**Statement of Research Interests**

**Shicheng Guo, Ph.D.**

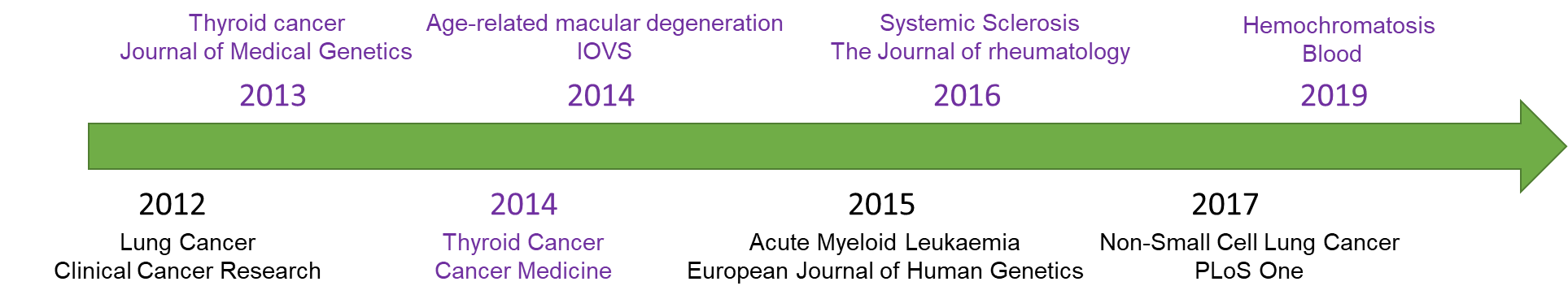
Human complex diseases are dynamically and progressively driven by a progression of genetic-epigenetic-environment-al interactions. These interactions contribute to bringing phenotypes from a susceptibility stage to clinical presentation. The most exciting ideas of precision medicine involve the consideration of numerous genetic susceptibility variants as potential baseline risk factors and track the epigenetic changes in pathogenesis of human diseases over time. Taking rheumatoid arthritis as an example, individuals carrying HLA-DRB1\*04:01 in combination with inflammatory processes and citrullination caused by factors such as smoking, are thought to initiate an autoimmune process. Early molecular and epigenetic biomarkers such as rheumatoid factor (RF), anticyclic citrullinated peptide (anti-CCP) and circulating cell-free DNA methylation signals caused by synovial cell apoptosis precede the manifestation of symptomology. I have a strong interest in delineating these genetic-epigenetic-environmental interactions to elucidate the etiologies of complex diseases. In the past 10 years, I have been working to identify genetic and epigenetic variants in multiple human complex diseases and authored >50 peer-reviewed publications, 17 of which are first/co-first/co-corresponding author publications. My extensive experience in conducting human genetics, epigenetics research combined with my bioinformatics skills in analyzing and interpreting data has prepared me to conduct comprehensive research studies in precision medicine. My future research will focus on 1) identifying and implementing detection of circulating cell-free DNA-based genetic and epigenetic biomarkers for early diagnosis, real-time monitoring and prognosis prediction; 2) Functional assessment of the genetic and epigenetic variants identified by GWAS, PheWAS, EWAS studies; and 3) developing epigenetics biomarkers applying a pharmaco-epigenetics (PeGx) approach to coordinate with PGx in guiding personalized treatment.

**Research Background**

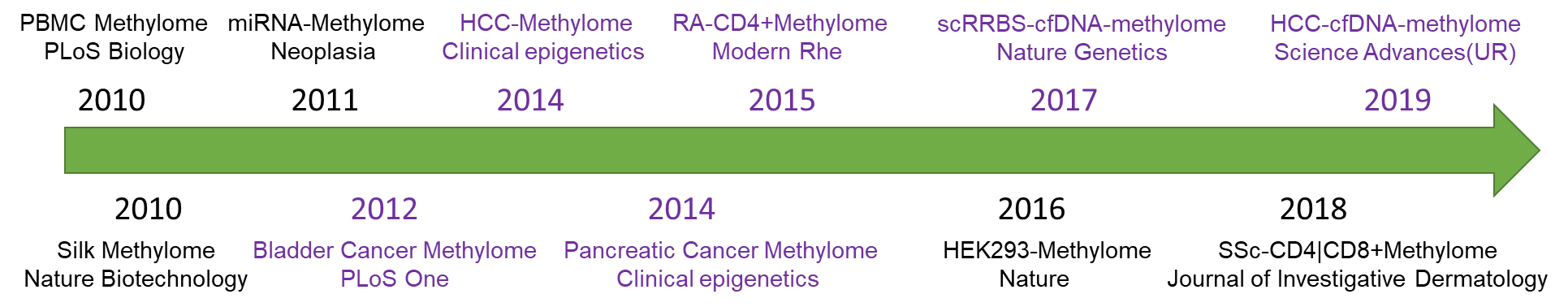
I have conducted a wide array of genetic and epigenetic studies which have utilized both my computational and bench science skills. My experience from these studies has given me the ability to investigate clinically important research questions using novel and multi-faceted approaches.

**Identification of Genetic Susceptibility Genes to Human Complex Diseases.**

Early in my career, leveraging GWAS and candidate gene studies, I identified multiple susceptibility genes, such as a CNV within *HLA-DQA1* and *APOBEC3A/3B* for systemic sclerosis and *FOXE1* for thyroid cancer. I also conducted a series of association studies investigating candidate genetic variants in miRNAs for human cancers. In these studies, miR-4293 was demonstrated to be significantly associated with lung cancer, and[miR-196a2/miR-499](javascript:void(0)) was clearly involved in esophageal cancer. In addition, I implemented a novel approach termed exome-wide gene-based recessive diplotype scanning. Compared with traditional methods, our method demonstrated higher power to identify recessive compound heterozygotes. I have applied the new method to 15 diseases and analyzed genetic data available for 20,000 subjects enrolled in the Personalized Medicine Research Project (PMRP) at the Marshfield Clinic. I identified a susceptibility gene, *FGF6,* for hemochromatosis. By conducting an evolutionary analysis, protein-protein network analysis, and examining both molecular and cellular evidence, we demonstrated *FGF6* plays an important role in iron metabolism which may be involved in multiple conditions underlying diseases associated with iron metabolism. The work was accepted in *Blood* in 2019. The following flowchart shows the timeline for publication of outcomes of my research efforts (Purple indicates 1st author).



**Epigenomic Research and Epigenetic Variations in Diagnosis and Prognosis Models for Complex Diseases.** Beginning in 2009, I investigated the epigenetics of human diseases with a particular focus on DNA methylation. I participated in several large projects to build a model depicting the epigenomic architecture for human cells and tissues under normal and disease conditions. Notable work includes the evaluation of the methylomes for normal human blood cells, animal model ‘silk’, CD4+ T-cells of patients with [rheumatoid arthritis](javascript:void(0)), pancreatic cancer cells, and hepatocellular carcinoma cells with WGBS and MBD-seq. Concurrently, I identified a large number of methylation-based markers with diagnostic and prognostic implications for lung cancer, bladder cancer, and pancreatic cancer. Since DNA methylation has different patterns for different tissue types, we proposed a predictive model to map the origin of cell-free DNA fragments based on tissue-specific methylation signals. This model provides a potentially non-invasive approach for the diagnosis of solid cancers. This work was published in *Nature Genetics* in 2017. Chronology of publication of this research is shown in the following flowchart below (purple indicates 1st author):



**Current Research**

My research in Center for Precision Medicine Research consists of several studies to elucidate genetics of disease with certain non-traditional statistical methods and to identify epigenetic variant-based diagnostic and prognostic biomarkers. The details for these projects are as follows:

**A gene-based recessive diplotype exome scan to identify disease genes for 15 PMRP phenotypes.**

This project was initiated in late 2017 and is supported by Dr. Schrodi’s MCRI grant (I am a Co-Investigator). We planned to map genes predisposing individuals to conditions including iron overload disorders, obesity, type 2 diabetes, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, premature myocardial infarction, and  obsessive-compulsive disorder based on 20,000 exome-chip data. Standard analyses applied in GWAS are well-designed to detect additive effects. However, the power for standard GWAS analyses to identify effects from recessive diplotypes is not typically high. I implemented a novel method named: gene-based recessive diplotype scanning. With this approach applied to iron overload, a strong association signal was identified between the fibroblast growth factor-encoding gene, *FGF6*, and hemochromatosis in the central Wisconsin population. As shown above, this work was published in *Blood, 2019.* Currently, I am working on a manuscript describing the remaining significant findings in the context of type 2 diabetes (*MTNR1B* and *KIF2C*), obesity (*SPTBN5*) and rheumatoid arthritis (*FSTL1*). I have completed all the bioinformatics, molecular function prediction analysis, molecular and cellular validation plans. Working together with collaborators, I have received some exciting biological validation resulting from knock-down and up-regulation experients. Finally, I am working on the manuscript of ”R Package for Compound Heterozygote Scanning to Identify Novel Disease Genes based on Exome-wide Genotype Data” which is near completion and is targeting submission to *Bioinformatics*.

**DNA Hypermethylation Mediated Epigenetic Silenced Diagnostic and Prognostic Biomarkers in Human Cancers.** This project started in early 2017 as a collaboration with Dr. Steven Schrodi and Dr. Minghua Wang. Leveraging my well-trained bioinformatics and data-mining skills, I have conducted a systematic analysis integrating well-known public data including the TCGA project, ENCODE project, Blueprint Epigenome project, Roadmap project, GTEx project (eQTL), 1000 Genome project, gnomAD dataset, and UK-biobank dataset to identify the most interesting differential DNA methylation regions using the ’differential DNA methylation region identification algorithm’ (fDMRI) for human cancers. I applied this method and identified ~50 fDMRs for esophageal cancer (ESCC) and cholangiocarcinoma. These regions are likely to have high utility for use as novel diagnostic and prognostic cancer biomarkers. Currently, biological validation has demonstrated that our method identifies novel interactions between DNA methylation and other genomic functional elements. For example, we found [epigenetic silencing of *ZNF132* mediated by methylation-sensitive Sp1 binding promotes cancer progression in ESCC](https://www.nature.com/articles/s41419-018-1236-z). Results were published in *Cell Death & Disease* in 2018. We will continue to validate the remaining fDMRs and discover the mechanisms surrounding these epigenetic factors in cancers.

**Proposed Research Program**

While human genetics opened the road to precision medicine, genetics is not sufficient to drive a revolution in individualized clinical treatment. Individual variability caused by genetics, epigenetics and environment should be concurrently considered in disease subtype, diagnosis, treatment and prevention. In addition, our understanding of the pathogenesis of complex disease is limited since the variants identified by traditional GWAS studies can only explain very limited proportions of the overall heritability. Epigenetic variation, non-additive but complex modes of inheritance, misdiagnosis, and rare SNP and CNV effects are the most promising solutions to the missing heritability problem. In addition, recently, a series of powerful and low-cost functional assessment and identification tools have been developed such as WES, ATAT-seq, scRRBS, BSPP, CUT&Tag, Fecal-seq, RAD-seq, cfMeDIP-seq which provide the best opportunity to investigate the role of genetic, epigenetic and interactions in the development of complex diseases. With diverse interdisciplinary training in human genetics, biostatistics, computational biology and epidemiology, my research will take advantage of my knowledge of these advanced technologies, which will be applied to the following three fields: 1) developing and implementing new approaches to identifying susceptibility, diagnosis, and prognosis-related genetic and epigenetic variants/biomarkers and novel drug targets for human complex diseases, especially cancers and autoimmune diseases; 2) applying computational and biological approaches for functional assessment of genetic and epigenetic disease variants identified by GWAS, EWAS and PheWAS projects; And 3) developing epigenetic biomarkers based pharmaco-epigenomics (PeGx) approaches that complement PGx to guide personalized medicine and treatment.

**Development and implementation of novel molecular diagnostic and prognostic approaches to human cancer.**

DNA methylation has been demonstrated to be one of the most promising diagnostic, prognostic and pharmaco-epigenomics biomarkers for human complex disease. This may be attributable to the fact that DNA methylation is partially stable and partially dynamic, compared to genetic variation (completely stable) and mRNA (highly dynamic). DNA methylation is involved in transcriptional regulation and therefore plays critical roles in differentiation, development and disease. Given its regulatory roles, DNA methylation changes usual occur earlier than other classes of molecular variation. A large number of DNA methylation-based diagnostic and prognostic biomarkers have been identified, such as *SEPT9* and *SHOX2* which have been approved by FDA for colon cancer and early lung cancer screening. However, DNA methylation diagnostic biomarkers for other cancer types remain to be defined. In my previous publication *(Nature Genetics*, 2017), I have demonstrated non-invasive cell-free DNA methylation haplotype-based tissue-of-origin prediction could be applied in cancer diagnosis. However, the study has a small sample size (N=134) and cancer types (N=2 includes lung cancer and colon cancer) and traditional statistic methodology. In this project, I will make full use of the abundant cancer samples in PMRP (cancer types >10 and sample size>400). Meanwhile, many plasma are collected before the diagnosis which provides me an opportunity to identify early biomarkers. Furthermore, I will integrate genetic (cancer risk alleles which have been generated for PMRP samples, somatic mutations), and other informative variables, such as cell-free DNA fragment distribution, metabolites in plasma. Finally, I will develop a multi-omics data based prediction platform with deep learning (artificial neural networks) based artificial intelligence approach. Following sufficient preliminary results, I will seek NIH funding for the expansion of this study.

**Functional assessment of human genetic and epigenomic variants with computational and biological approach.**

In the past decades, GWAS, EWAS and PheWAS have identified hundreds of significant disease-associated genetic and epigenetic loci. However, there is an enormous gap between the statistically- significant associations and human phenotypes. Functional assessment of genetic and epigenomic variants will provide the opportunity to advance understanding of the functional significance of these loci and how they contribute to disease risk and the pathogenesis of complex diseases. As the accumulation of multi-dimensional data including genetic, gene expression, epigenetic, eQTL, pathway and drugbank database, computational approach could provide an inspiring solution to identify the disease causal variants and to design the most efficient functional discovery approach for the variants. In this study, computational assessment will be conducted leveraging publically available databases and tools such as ENCODE, GTEx, ExAC, FANTOM, Roadmap, Blueprint, RegulomeDB, BioGPS, STRING, Reactome and KEGG pathway. In my previous publication (*Blood*, 2019), I have demonstrated that comprehensive computational and evolutionary analysis could provide highly efficient functional, protein structure-related, interactive network prediction and cell-of-origin for understanding the pathogenic role of candidate genes and variants. To date, I have conducted these analyses on 15 phenotypes of 20,000 PMRP samples and identified several very interesting candidates, such as *FSTL1* in rheumatoid arthritis, *MTNR1B* in type 2 diabetes, and *SLX4* and *SPTBN5* in morbid obesity and the full functional predictions have been completed. In the next stage, I will validate these discoveries collaborated with scientists in UW-Madison. I will apply gene editing approach to investigate the function for *FSTL1*, *SLX4* and *SPTB5* in the corresponding disease traits to confirm how these risk alleles are involved in the disease pathway and disease phenotypes. External funding will be sought to support these efforts and look forward to building synergistic collaborations with UW-Madison investigators.

**Phenome-wide association study of genetic variation in epigenetic factors.** This project is a collaboration with Dr. Steven Schrodi (Marshfield Clinic and Laboratory of Genetics, UW-Madison) and Dr. Mark Craven (Department of Biostatistics and Medical Informatics, UW-Madison) that is currently in the initial stages as part of the CIBM training program supported by UW-Madison and National Library of Medicine (NLM). Human complex disease is generated by the interaction between genetics, epigenetics and the environment. While the rationale for genetic association studies have been supported by different fundamental observations such as heritability estimates from twin studies, there is no fundamental research to illustrate whether epigenetic changes are involved in disease heritability, although we know that epigenetic elements are an important interface between genetics and the environment. In this study, I hypothesize that genetic variants in epigenetic genes are a proxy to infer the epigenetic involvement in phenotypes. We will apply a phenome-wide association study (PheWAS) approach to test the association between a panel of epigenetic factors against 6,221 clinical traits within the PMRP dataset. This will enable us to identify all the significant phenotypes whose pathology are potentially driven by epigenetic changes and apply the measurement of genome-wide DNA methylation levels in the corresponding phenotypes to validate the above findings. Such studies have not been previously conducted and the results of this work will provide insight into epigenetic architecture underlying important clinical traits. If offered the position, I anticipate transitioning this study to a K-award mechanism.

**Developing pharmaco-epigenomic (PeGx) biomarkers to improve PGx-based treatment.** Genetic variation in drug-metabolizing enzymes and transporter (DMET) genes modifying drug response is the main focus of pharmaco-genetics (PGx) laboratories. Multiple studies have shown that SNP-based, even on a genome-wide scale, is only partially informative for the prediction of drug response (typical AUC<0.8). Hence, transcriptional regulation of DMET genes may also play an important role in drug response, drug resistance and adverse drug events (ADE). In my previous studies, I have developed and implemented multiple epigenetic approaches to investigate how DNA methylation plays a role in drug response. My hypothesis is that epigenetic status could significantly improve prediction of drug response. In this study, I plan to apply epigenetic approaches to identify blood-based DNA methylation biomarkers for Methotrexate (MTX) is one of the chemotherapy drugs for cancer treatment and also the gold standard dug in rheumatoid arthritis (RA) which is widely used by 350 million RA patients. Even though several polymorphisms have been identified to be associated with MTX efficacy and toxicity, the outcome prediction performance is very disappointed. In this study, I will collaborate with UW-Madison scientists to conduct a prediction system based on genetics and epigenetics of 66 very important pharmacogenes (PharmGKB). Similar research strategy can be applied in other drugs whose samples are easy to collect in UW-Madison. My aim is to achieve a further paradigm that informs advancement of precision medicine.