Statement of Research Interests

Shicheng Guo

School of Life Sciences, Fudan University, Shanghai, China

Division of Biostatistics, The University of Texas Health Science Center at Houston

July 21, 2014

My research interests and activities mainly centered around the following areas: 1) Susceptibility or heritability components of Lung cancer derived from genetic (SNP, CNV) and epigenetic variation (DNA methylation). 2) Screening, early diagnosis and prognosis model for lung cancer identification or surveillance. 3) Lung cancer biomarker research strategy leveraged by high dimensional genome-wide genetic or epigenetic dataset based on microarray or next generation sequencing. 4) Novel Lung cancer related DNA methylation aberrant detection methods develop (bisulfite free detection) and analysis method (Functional based analysis method). 5) Relationship between elements of genome and epigenome and the interaction routine among environment factors with genetic and epigenetic variations in lung cancer. Additional the following three fields have been stepping into my research list in the present.

**Methylation and SNP Double Model in Next Generation eQTL Research**

A fundamental challenge in the post-genome era is to understand and annotate the consequences of genetic variation, particularly within the context of human tissues, for example, annotation to expression quantitative trait loci (eQTL). eQTL are most important genomic variations which have great biological regulation power. However, traditional research usually just focuses on SNP variations in human genome. CpG methylation derived expression quantitative trait loci are another important source of the gene expression variation. Genome variation, DNA methylation and gene expression have complicated relationship. Both DNA methylation and DNA methylation could cause high or low gene expression for specific gene, also, they might bring alternative splicing so that give complex biological phenotypes or traits. TCGA has provided large number DNA methylation, SNPs, CNVs and gene expression for same individual, which means, we can conducted our above analysis right now. Association between SNP and DNA methylation, DNA methylation and gene expression, SNP and gene expression can be easily validated by current molecular biological technique. The identification of such relationship would provide valuable information for pharmaceutical drug design, personalized medicine and fundamental of molecular/cellular biology.

**Function based next generation sequencing analysis in lung cancer research**

RNA-seq technology provides huge biological information of gene expression and alternative splicing for biologist and medical scientists to discover diagnostic or prognostic biomarkers. Cumulative sum method was generally adopted in current RNA-seq analysis. However this analysis would ignore alternative splicing information which would play important role in the pathogenesis of the complex disease. He provided an effective novel pipeline to analysis of next-generation RNA-seq data based on Functional PCA which can identify aberrant alternative splicing in specific disease or conditions and can discover specific biological variation/subtype, such as cancer or normal, drug response status. This methodology takes the spatial information in the RNA expression characteristic into the consideration, which would be a great innovation in RNA-seq analysis and biological theoretical.

**Genome-wide Epigenetic Association Study between Methylation and Middle Heritability Disease**

In the past decades, population genetics has been unprecedentedly developed, especially, in complex disease. Hundreds of susceptibility genes were identified by genome-wide association study. 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. Missing heritability mainly derived from epigenetic variations, has been proposed by large number of genetic epidemiologists. Genome-wide epigenetic association study (eGWAS) or genome-wide DNA methylation association study provided powerful ability to discover disease association epigenetic pathological or etiological factor for middle or low heritability disease. Current DNA methylation high throughput technology, such as MBD-seq, Methylation microarray, has equmented such ability to apply eGWAS or mGWAS on the 1000-2000 population size with case-control design or cohort study (Samples could be obtained from our cohort population in Taizhou, Jiangsu). Some interesting binary outcome disease such as caner/normal, or quantitative trait, such as body-mass index (BMI), relative lymphocyte proportions (RLP), blood pressure (RP), intelligence quotient (IQ) can be considered in our future research proposal in China or U.S or as the International collaboration project.

Key words: Lung cancer, susceptibility, heritability, genetic and epigenetic variation, diagnosis model, biomarker, eGWAS and GWAS.