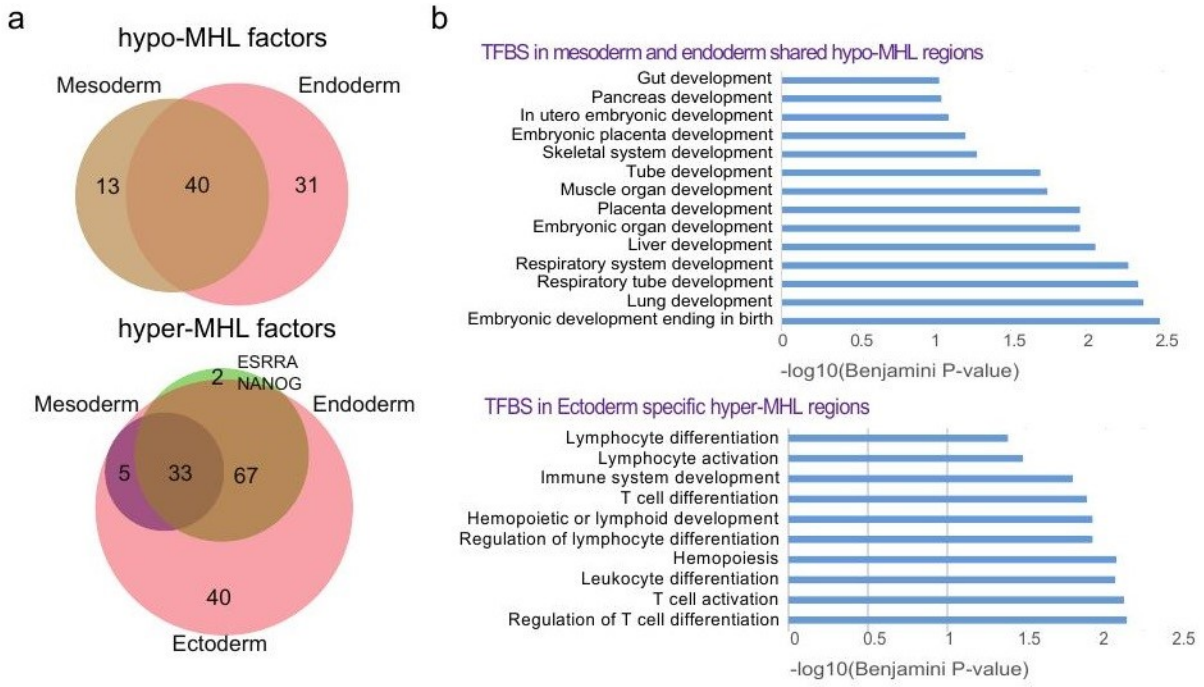
X-axis denote the distances from the centers of MHBs (+/- 1000 bp) and y-axis denotes the average reads density in RPKM (input normalized reads per kilobase per million). Epigenomics Roadmap histones data were downloaded from NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/).



Supplementary Figure 5

PCA analysis of human tissues and cells based on methylation haplotype loads for MHB regions.

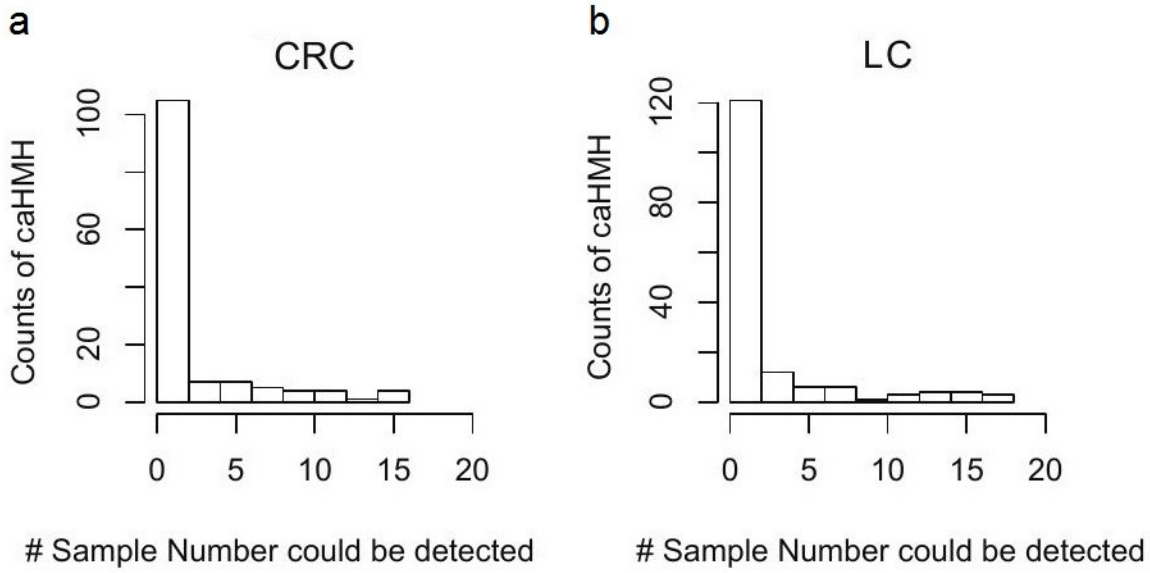
Tissues and cells from WGBS datasets were from 5 other studies and 10 adult tissues WGBS were from this study.



Supplementary Figure 6

Distinctive patterns of functional enrichment for transcription factor binding sites (TFBS) associated with layer-specific MHBs.

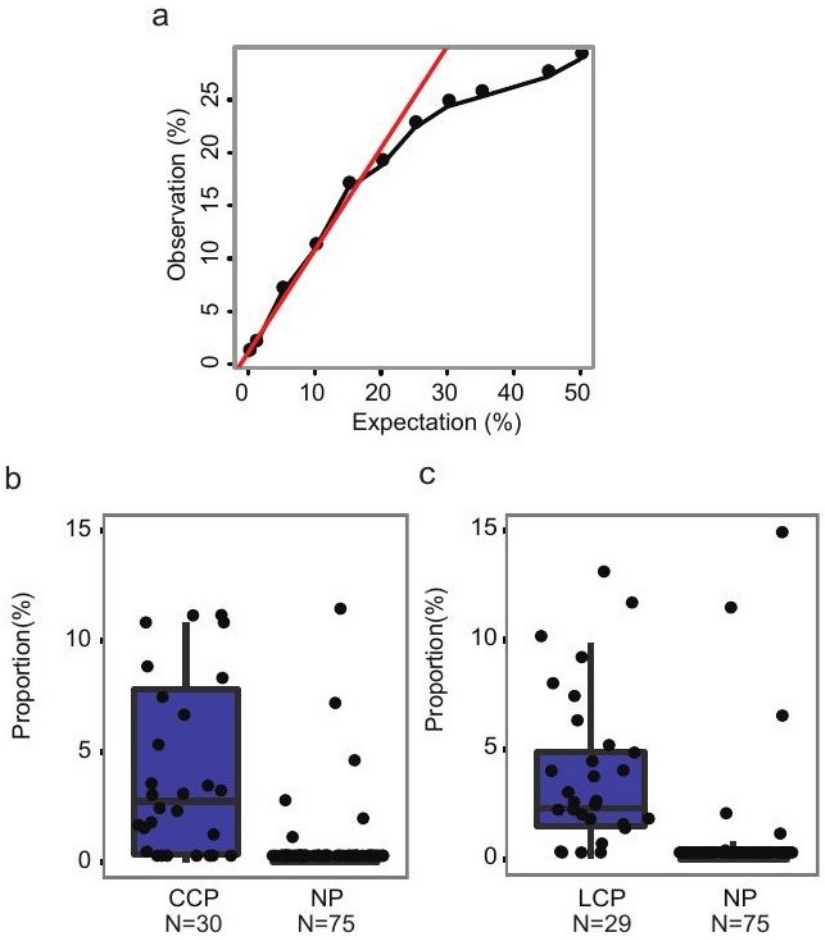
(a) Venn diagrams of transcription factors (TF) with binding sites associated with layer specific hypo- or hyper- MHL regions. (b) Functional enrichment analysis of associated TFBS using GREAT (http://bejerano.stanford.edu/great/public/html/).



Supplementary Figure 7

Distribution of incidence of cancer-associated HMH in CRC and LC plasma samples.

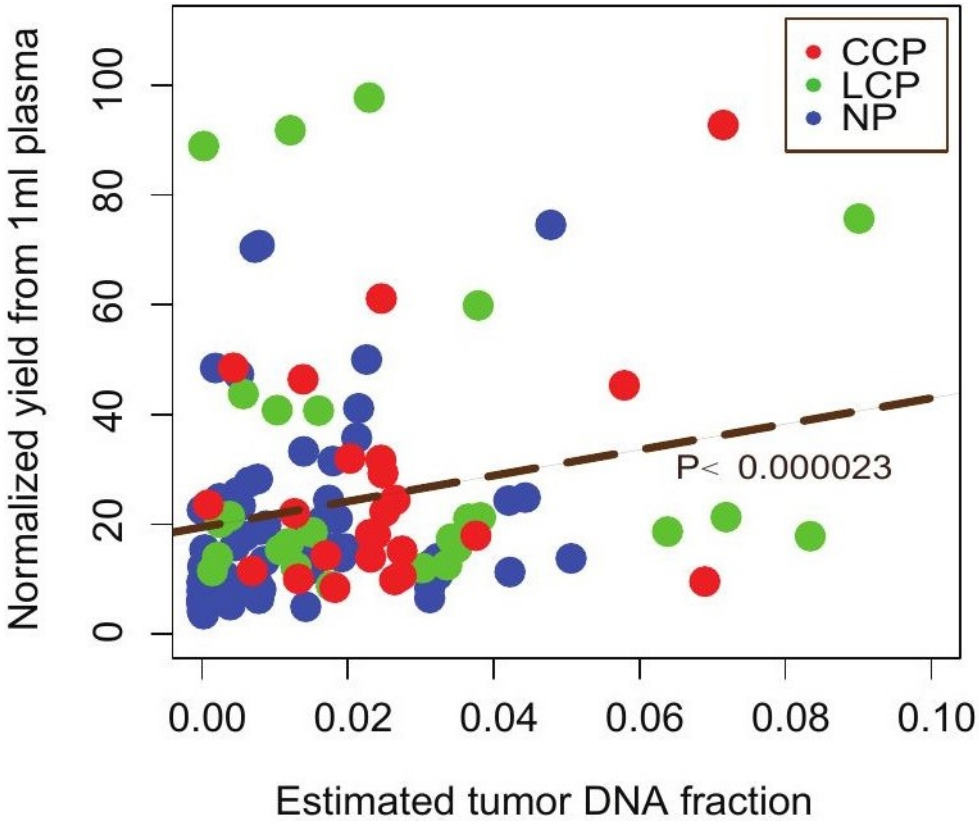
Y-axis denotes the frequency of caHMH and x-axis denotes the incidence (sample number) of the caHMH in CRC plasma samples (a) or LC plasma samples (b). The majority of caHMH are patient specific while a few have high incidence among the cancer plasma samples.



Supplementary Figure 8

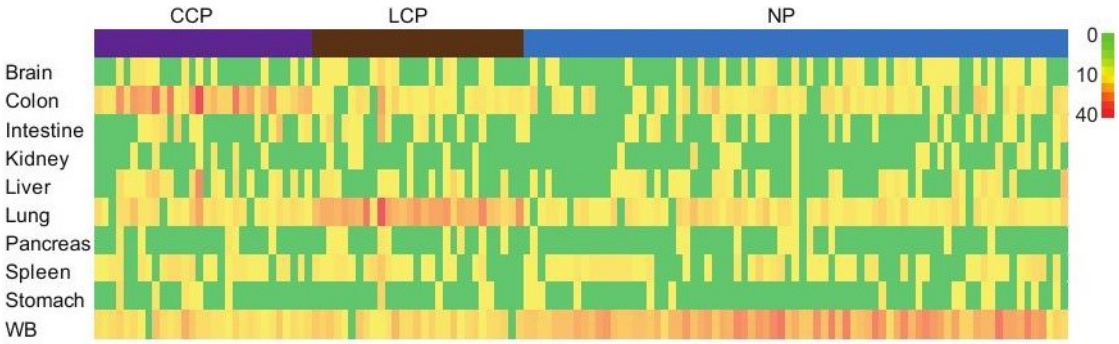
Deconvolution of cancer and normal plasma samples using non-negative decomposition with quadratic programming.

(a) Deconvolution accuracy as a function of tumor fraction using simulated data. (b) Cancer DNA proportions estimated by deconvolution of plasma samples using CCT or LCT as the tumor reference.



Supplementary Figure 9

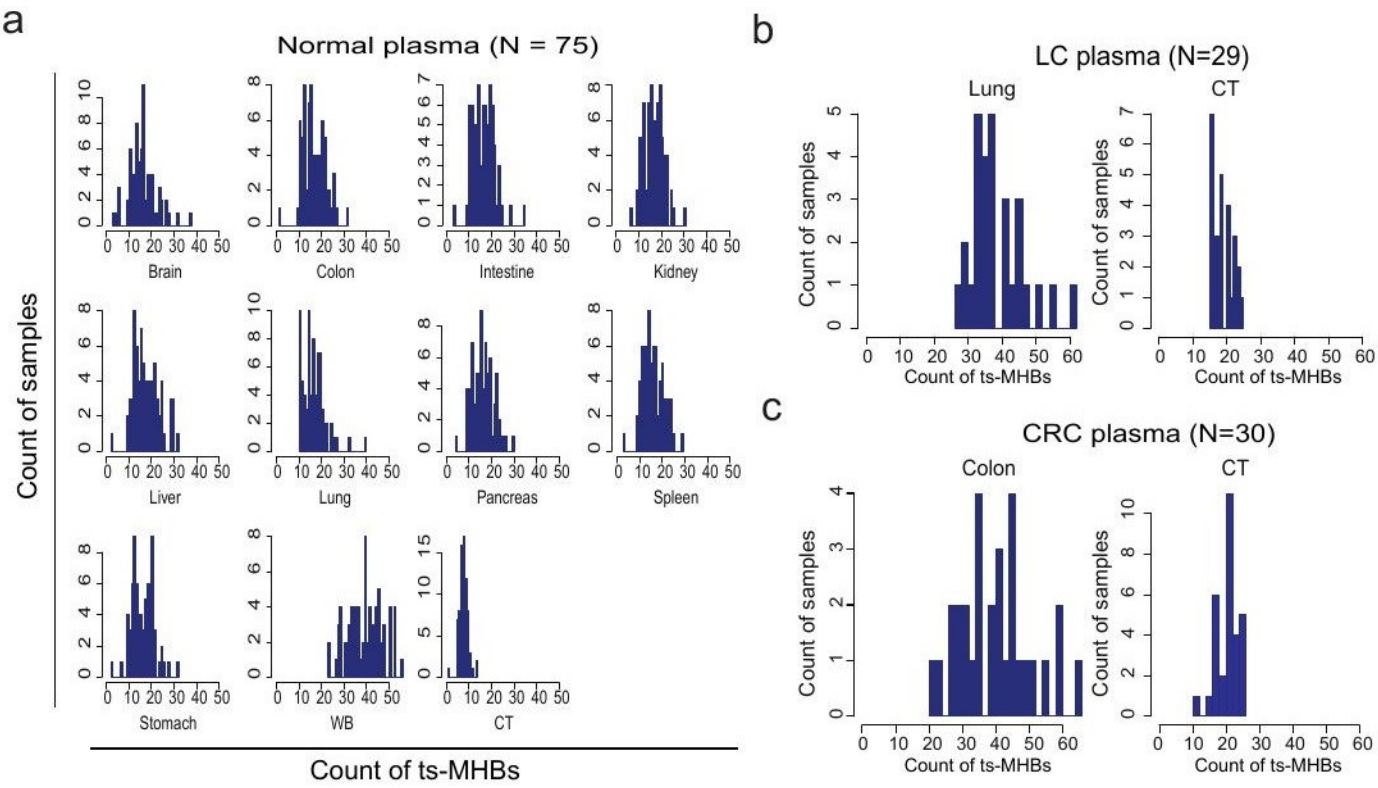
Estimated tumor fraction in plasma correlated with the normalized yield of DNA extraction from plasma.



Supplementary Figure 10

Tissue specific MHBs counting.

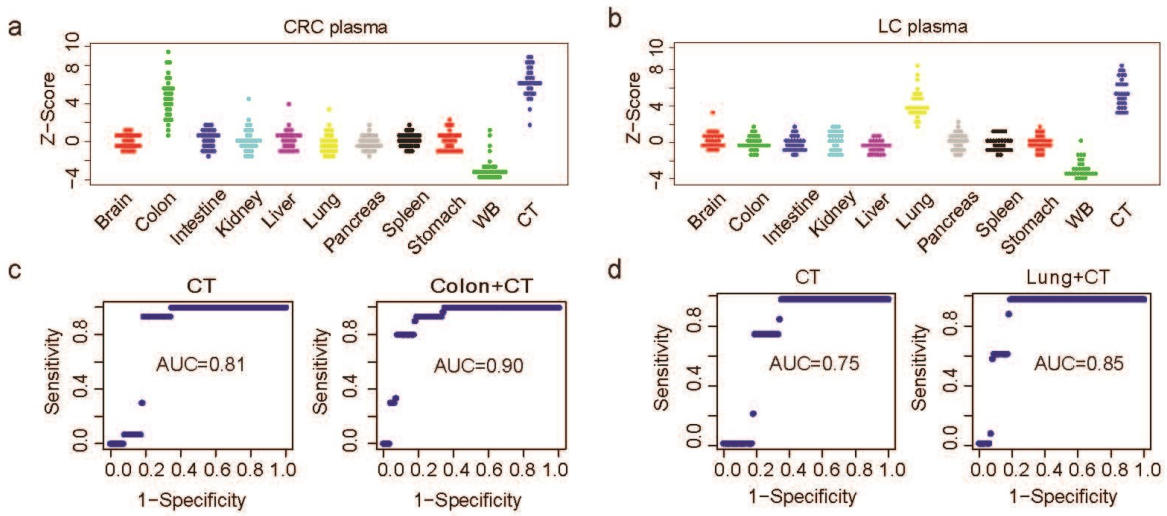
Color bar represents the number of tissue specific MHBs (for each respective tissue) over the MHL threshold in each plasma sample.



Supplementary Figure 11

Distributions of count of ts-MHBs above MHL thresholds in plasma samples.

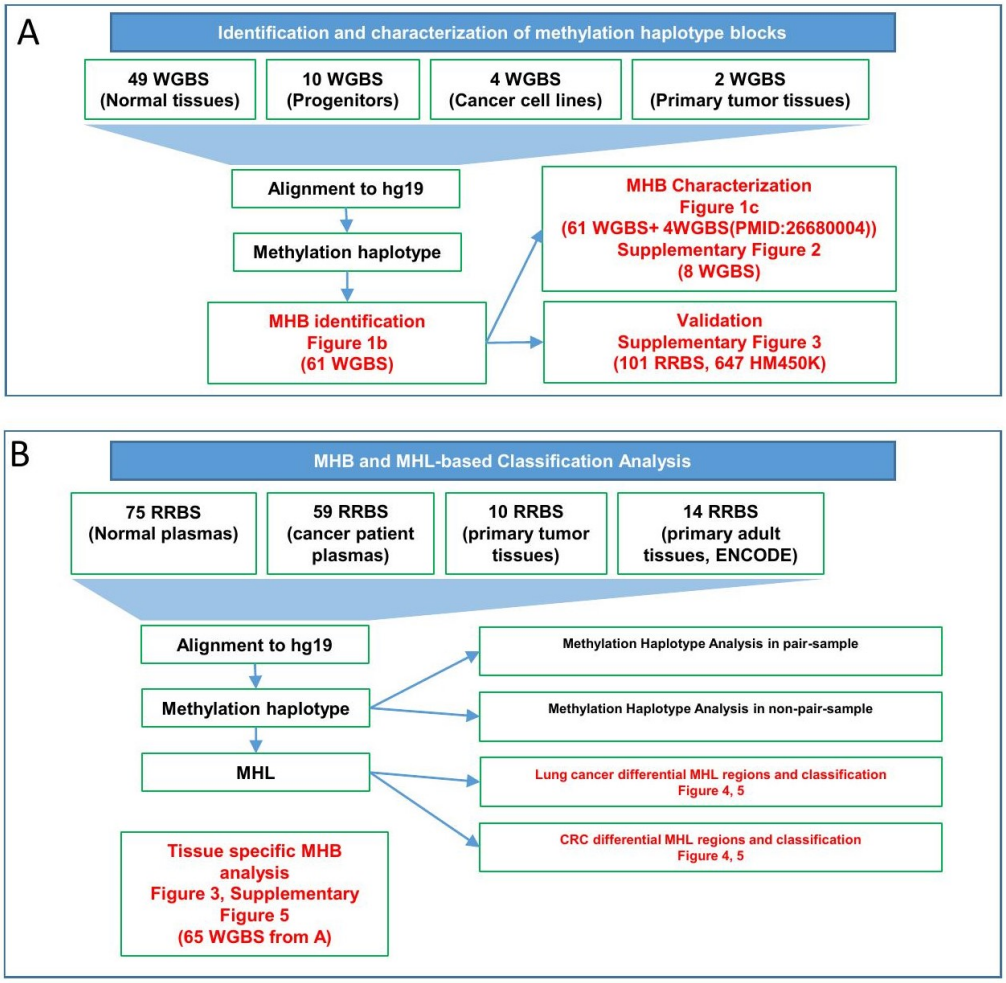
(a) Distributions of count of ts-MHBs for normal plasma samples to 11 reference tissues. (b) Distributions of count of ts-MHBs for lung cancer plasma samples to lung tissue or to pan-cancer tissue (CT). (c) Distributions of count of ts-MHBs for colorectal cancer plasma samples to colon tissue or to pan-cancer tissue (CT).



Supplementary Figure 12

Joint prediction of cancer status and tissue-of-origin from plasma.

Distribution of Z-score in each set of reference-specific MHBs for colon cancer plasma samples (a) and lung cancer plasma samples (b). Integrating signatures from cancer and tissue-of-origin (Colon+CT; Lung+CT) improves the prediction accuracy (c,d), on both types of plasma samples, over focusing on cancer signatures alone (CT). The ROC curves were created by adjusting the Z-score cutoff for calculating specificities and sensitivities. AUC denotes area under the curve.



Supplementary Figure 13

Flowchart of data processing and samples usage.