**Job Interview for TTAP Faculty Position in UW-Madison**

**The positon focused on novel approaches to the functional assessment of genomic and/or epigenomic variants associated with cancer and other human diseases.**

**Please introduce yourself:**

I received Ph.D degree on Human Genetics in 2015 from Fudan University. During my Ph.D training, I received well trainings in population genetics and epigenetics, especially in DNA methylation. Between 2012 and 2014, I received a fellowship to join a Ph.D exchange program between School of Public Health, University of Texas and Fudan University. After that, I come back to Fudan and received my Ph.D degree. Then in 2015, I returned to US for my Postdoc training in School of Public Health, University of Texas, Department of Bioengineering in UCSD and Center for Precision Medicine Research In Marshfield Clinic Research Institute.

During I was working in University of Texas, I developed a novel functional principal components analysis to accelerate the analysis of high-dimension data including NGS data and image data. In addition, I made important contributions to RNA-seq data processing including raw data calling and created solutions to big data storage.

During in UCSD, I conducted a project in which we want to use DNA methylation signal of cell-free DNA to predict whether you can cancer or not and where this cancer signal come from. I proposed the concept of a methylation haplotype load (MHL) to quantify both methylation level and haplotype complexity simultaneously and to infer the tissue-of-origin for circulating cell-free DNA methylation signals. My study demonstrated MHL is a powerful metric to estimate the methylation signals in cell-free DNA compared with the traditional methylation level, methylation entropy, or epi-polymorphism. In summary, my work provided a new approach to apply cell-free DNA methylation to serve as an effective biomarker for tissue-of-origin mapping in cancer screening.

At MCRI, I leveraged my computational and biological skills to identify novel disease genes within the 20,000 sample Personalized Medicine Research Project using exome-chip data. I implemented a novel genetic model to discover combinations of recessively-acting susceptibility variants for 15 complex diseases. The study identified a novel hemochromatosis gene, FGF6, and I conducted a series of protein-protein network analyses and molecular evolution analysis, providing strong evidence supporting FGF6 being involved in iron overload. Subsequently, I designed functional studies to assess additional evidence for the role of Fgf6 in iron metabolism.

**Why are you interested in our department (Why did you consider our University and why should we hire you)?**

This position is quite match the PHD and Postdoc training. Integrative training in both genetics and epigenetics, computational skills and web lab experiences provide me an opportunity to implement genomic variants in the implementation of precision medicine. I have enough experience to deal with all kinds of data and ability to design the molecular and cellular study to validate the hypothesis created in the computation analysis section.

In addition, the excellent resource from Madison such as SickKids Clinic Genome Project, TSB-biobank at Carbone Cancer Center. I am also quite familiar with the HTcondor scalable computing resource that is supported by NSF, NIH, and DOE. I have used it for large number of GWAS, PheWAS and EWAS analysis. Meanwhile, UW-Madison have multiple facilities and cores to support my research such as Circulating biomarker core (Dr. Josh Lang) at and [Biotechnology Center which can provided all kinds of sequencing in my research such as WGBS, ATAC-seq, ChIP-seq.](https://www.biotech.wisc.edu/)

Finally, I have already built widely collaboration with lots of scientist in UW-Madison, such as [Dr. Mark Craven](https://www.biostat.wisc.edu/~craven/) in department of Biostatistics and Medical informatics. Dr. Judy Smith in Medical Microbiology and Immunology. Current, I am working together with Dr. Randy Kimple in Department of Human Oncology to prepare an ICTR proposal on “Cell-free DNA methylation research in oral cancer”. Meanwhile, I have been discussing with Dr. Qiang Chang in Waisman Center to have some methylation research together.

**Tell us a little bit about your research plans?**

I am extremely interested in this position at UW-Madison, where I believe my Ph.D and Postdoctoral training in human genetics, clinical epigenetics and bioinformatics will allow me to meet the challenge of precision medicine in which genetics, epigenetics and computatio-nal biology are combined to address important clinical problems. I view this position as an opportunity to apply my computational and wet-lab skills to develop a multi-faceted research program in precision medicine, investigating the interactions between genetic and epigenetic factors that underlie complex diseases.

On one side, current I was involved in several collaborations between MCRI and UW-Madison. For example, we have an ICTR proposal which is under review to investigate the cell-free DNA methylation signal and fragmentation size in oral cancer and another collaboration with Steven Schordi and Judy Smith on the GWAS study between genetic variation and cytokine.

On the another side, I will make full use of UW-Madison resource such as SickKids project, TSB-biobank at Carbone Cancer Center

Take Sickkids project as example, as we know, we will identify 100-1000 potential disease or risk associated genetic variations for the specific family or trios. How to determine the real causal variation and how to functional evaluate these genetic variants are great challenge. As we know from the GWAS study, more than 80% of the significant hit are not protein-coding variations. In order to provide better understanding to these diseases associated variations, we need pay enough attention to these non-coding variation or regulatory variations. However, the current functional evaluation are seriously lack of regulatory annotation to genetic variants. As an expertise on DNA methylation, current I am exactly working on the regulatory genetic variants mediated by epigenetic modifications. For example, for human, DNA methylation almost only occurred in CpG dinucleotide, any genetic variations break or create CpG dinucleotide will change the ability of DNA methylation or not and then will influence the gene expression or splicing. Another example in genetic variations located in seed region of microRNA or Piwi-RNA will change the miRNA-binding target and then change the gene expression network. Finally, genetic variations occurred in transcript-factor binding site will also change the chromatin status and further gene expression. Only when we recruited all these variations into the association model or prediction model, we can provide better interpretation with genetic variants to phenotypes. Meanwhile, when we enrolled all these functional variants, we can have better gene based association model, such as recessive model, compound heterozygote model and further high level models such as pathway based analysis model. With these features, we can have better prediction ability to mapping the variants to potential tissues or organ systems. Current, working with Dr. Steven Schrodi, I almost finished an updated compound heterozygote analysis pipeline, which was created in my previous blood paper, in which eQTL and loss-of-function were integrated to increase the power of the association study. After I completed the functional CpG-SNP, functional miRNA-seq and functional TFBS-Seq, I can integrate all these feature in my analysis pipeline and apply it in SickKids project to identify novel functional disease variants. Finally, I will also introduce the novel approach such as WGBS, ATAC-seq, histone ChIP-seq to extend the evaluation pipeline to regulatory SNPs.

Meanwhile, I always investigate epigenetics with genetic concepts. It has been shown with my previous research, such as I proposed methylation haplotype to reflect methylation level and methylation diversity at the same time. I also want to extent this kind of research strategy in UW-Madison. For example, I will start a pilot study in pharmaco-epigenomics research together with current running pharmacogenomics research. My hypothesis is pharcogenetic will provide baseline prediction potential for the drugs while pharmaco-epigenomics could provide recent prediction ability. It will be have higher prediction ability combined these together. I know several scientist such as Dr. [Arash Bashirullah](https://apps.pharmacy.wisc.edu/sopdir/arash_bashirullah/index.php) are working on pharcogenetic research to leukemia, I can proposal an integrative proposal and try to submit the proposal to Pharmacology Study Section.

Where do you see yourself applying for funding?

My potential funding will be from NCI, Cancer Etiology Study Section, Cancer Genetics Study Section. Meanwhile, I will also pay attention to all kinds of Special Emphasis Panel that is DNA methylation related projects. Since I have strong bioinformatics and data analysis skills I will actively participate into the proposals from different collaborators to practice the written skills.

Currently, as a postdoc, I have tried to prepare several proposal to foundation and pharmaceutical company:

Another grant that I have submitted is termed “Genomic Innovation Grants” supported by Intermountain Healthcare. In this grant, they want support the application to apply different sequencing technique to solve interesting clinical and basic research questions. They have 4.5 million archived patient tissue samples and they provided all kinds of sequencing service. You only need to proposal ideas and question want to solve. I have proposed to generate DNA methylation reference for 9 different cells from synovial joint derived from rheumatoid arthritis patents. In the past several month, single-cell RNA-seq have been applied in rheumatoid arthritis synovial joint and identified several new cell types. However, neither known cell type nor new identified cell types, DNA methylation reference for these cells are not existed. It is quite important to provide the methylation reference so that deconvolution analysis to the mixture tissues can be conducted to accelerated biomarker or mechanism research for rheumatoid arthritis

For example, I am preparing a proposal to Boehringer Ingelheim, which is one of the world's largest pharmaceutical companies. The funding is $200,000 European dollars. They want to extend one of their drug to new diseases. This drug is a SOS1:: KRAS interaction inhibitor which is used in colon cancer therapy and they want to know whether this drug can be used in other diseases. In one of my project, I am analyzing a rheumatoid arthritis synovial tissue expression project. And I found in the synovial cell KRAS are significantly activated, therefore, their drug maybe can be used in the therapy of rheumatoid arthritis. I have prepared the proposal and the deadline is December 13.

Can you talk a bit more broadly about your research vision?

Precision medicine requires unprecedented multidisciplinary collaboration and require multi-omics data to provide a full architecture to the diseases or phenotypes. Full consideration to Interaction between genetics and epigenetics is very important to understand the phenotypes for the complex diseases. For these research, we need to strong ability to investigate every details to the raw data generated in the project which might be ignored by the facilities or cores. We should be careful to every step of the analysis or else we might loss the great findings.

What would you need equipment-wise to get things going?

My lab will be an integrative genomics lab including computational work and wet lab work. **For the computational work**, I require GPU and parallel computing systems for raw data alignment, variants calling, association and causal analysis and deep learning. **For the wet lab work**, I require the routine molecular biology equipment such as different kinds of refrigerator, real-time PCR machine. Cell biology related equipment to investigate the gene expression regulation; I think we should have shared facilities. Some other equipment can be shared by numerous facilities in UW-Madison such as [Flow Cytometry Facility](https://mcardle.wisc.edu/what-we-do/resources/facilities#flow_cytometry) in UWCCC.

Whom are you looking to hire right out of the gate: Post-doc? Technician, graduate student? In addition, what kind of person would you look for?

As the beginning, I think my lab will be dominant with Ph.D and Master Students. Since I just finish two round postdoc training, I am quite familiar with computational and wet-lab platform. I will spend 1-2 years to let my lab have excellent working environment and platform, then from the 3rd year, I will hire dedicate postdoc to join the lab to increase the efficiency of the works.

Is there any question for us?

1. Can you give me more information what kinds of resource that will be available to me such as samples, data, computational resource?
2. How to deal with my current on-going project? Can I take parts of them and continuous these research?

Why did you consider our university and why should we hire you?

This position is quite match the PHD and Postdoc training. Integrative training in both genetics and epigenetics, computational skills and web lab experiences provide me an opportunity to implement genomic variants in the implementation of precision medicine. I have enough experience to deal with all kinds of data and ability to design the molecular and cellular study to validate the hypothesis created in the computation analysis section.

In addition, the excellent resource from Madison such as SickKids Clinic Genome Project, TSB-biobank at Carbone Cancer Center. I am also quite familiar with the HTcondor scalable computing resource that is supported by NSF, NIH, and DOE. I have used it for large number of GWAS, PheWAS and EWAS analysis. Meanwhile, UW-Madison have multiple facilities and cores to support my research such as Circulating biomarker core (Dr. Josh Lang) at and [Biotechnology Center which can provided all kinds of sequencing in my research such as WGBS, ATAC-seq, ChIP-seq.](https://www.biotech.wisc.edu/)

Finally, I have already built widely collaboration with lots of scientist in UW-Madison, such as [Dr. Mark Craven](https://www.biostat.wisc.edu/~craven/) in department of Biostatistics and Medical informatics. Dr. Judy Smith in Medical Microbiology and Immunology. Current, I am working together with Dr. Randy Kimple in Department of Human Oncology to prepare an ICTR proposal on “Cell-free DNA methylation research in oral cancer”. Meanwhile, I have been discussing with Dr. Qiang Chang in Waisman Center to have some methylation research together.

Compared with other candidate, I think my advantage is I have multiple trainings on different research strategies so that I can handle diversity data in precision medicine research including multi-Omics data and molecular/cellular studies, both genetics and epigenetics variants. I have ability to analysis the data with the most raw data level. For example, my DNA methylation analysis pipeline is just start with the raw data and very different with the traditional pipeline in which I can integrate all other information/features, such as I can integrate genetic variants, copy number variation into my methylation haplotype analysis platform. I have built well collaboration network between UW-Madison, MCRI. I am quite familiar with the paralleled computing platforms, which is largest scalable computational resource in USA. My computational skills will provide me multiple opportunities to collaborate with other faculties in our center and external centers.

How can I follow the proposed topics and how will get the funding?

Is there anything beyond what I had already shared in my application that I would like to tell now?