Re: Tenure-track Informatics Scientist in Center for Precision Medicine Research
Marshfield Clinic Research Institute

Dear Dr. Brilliant and Search Committee:

I am writing to apply for the open Tenure-track Informatics Scientist Position in Center for Precision Medicine Research (CPMR), Marshfield Clinic Research Institute (MCRI). I am extremely interested in this position at MCRI, where I believe my Ph.D and Postdoc training in bioinformatics, genetic epidemiology and clinical epigenetics will allow me to meet the challenge of precision medicine in which demographical, clinical, genetic, epigenetic and pharmaceutical information is combined. I view this position as an opportunity to use my informatics and molecular skills to develop an independent and multi-faceted research program in precision medicine, investigating the genetic and epigenetic interactions that underlie complex diseases. Currently, I am a Postdoctoral Research Fellow in the CPMR at the MCRI where I collaborate with Dr. Steven Schrodi working on (i) Identify novel susceptibility and functional genetic variants for autoimmune diseases (ii) Identify the genetic and epigenetic interaction in human complex diseases. (iii) Circulating cell-free DNA methylation based diagnosis and prognosis biomarkers for autoimmune diseases with powerful machine learning methods mediated by different feature selection strategies.

My formal training is in human population genetics and bioinformatics. As a graduate student, I design and conducted a series of genetic epidemiology association studies and genome-wide epigenetic profiling studies for cancer and autoimmune disease. In these projects, I participated the whole process including study design, data collection, data analysis and manuscript preparation. Therefore, I am well trained both in computational and molecular experiments. With regard to computational expertise, I have 8 years of programing experience in Perl, portable batch system (PBS) and Linux system (Ubuntu and Centos), 5 years of programing experience in R and Matlab and 2 years of experience with Python. Additionally, I have the basic skills on MYSQL and SAS to deal with database architecture. I have also completed coursework in text mining and have considerable experience in text mining for non-structured dataset. All these abilities are displated in my previous work, such as, in 2014, I applied a multiple prediction model including 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 to evaluate the prediction performance of GWAS significant SNPs for thyroid cancer. With regard to experimental work, I have obtained systemic training on molecular experiments including DNA Sequencing and DNA methylation assays (such as methylation-specific PCR (MSP) and bisulfite sequencing PCR (BSP), methylation binding domain sequencing (MBD-Seq), reduced representation bisulfite sequencing (RRBS), and whole-genome bisulfite sequencing (WGBS)). In this part, I integrated three different GEO methylation datasets and then applied a batch effect correction method to identify differential methylated loci between NSCLC and normal tissues. Meanwhile, I also designed a novel median-throughput methylation assay: MSD-SNuPET to validate the previous findings. Lastly, I proposed a novel biomarker panel for the diagnoses of NSCLC with AUC=0.91.

During my training within the School of Public Health, University of Texas (UThealth), I participated in several studies where I developed a novel functional principal components analysis to RNA-seq and image data. I made important contributions to RNA-seq data processing including raw data calling and created solutions to big data storage problems. I proposed a new binary data format to save the hard-disk storage for the functional data. Meanwhile, I contributed to genetic and biological significance to the analysis. During the postdoc training in Department of Bioengineering, University of California, San Diego (UCSD), I finished another project that applied Methylation Haplotype Load (MHL) to quantify methylation level and complexity to infer the tissue-of-origin for the circulating DNA methylation signals which was published in Nature Genetics (I was first author). Our study demonstrated MHL is a good index to estimate the methylation signals in the cell-free DNA compared with traditional methylation level, methylation entropy, Epi-polymorphism. Meanwhile, our result showed that cell-free DNA methylation could be applied to serve as an effective biomarker for tissue-of-origin mapping in different cancers.

I would enjoy discussing this position with you in the weeks to come. Meanwhile, I enclosed my curriculum vitae and statement of research interests. Letters of recommendation will arrive sent by each referee. Please don't hesitate to contact me if you require any additional materials or information. Thank you very much for your consideration.

Sincerely,

Shicheng Guo, Ph.D. Postdoctoral Research Fellow
Center for Precision Medicine Research
Marshfield Clinic Research Institute
1000 North Oak Avenue, Marshfield, WI 54449
Tel: (281) 685-5882, Email: Guo.Shicheng@Marshfieldreasearch.org