**Statement of Research Interests**

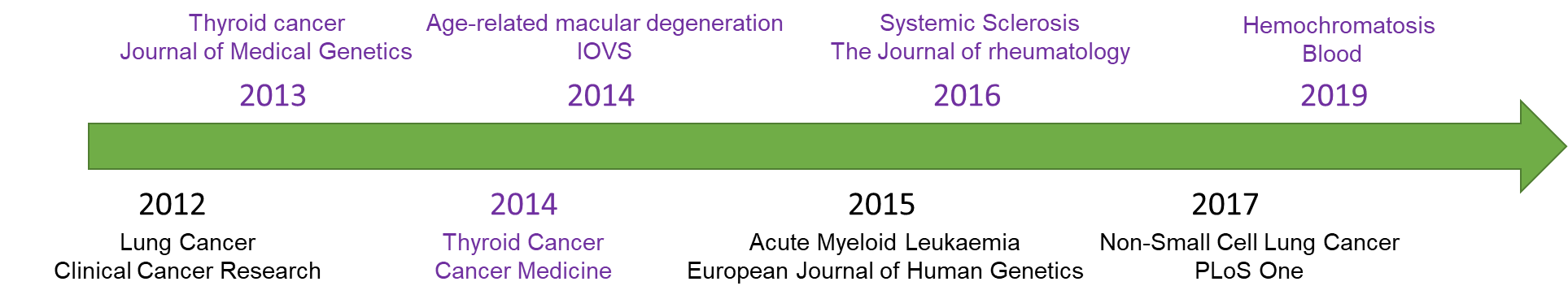
**Shicheng Guo, Ph.D.**

Human complex diseases dynamically and progressively driven by sequences of genetic-epigenetic-environment interactions. Each of these components contributes to bring phenotypes from a susceptibility stage to clinical presentation. The most exciting ideas of precision medicine require scientists to consider numerous susceptibility genetic variants as the baseline risk and to track the epigenetic changes in pathogenesis of human diseases. Taking rheumatoid arthritis (prevalence=0.5% in US) as an example, individuals carrying HLA-DRB1\*04:01 together with triggering inflammation and citrullination caused by factors such as smoking is thought to initiate an autoimmune reaction. Early molecular and epigenetic biomarkers such as rheumatoid factor (RF), anticyclic citrullinated peptide (anti-CCP) and circulating cell-free DNA methylation (cfDNAmeth) signals caused by synovial cell apoptosis occurred before the symptoms. In the past 8 years, I have been working to identify genetic and epigenetic variants/biomarkers in multiple human complex diseases and authored more than 50 SCI publications, 17 of which are first/co-first/co-corresponding author publications. The abundant experience in human genetics, epigenetics and bioinformatics skills has prepared me to conduct comprehensive research studies in precision medicine. My future research will focus on 1) developing and implementing circulating cell-free DNA based genetic and epigenetic biomarkers for early diagnosis, real-time monitoring and prognosis prediction. 2) Discovering functional mechanisms to the genetic and epigenetic variation identified by GWAS, pheWAS, EWAS studies. 3) Developing epigenetic (DNA methylation) based biomarker (pharmaco-epigenetics, PeGx) to cooperate PGx to guide personalized treatment and precision medicine.

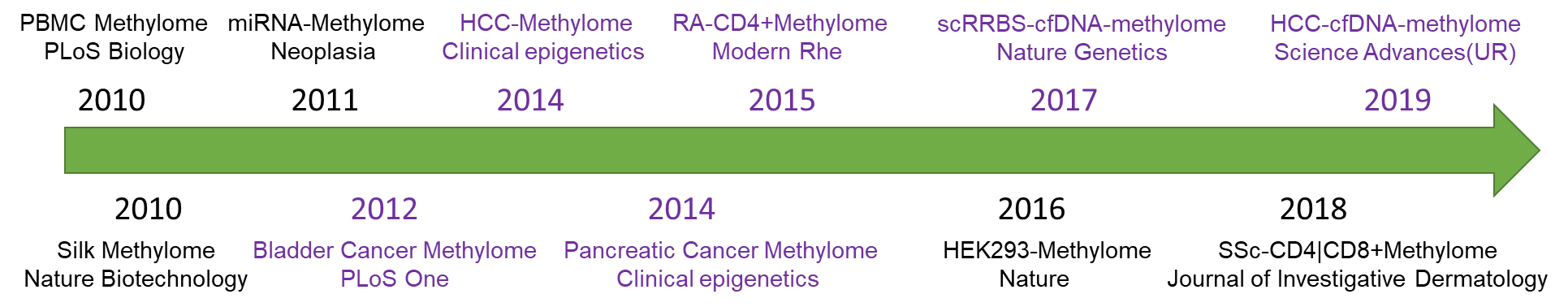
**Research Background**

I have conducted a wide array of genetic and epigenetic studies which are required both computational and wet-lab works. The 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 study, I identified multiple susceptibility genes, such as CNV within *HLA-DQA1* and *APOBEC3A/3B* for SSc, *CFH* for age-related macular degeneration, and *FOXE1* for thyroid cancer. I also conducted serials of association study interrogating genetic variants in miRNA for human cancer. In these study, miR-4293 was demonstrated to be significantly associated with NSCLC, and[miR-196a2/miR-499](javascript:void(0)) involved in [ESCC](javascript:void(0)). In addition, I implemented a novel approach termed exome-wide gene-based recessive diplotype scanning. Compared with traditional method, our method has higher power to identify recessive compound heterozygotes. I apply the new method to 15 diseases from 20,000 Marshfield Clinic PMRP data. I identified a susceptibility gene FGF6 for hemochromatosis. Together with evolution analysis, protein-protein network, molecular and cellular evidence, we demonstrated FGF6 play important role in iron metabolism which may be involved in multiple iron related diseases. The work was accepted in Blood, 2019. These works were shown in the following flowchart (purple indicates 1st author)



**Epigenomic Research and Epigenetic Variations in Diagnosis and Prognosis Models for Complex Diseases.** Starting from 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 of the epigenomic architecture for human cells and tissues under normal and disease conditions. Notable work includes evaluating the genomic methylation profiles (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 different methylation methods, such as BS-seq 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 has been published in *Nature Genetics* in 2017. These works were shown in the following flowchart (purple indicates 1st author)



**Current Research**

My research in Center for Precision Medicine Research is composed of several studies to elucidate disease genetics 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 started from late 2017 and supported by Dr. Schrodi’s MCRI grant. We planed to map disease genes in iron overload disorders, obesity, type 2 diabtetes, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, premature myocardial infraction, and obsessive compulsive disorder using a novel approach of gene-based recessive diplotype analysis 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. 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 I mentioned above, this works have been published in *Blood, 2019.* Currently, I am working on the remaining findings in type 2 diabetes (*MTNR1B* and *KIF2C*), obesity (*SPTBN5*) and rheumatoid athritis (*FSTL1*). I have completed all the bioinformatics, molecular function preditcion works, molecluar and celluar validation plans. Working together with my collabortors, I have received some exciting biological validation result throught knocking-down and up-regulated experients. Finally, I am working on the bioinformatics manuscript of ”R Packge for Compound Heterozygote Scanning to Identify Novel Disease Genes based on Exome-wide Genotype Data” which is almost done and hope to be published in *Bioinformatics*.

**DNA Hypermethylation Mediated Epigenetic Silenced Diagnostic and Prognostic Biomarkers in Human Cancers.** This project started in early 2017 and collabrate with Dr. Schrodi and Dr. Minghua Wang. Leaveraging my well-trained bioinformatics and data-mining skills, I have conducted a systemic analysis integrating most well-known public data such as TCGA project, ENCODE project, Blueprint Epigenome project, Roadmap project, GTEx project (eQTL), 1000 Genome project, gnomAD dataset, UK-biobank dataset to identify most interesting differential DNA methylation region identification algorithm (fDMRI) for human cancers which would be considered as important and novel diagnotic and prognositc biomarkers. I applied this method and identified ~50 fDMR for esophageal squamous-cell carcinomas (ESCC) and Cholangiocarcinoma (CCA). Currently, biological validation demonstrated that our method could identify 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) which have been published in Cell Death & Disease in 2018. We will be continue to validate the remaining fDMRs and discover the mechanisms that epigenetic factors in the human cancers.

**Genetics and Epigenetics Interaction and Precision Medicine in PMRP cohort**. Precision medicine require the cooperation beween genetics and epigenetics which provides different informations from two different dimentions. Recently, inexpensive gentopying or sequencing technologies are now being introduced to identify genetic variants. For example, MCRI already have genotype data for more than 20,000 samples in PMRP cohort. However, epigenetic data for PMRP has not yet been generated—a situation which impacts the progression of precision medicine. In order to integrate genetic and epigenetic variants for better precision medicine and for better phenotype prediction. With the help from CIBM program (supported by UW-Madision, NLM and MCRI), I generated genome-wide DNA methylation profiles for 4 phenotypes (N=100 samples) which are signficant associated with genetic variants occured in DNMT1/3a/3b genes. We have sent all the samples to University of Chicago for genome-wide DNA methylation profiling with illumina methylation EPIC array. In this project, 1) I will investigate the influence of epigenetics from genetic variants of DNMT genes. 2) Estimate the methylation based biological age and investiage biological age among different phenotypes. 2) estimate T-cell subtype propotion and compare them among different phentoypes. 3) We will also compare the methylation based biological age with telomere length since all the telomere length information for PMRP have been generated before by Dr. Scott J Hebbring.

**Proposed Research Program**

While human genetics opened the road to precision medicine, genetics is not enough for the achievement of precision medicine. Individual variability caused by genetics, epigenetics and environment should be concurrently considered in disease subtype, diagnosis, treatment and prevention. In addtion, our understanding to the pathogenesis of complex disease is limited since the variants identified bt 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 serials 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 provided 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 to these advanced technique on the following three fields: 1) developing and implementing new approach to identify susceptibility, diagnosis, prognosis related genetic and epigenetic variants/biomarkers and novel drug targets for human complex diseases, especially cancers and autoimmune diseases. 2) Functional assessment of genetic and epigenetic diseases variants identified by GWAS, EWAS and pheWAS projects by computational and biological approaches. 3) Developing epigenetic variants or biomarkers based pharmaco-epigenomics (PeGx) to cooperate PGx to guide personalized medicine and treatment.

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

DNA methylation has been demonstrated to be one of the most promsing 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 regulation 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 earlty lung cancer screening. However, DNA methylation diagnosis biomarkers for other cancer types are still blank. In my previous publication (Nature Genetics, 2017), I have demenstrated non-invasive cell-free DNA methylation haplotype based tissue-of-origin prediction could be applied in cancer diagnosis. With my expertise on DNA methylation, I will conducted a novel cel-free DNA based non-invasive cancer diagnosis and drug response prediction study. In this project, I will integrate genetic (cancer risk allele, somatic mutation), epigenetic (DNA methylation) and other informative variables, such as cell-free DNA fragment distribution, metabolics in plasma to develop a systemic prediction platform with artificial intelligence (netural network). Meanwhile, I will take advantage of the preliminary data and samples in MCRI where I have collected. Since MCRI have colllected blood, plasma and sera samples since 2005, huge number of samples were collected before the diagnosis which will be valuable to evaluate the earliest occurence time of the epigenetic/matebolic signal for the progressive diseases such as cancer and autoimmnue diseases. In summary, I will prepare a RO1 proposal to implement this study with collaboration MCRI and UW-Madison dedicate scientists.

**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 signficant diseases associated genetic and epigenetic loci. However, there is a huge gap between the statistical signficant associations linking locus and human phenotypes. Functional assessment of genetic and epigenomic variants will provide the opportunity to functional understanding of the biology underlying disease risk and pathology of the complex disease. In this study, I will propose serials of functional assessment study to these human genetic and epigenomic variants with both computational and biological approach simultaneously. In the first stage, computational assesement will conducted with the assistant from different public available database or tools such as Encode, GTEx, ExAC, FANTOM, Roadmap Epigenomics, LINCS, Blueprint, RegulomeDB, BioGPS, STRING, Reactome pathway, KEGG, ANNOVA, VEP. In my previous publication (Blood, 2019), I have demenstrated comprehensive computational and evolutional analysis could provide highly efficient functional, protein structural and interactional network prediction to candidate genes and variants. In the second stage, I will apply biological approach such as CRISPR-Cas9 to validate the hypothesis generated with above method. Actually, I have conducted these analysis to 15 phenotypes to 20,000 PMRP samples and identified several very interesting candidates, such as *FSTL1* in rheumatoid arthritis, *MTNR1B* in type 2 diabetes, SLX4 and SPTBN5 in obeisty. In summary, I will prepare a RO1 or KL2 proposal to continue these study with collaboration with UW-Madison scientists.

**Phenome-wide association study of genetic variation in epigenetic factors.** This project is a collaboration with Dr. Steven Schrodi and Dr. Mark Craven as part of the Computation and Informatics in Biology and Medicine (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. This project will also feature a collaboration with Dr. Scott J Hebbring. We will share the DNA methylation dataset with Dr. Hebbring so that he will investigate the relationship between aging, telomeres length and genome-wide DNA methylation. To date, there has not been this type of study conducted and the results of this work will provide insight into epigenetic architecture underlying important clinical traits. I will prepare a RO1 proposal to continue these study with collaboration with UW-Madison scientists.

**Developing epigenetic variants or biomarkers based pharmaco-epigenomics (PeGx) to cooperate PGx to guide personalized medicine and treatment.**  Genetic variation in drug-metabolizing enzymes and transporters (DMET) genes in relation to drug response has been the main focus of pharmaco-genetics (PGx) laboratories. However, transcriptional regulation of DMET genes also played important roles in drug response, drug resistance and side effects. In my previous studies, I have developed and implemented multiple epigenetic approach to investigate genome-wide DNA methylation and gene-panel based methylation profiling, such as single-cell reduced-representation bisulfite sequencing (scRRBS), methylation status determined single nucleotide primer extension technique (MSD-SNuPET), Multiplex PCR targeted bisulfite sequencing (MTBseq). Multiple studies have shown that even we included genome-wide SNPs, the drug response prediction performance is still quite limited (AUC<0.8). Actually, genetic background only provides a baseline variable to predict drug response and adverse event. My hypothesis is epigenetics status could provide recent signals for better prediction. In this study, I plan to apply epigenetic approach to identify blood based DNA methylation biomarkers to cooperate PGx to achieve another paradigm for precision medicine.