**Uniform tissue-of-origin and pathology status assignment with distribution of reference specific MHL**

Additional, as we known, the cell-free DNA based diagnosis or prognosis should provide tissue-of-origin and tissue pathology status at the same time. We proposed a unified statistic frame based on distribution of the reference specific biomarker in background references (with tissue- and pathology specific MHBs) to assign the tissue-of-origin and pathology status with corresponding P-value and false discover ratio (FDR). Our proposed method could provide three different applications. 1, tissue-of-origin assignment based on maximum Z-scores among all the references. 2, significant higher biomarker occurrence based on Z-scores for each sample to each reference. 3. Multi-hitting of the biomarkers to one tissues for complex disease or multiple disease in one person (with tissue-of-origin markers and pathology markers).

In the first stage of our method, normal samples (RRBS) were collected to construct the distribution of the reference-specific MHLs of normal plasma in different references (tissue and pathology). We assign the tissue-of-origin and pathology status for the cancer plasmas given the FDR<0.01 in normal plasma and assign the tissue source to the maximum ts-MHLs in 11 reference (Figure 6A). And then, with different false discover ratio setting (FDR=0.01, 0.05, 0.10, 0.5, 0.75, 0.95, 0.99, 1), we could collect the corresponding threshold to predict tissue-of-origin and pathology status and provide ROC curve for each reference (Figure B and C). We found only the tissue-of-origin and pathology status could show good ROC curves (AUC>0.7, Figure B and C) while combining of the two signals will provide excellent performance (AUC>0.85). In addition, Z-score based P-value for samples to each reference provide us the opportunity to infer multiple disease for one person with the exact P-value which would be useful for complicated diseases (Figure D, E, F). by integrating both types of signals, we achieved a 99% specificity (NCP54 excluded) and 79.65% sensitivity in predicting cancer and 89.8% accuracy in predicting the tissue-of-origin for the 59 cancer patients controlled by 1,000 background sample resampling validation. Summary, we established a unified approach based on Gaussian distribution to conduct tissue-of-origin and pathology status and demonstrate that combining of multiple reference will have higher performance to mapping the cell-free DNA methylation aberrant to its disease focus.