Dear Prof. Craven,

I am currently a postdoctoral research fellow in Steve Schrodi’s lab at the Marshfield Clinic (2017-Present). I am writing this letter to ask if you would be willing to serve as my co-advisor for the Computation and Informatics in Biology and Medicine (CIBM) Training Program. Dr. Schrodi highly recommends you and your expertise in machine learning and informatics is extremely complementary to my work. I finished my Ph.D. training at Fudan University (Shanghai, China) in 2015 and completed a 2 year postdoc training at University of Texas Health Science Center in Houston (UThealth) and UCSD doing work on genetic epidemiology and human epigenetics in cancer and autoimmune diseases. I have more than 6 years’ Perl and R programming experiences as well as [portable batch system (PBS).](https://en.wikipedia.org/wiki/Portable_Batch_System)

My proposed CIBM training project is “Phenome-wide Association Study Maps Genetic Variation in Epigenetic Factors with Human Complex Disease”. Massive evidences showed complex diseases are caused by the interaction between genetics and epigenetics. However, the effect of the genetic and epigenetic interaction on different phenotypes have not been widely evaluated. In this project, I will investigate all the genetic variants within 250 human epigenetic-related proteins and correlate with 6,221 phenotypes in Marshfield Clinic Personalized Medicine Research Cohort (PMRP). In the PMRP dataset, we have exome-chip data for 18,000 individuals with comprehensive clinical and demographical information (including > 6000 phenotypes). I want to apply different statistical and genetics methods, such as genotype-based and gene-based association methods across different genetic models (dominant, additive, recessive and multiplicative) to identify the phenotypes, which are significantly associated with the panel of epigenetic factor variants. These data will allow us to infer genetic-epigenetics interactions across the phenotypes. Subsequently, I will validate our findings with target bisulfite sequencing to measure methylation patterns that may contribute to the most significant clinical phenotypes.

In this project, medical information, machine learning and big data analysis approaches will be widely used, therefore, I really hope you can be my co-advisor to help me complete this project.

I have attached my CV. I hope this is helpful for your evaluation.

Thank you for your consideration,

Shicheng Guo