**Supplementary Figure 9.** Circulating tumor DNA components de-convoluted by DNA methylation haplotype. High-methylation haplotype of cancer plasma and normal plasma were compared with normal tissues, cancer primary tissue and WB. The most predominant contribution was removed one by one to estimate the contribution from different tissue or cells.

We identified large number of significant different methylated MHB fragments in cancer plasmas and large number of them have been previously identified as the promising DNA methylation predictor for cancer non-invasive diagnosis were validated to be high frequently methylated in cancer plasmas. We also demonstrate that MHL have powerful ability to detect rare (<1.0%) long continuous methylated fragment compared with 5mC level and the power positively correlated with the number of continuous methylated CpGs. We observed that the MHL signal will be 10-20 time higher than 5mC level when the number of continuous CpG site come up to 6 CpGs. Our result provided a model to detect rare cancer specific long continuous methylated fragment. The prediction performances based on MHL and average 5mC within in genomic interval are better than the single CpG site analysis which mainly caused by high variation and high missing ration for single CpGs among the samples of same groups which indicating the methylation predictor would be more powerful when more continuous CpG site were measured and analysis together.

## Summary

We apply linkage disequilibrium to the methylation alleles on a single DNA molecule and defined methylation haplotype blocks in human genomes. We demonstrate MHB represent a distinct class of genomic feature with gene regulation and tissue differentiation. Furthermore, we proposed methylation haplotype load (MHL), an average methylation weighted by order of the CpGs within MHB, to quantitatively measure both DNA methylation level and methylation complexity simultaneously. Eventually, we demonstrate MHL based genome-wide DNA methylation analysis would provide a novel strategy in cancer biomarker identification and tumor-of-origin prediction to cancer plasma samples.