Shicheng Guo
Fudan University
220 Handan Road
Shanghai, China, 200433
+1-832-538-7052 (Mobile)

Dear Sir/Madam,

I am writing to inquire about the possibility of obtaining an associate research fellow position in School of life sciences, Tongji University. I will graduate from the Fudan University's Genetic Epidemiology Ph.D. program (Shanghai, China) in Dec, 2014. As a joint-cultivated doctoral student, I'm under the guidance of Dr. Li Jin (Fudan University, School of Life Sciences), Dr. Momiao Xiong (University of Texas, School of Public Health) and Dr Jiucun Wang (Fudan University, School of Life Sciences).

My thesis work focused on discovering disease relevant genomic variations, especially DNA methylation, Single-nucleotide polymorphism (SNP), copy number variation (CNV), miRNA and so on in the genome of human cancer and other complex diseases with all kinds of high-throughput genetic data (RNA-seq, MeDIP-seq, Methylation and RNA microarray) to expand the biomarkers which can be used in the screening, diagnosis, prognosis and drug response. And then to establish prediction or diagnosis models to assess the potential application value of the biomarkers for risk prediction or diagnosis. Majority of the designs in my thesis were based on population based case-control study. Batch effect removing, miss value imputation, alignment or mapping, genome function annotation, risk assessment, prediction model establishment, association analysis were deeply applied in my Ph. D training. 5 years Linux, Perl, R and programming experience provide me effective guarantees to finish the analysis with high replicability, especially for big dataset.

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; He re-analyzed his previous published association data and evaluated the prediction ability with several significant SNPs which is identified by GWAS studies. Since thyroid is one of complex diseases with highest heritability among all the cancers, his result has an important implication that the role of the prediction ability based on few significant SNPs from GWAS study is limited. 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 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.

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. He 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. Three high-throughput DNA methylation microarray datasets were collected in the discover stage. 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 reviewed by *Cancer Research* (CAN-14-1687). I fully believe that his works in these areas are of great importance and success.

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

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, 5) Novel DNA methylation detection (bisulfite free detection) and 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.

1) Methylation and SNP based double 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.

2) 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.

3) 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 metabolomic 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, we are going to focus on the rigorous evaluation, interpretation and dissemination of health outcomes information.

I would greatly appreciate the opportunity to talk to you more about the possibility of a statistics staff scientist position in your lab. I have attached my resume to this email for your review. Three references were provided as the follwoing, please do not hesitate to send email to them to achieve more information for me.

References

Li Jin Dr.	Jiucun Wang Dr.	Momiao Xiong Dr.
School of Life Sciences, Fudan University	School of Life Sciences, Fudan University	School of Public Health, UT
Lijin.fudan@gmail.com	Jc.wang@fudan.edu.cn	Momiao.xiong@uth.tmc.edu
+86-21-55664474	+86-21-55665499	(713)-500-9894

Thank you for your consideration.

Sincerely,

Shicheng Guo

July 9, 2014