In the following two sections, Research experience, accomplishments and research interests were provided to show that the interest and experience were highly fitted to your lab. Really hope to have a chance to obtain a postdoctoral position in your lab.

**Research experience and research accomplishments**

Over the past five year, I have participated in numeric model construction project which included risk prediction model based on genetic variations (SNP and CNV) for cancers, rheumatoid arthritis and systemic lupus erythematosus; diagnostic biomarker discovery projects, such as methylation biomarker panel for prostate cancer, lung cancer. PLZF hyper-methylation in non-small cell lung cancer, association between CNV variation in HLA region with ankylosing spondylitis, systemic lupus erythematosus and prediction model based on significant GWAS SNP to thyroid cancer. All of these have been published on SCI journals which have been listed in my curriculum vitae.  
Although great progress in genome-wide association studies (GWAS) has been made, the performance of the most current SNP-based disease risk prediction methods was not assessed accurately. The prediction ability with several significant SNPs which is identified by GWAS studies was evaluated with multiple popular and powerful prediction model which included K-nearest neighbors (KNN), Logistic regression (LR), Naïve Bayes (NB), Random forests classification (RF), Support vector machine (SVM), Bayesian additive regression trees, Boosting, Recursive partitioning, Fuzzy rule-based system. Since thyroid is one of complex diseases with highest heritability among all the cancers, the result has an important implication that the role of the prediction ability based on few significant SNPs from GWAS study is limited (Cancer Medicine, 2014). Also, this research re-emphasized that associations identified by GWAS account for only a few percent of the genetic variance, more attentions should be paid to the missing heritability caused by proteomic and metabolomic variations and epigenetic variations such DNA methylation and histone modifications and temporal and spatial specific gene expression.

Some autoimmune disorders such as rheumatoid arthritis and systemic lupus erythematosus have been demonstrated to be highly associated with the genetic variation in HLA region. Between 2012 and 2014, we found serial of HLA related copy number variations were significantly associated with autoimmune disorders which included HLA-DQA1 with RA (Journal of Genetics, 2014) and ankylosing spondylitis (Genes and Immunity, 2013), HLA-DRB5 with SLE (Acta biochimica et biophysica sinica, 2014), APOBEC3A/3B and systemic sclerosis (Arthritis and Rheumatology, 2014, under review, AR-14-0879), in Chinese population. It would be provide great help on the risk prediction model for autoimmune disorders. HLA-region is most important components for personalized medicine. These evidences demonstrate the importance of the copy number variation in the risk prediction model for complex disease and therefore would be potential target or candidate components in the personalized medicine. In addition, the manuscript about the association between NOTCH4, TAP2, CCHCR1 and systemic sclerosis has been completed and will be submit soon.

DNA methylation was suggested as the promising biomarker for early diagnosis of lung cancer. However, it is a great challenge to search for the optimized combination of the methylation biomarkers to obtain the maximum diagnosis performance.  I developed a panel of DNA methylation biomarkers and validated their diagnostic efficiency for non-small cell lung cancer (NSCLC) in a large Chinese Han NSCLC retrospective cohort. In the discover stage, three high-throughput DNA methylation microarray datasets were collected from public database (GEO, ArrayExpress). After normalization, batch effect elimination and integration, significant differential methylated genes and best combination of the biomarkers were determined with leave-one-out support vector machine (SVM) feature selection operation. Then candidate promoters were examined by methylation status determined single nucleotide primer extension technology (MSD-SNuPET) in an independent set of NSCLC/normal tissues. He proposed an effective DNA methylation-based biomarker discover pipeline and identified a promising panel for NSCLC diagnosis. High throughput DNA methylation microarray dataset followed by batch effect elimination can be a good method to discover optimized DNA methylation diagnostic panels. Methylation profiles of AGTR1, GALR1, SLC5A8, ZMYND10 and NTSR1, could be an effective methylation-based assay for the NSCLC diagnosis. The works has been accepted by Clinical Epigenetics (2015,7(1):3).

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.  The work now is being prepared to the manuscript and will be submitted to Proceedings of the National Academy of Sciences (PNAS).

**Research interests**

With his diverse interdisciplinary training in genomics, human genetics, biostatistic and epidemiology, Dr. Guo's research has focused on identification of genetic and epigenetic interactions and their roles in diagnosis, prognosis and personalized medicine for complex dieases.

My future professional research interests and activities will concentrate in the following areas: 1) Disease susceptibility/heritability derived from DNA methylation, 2) population epigenetics, 3) DNA Methylation based biomarker identification leverage high dimensional dataset, 4) Interaction between DNA methylation and genetic variations and 5) noval methylation analysis method (Functional based analysis method). 6) Relationship between elements of epigenome (5m, 5mc, 5hmc, histone modifications) and the application on human disease. Human genome project, 1000 genome project and Human cancer project have been completed one by one. More and more evidences show that the contribution from the epigenetic variation is more important than genetic variations. Actually, I think it suggest that the environment plays the most part of the role on the pathological or etiological of the majority complex disease, such as cancer, coronary artery disease, autoimmune diseases and so on. Some detail information can be seen as the following description.

1) Genome-wide Epigenetic Association Study between Methylation and Median 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, Fudan University, Shanghai). 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.

2) Population epigenetics, allele specific DNA methylation and the relevance on the epidemiology of human diseases.

Our knowledge of the mechanisms that specify and propagate epigenetic states of gene expression is expanding rapidly; however, the significance of variation in epigenetic states at the population level remains unexplored. Population epigenetics would become an active subfield at the interface of molecular genetics, genomics, and population biology to addresses questions concerning the prevalence and importance of epigenetic variation in the natural world and disease status. Some high similar disease can be clustered together and share similar therapy based on genome-wide DNA methylation profile.

3) Methylation and SNP based interaction effects 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.

4) Genetic and epigenetic based personalized medicine outcomes evaluation and assessment

Healthcare is about to undergo a revolution as new technologies to predict, detect, monitor, and treat disease at the molecular level become available. Increasingly, inexpensive genomic technologies are now being introduced to identify genetic variants, inform preventative care and guide disease therapy. New proteomic and metabolomics technologies will soon provide accurate diagnostics to monitor the wellness of organs in the body, to detect disease and to determine efficacious, non-toxic therapy. Developments in the relatively new areas of the epigenome and microbiome will provide additional insight into disease therapy and maintenance of wellness. The advent of healthcare based on the unique molecular makeup of each individual, commonly termed personalized medicine, will allow effective preventative care, will improve the safety (reduce adverse drug reactions), efficiency (only treat patients with medicines that work on them) and effectiveness (more accurate diagnoses, better matching of treatment to disease) of the healthcare system and has potential for reduced per capita costs and a dramatically improved patient and provider healthcare experience. Despite the costs associated with new genetic-based tests and treatment, the appropriate uptake of PM strategies can potentially decrease health care costs and improve patient outcomes. However, without appropriate health technology assessment (HTA) and technology diffusion, PM could significantly increase costs without improving health outcomes. With this increasing need for HTA specifically for PM-specific drugs and technologies, I want to do some jobs on the rigorous evaluation, interpretation and dissemination of health outcomes information.