

**Annual Report 2018-2019**  
**Dave House Endowed Professorship**  
**in Statistics, Data Mining and Data Analytics**

**Kui Zhang**

**November 26, 2019**

This is 2018-2019 annual report to describe the research activities that were supported by the funds from Dave House Endowed Professorship in Statistics, Data Mining and Data Analytics. I would like to thank the generous endowment from Mr. Dave House. His continuous support to Michigan Technological University and to Department of Mathematical Sciences has significantly improved the learning experience of graduate students in my department.

Between September 2018 to November 2019, the funds generated from this endowment were solely used to pay the stipend and tuition and part of travel expense for Cheng Gao, a PhD student in statistical in Department of Mathematical Sciences. With this fund, Mr. Gao was able to focus on his research and was able to present his work in a national meeting.

The methodological research interests of my group focus on the development of novel statistical methods and efficient computational and bioinformatics tools to address scientific problems in biomedical fields, especially in statistical genetics and genomics. Specifically, these include developing new methods and designing novel algorithms for analyzing data from genetic studies and finding genetic variants and environmental factors that are associated with complex human diseases, such as diabetes, hypertension, obesity, etc. The collaborative research interests of my group are to apply powerful and innovative statistical and computational methods to address problems from genetic studies and other types of studies in biomedical research. We hope our research will ultimately lead to the better preventions, diagnoses, and treatments for human diseases.

During the time between September 2018 to November 2019, Cheng Gao focused on developing a statistical method that can detect the association between multiple genetic variants and multiple trait using family data. Mr. Gao was able to implement the developed method to a R program, conduct simulation studies to evaluate its performance, and applied it to The Genetics of Kidneys in Diabetes (GoKind) study to illustrate its usefulness. The results show that our method outperforms several available methods in terms of power. Mr. Gao presented his work in the annual meeting of American Society of Human Genetic in October 2019. He has prepared a manuscript titled “MF-TOWMuT: Testing an Optimally Weighted Combination of Common and Rare Variants with Multiple Traits Using Family” based on his research work and we expect to submit it for publication in next month.

During the summer and this fall semester, Cheng Gao conducted the research work on polygenic risk score (PRS). The polygenic risk score of a trait is a number based on genetic variants and their association with the trait. Polygenic risk scores are widely used in animal, plant, and human to predict the traits when multiple genetic variants are taken into account. Mr. Cheng illustrated that the use of multiple polygenic risk scores can improve the dissection of sub-phenotypes of bipolar disorder. This work was summarized in a manuscript titled “Dissecting sub-phenotypes of bipolar disorder using multiple polygenic risk scores” which has been submitted for publication. In addition, Mr. Gao is working on two manuscripts that are related to the polygenic risk score. One is the comparison of genetic risk prediction approaches in the presence of heterogeneity. The other is the comparison of methods for constructing the polygenic risk score using comprehensive real data. Both manuscripts will be submitted for publication in the next few months.

As the holder of Dave House Endowed Professorship in Statistics, Data Mining and Data Analytics, I would like to take this opportunity to highlight some of my achievements during 2018 to 2019. With respect to teaching activities, I have graduated a MS student in statistics in December 2018, is currently advising three PhD students, have taught three courses, and is developing an online course, Design and Analysis of Experiments for the MS program in Applied Statistics. With respect to research activities, I have published one journal article and submitted one manuscript for publication, has submitted one grant as co-PI to NSF ABI Innovation (not funded), one pre-proposal to DOE (not selected), one R15 grant as co-investigator to NIH with a potentially fundable impact score of 27, and one MTU internal grant as co-PI to Portage Health Foundation (will be funded). With respect to service activities, I have served as a member of the departmental graduate committee, have served as an ad-hoc reviewer of NIH BMRD study section in February 2019.

In next year, I am planning to use the funds to continuously pay the stipend and tuition and part of travel expense for Cheng Gao. Mr. Gao will continue on his work on statistical genetic and genomics. He is a bright and hardworking scholar and has the potential to make an impact in his research area. I will identify and support an additional PhD student in statistics if there are sufficient funds left after I support Cheng Gao. Again, I would like to thank the generous support from Mr. Dave House. Without his continuous support, it would not be possible to provide such wealth of learning opportunities to our students.