**Tables and Figure legends,**

Figure 1. DNA methylation haplotype in human genome inferred by different tissues in different platforms. (A) Workflow diagram for methylation haplotype calling and its application in cancer diagnosis and tumor-of-origin prediction for cancer plasma. (B) DNA methylation blocks were inferred with genome-wide bisulfite-sequencing.The pearson correlation was calculated and empirical threshold were applied for the methylation block inference genome-widely.(C) linkage disequilibrium between adjacent CpG loci in different cells including stem cells and progenitors, normal adult tissues as well as primary tumor. (D). The distribution characteristics of the methylation blocks in human genome including gene region and CpG high density regions. (E). Enrichment analysis to MHB in known regulatory elements. (F). MHB inferred by WGBS data were over-represented in MHB inferred by RRBS and MH450K data. (G). MHB inferred by MH450K array were highly overlapped in solid tissue and circulating PBMC.

DNA methylation blocks were inferred with genome-wide bisulfite-sequencing. The Pearson correlation was calculated and empirical threshold were applied for the methylation block inference genome-widely. (B) The distribution characteristics of the methylation blocks in human genome. The left and right were showed with or without intergenic regions. (C) Methylation block regions identified by BS-seq data of normal tissues were significantly over-represented in different platforms such as Methylation 450K microarray and RRBS. The right figure showed the regions from methylation 450K and RRBS within the methylation blocks shown higher correlation compared with regions without the methylation block regions. (D) Methylation blocks were shown highly conservative in 11 normal tissues from methylation 450K array. (D). large number of dataset from GEO also validated the conservative of methylation blocks. Meanwhile (D and E) shown the conservativity of methylation blocks were not influenced by CpG or not.

Figure 2. The performance of methylation haplotype load in distinguishing the methylation complexity and methylation frequency compared with other metrics in different scenarios. methylation entropy (ME) and epi-polymorphism were same with previous report and the formula were shown in method section.

Figure 3. Methylation haplotype load has ability in tissue origin, development layer and disease status distinguishing. (A) Unsupervised cluster analysis shown genome-wide DNA methylation haplotype load could represent the sample relationship. The sample with same origin were cluster together preferentially. (B) Layer specific methylation haplotype region showed the development layer relationship. (C) Tissues specific methylation haplotype loading (MHL) shown advantages in tissues distinguish compared with average methylation frequency (AMF) and methylation for all CpG site (MAS). Tissue specificity value (TSV) was the average MHL for the corresponding tissue specific MHL in the correct samples while the background value (BV) were the average MHL in mis-assigned samples. Contract value were the division of TSV by BV. (D). TFBS located in layer specific MHB regions shown different mechanism which MHB involved in layer development. (E). Genome ontology analysis to layer specific and share transcript factors to infer the roles of MHB in layer development.

Figure 4. Plasma signature by MHL with different analysis strategy. (A). schematic diagram to show tissue specific MHL signature were existed in normal tissue, solid tissue and tumor plasma while they are clean in normal plasma and whole blood sample in colon cancer, lung cancer and pancreatic cancer. (B) Tissue specific haplotype regions in plasma show increased MHL level in cancer plasma compared with normal plasma. (C, D) Random forest prediction model for tumor-of-origin prediction based on tissue specific MHL regions. WGBS and RRBS dataset with balanced samples for each tissue were collected to train the prediction model and our lab RRBS dataset were used to validate the prediction performance.

**Tables**

Table 1. Relationship between methylation haplotype and LDA, VMR, LOCK

Table 2. Methylation haplotype regions useful for colon, lung, pancreatic cancer mapping.

Table 3. Cancer plasma mapping accuracy (discover dataset and validation dataset)