Phenome-wide Association Study Maps Genetic Variation in Epigenetic Factors with Human Complex Disease

In the past decades, population genetics based genome-wide association study has been unprecedentedly developed to identify the susceptibility genes for human complex disease. Thousands of susceptibility variants or genes were identified GWAS or candidate genes based fine mapping strategy. However, my previous study showed the prediction ability was severely limited with significant SNPs identified by GWAS study even for some high familial risk disease (Cancer Medicine, 2014). Recently, some evidences shown genetic risk score would help to have improved distinguish ability between genetic disease and normal individuals, however, the effect were still limited. One the other side, majority of the GWAS identified significant SNPs were non-coding variants, which are locating in intergenic, intron or UTR regions. These evidences demonstrate we need pay more attention to human epigenetic variants and these variants might play quite important roles in the susceptibility and pathology of human disease. Among all the epigenetics variants, DNA methylation is the most common investigated factors since the technique is relatively matured and the biological function of DNA methylation is deeply understudied. In order to investigate these epigenetic variants, Genome-wide epigenetic association study (EWAS) or genome-wide DNA methylation association study have been conducted in some mediate or low heritability human disease, such as rheumatoid arthritis, type 2 diabetes, obesity and human cancers. EWAS study could provide the fundamental evidence that whether human epigenetics play important roles for the pathology of the disease and then more case-control study can be conducted to find exact epigenetic variants in disease-origin tissues, which is determined by the truth that the epigenetic profiles have strong tissue-specificity. However, It is difficult to be extend to more diseases since the cost of EWAS is almost 10 times higher compared with GWAS study. What’s more, GWAS study could apply linkage disequilibrium to select tag-SNPs to cover the whole genomic regions, however, EWAS cannot rely this mechanism. Current, even by the most latest DNA methylation 850K array, we can only cover 3.01% total CpGs in human genome. In this study, we will evaluate the roles of epigenetic variants in more than 6,221 human phenotypes by investigate the genomic variants in the whole panel of epigenetic factors which including about 250 human epigenetic modification factors, such as DNMT1, DNMT3A/3B, TET1/2, DOT1L and so on.

The whole project will contain two stages: discovery and validation stage. In the discovery stage, we will conducted a phenome-wide association study (PhenWAS) to epigenetic factors with 6,221 human traits based on Marshfield Clinic PMRP dataset. PMRP dataset contains phase I and phase II data which containing 10,124 and 8,258 human genome-wide exom-chip data. We will extract all the variants nearby (high linkage disequilibrium) or located in the epigenetics factors. We will also include all the epigentic factor trans-eQTL variants in this study to make full investigation of the epigentic factors to human disease. In my preliminary selection, at least 5,873 SNPs will be enrolled in our study which might be increased since novel epigenetic genes might be identified in the coming several month. On the other side, I will impute the current raw dataset to the whole genome with beagle and then the totally candidate SNPs might be ~50K. The phenotype information for these 10,124 and 8,258 samples will be mapped to ICD9 and PheWAS code. Power analysis will be conducted to evaluate the sample size for each phenotypes. We will apply different genetic model to make a full analysis to the association including dominant model, recessive model, additive model, weighted burden and compound heterozygote test.