Haotian Zheng

211 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, USA

(+1) 215-252-6728 | haotianz@pennmedicine.upenn.edu

EDUCATION

Ph.D. in Biostatistics   *Aug 2020 – Present (Expected 2023)*

University of Pennsylvania   Philadelphia, PA

➢ Advisor: Hongzhe Li, Ph.D.

M.S. in Biostatistics   *Aug 2018 – May 2020*

University of Pennsylvania   Philadelphia, PA

➢ Advisor: Hongzhe Li, Ph.D.

➢ Thesis: Proteomics association analysis and risk prediction for cardiovascular disease in CRIC cohort

➢ Passed Doctoral Qualifications Examination in June 2019 (1st Place among Ph.D. and Master's students)

B.S. in Mathematics   *Aug 2014 - Jul 2018*

Tsinghua University   Beijing, China

➢ Advisor: Ke Deng, Ph.D. and Ying Yang, Ph.D.

➢ Thesis: Statistical Methods for Single-Cell Classification Based on Gene Sequencing

AWARDS

➢ Graduate Student Travel Award (Biomedical Graduate Studies)   *2022*

➢ Graduate Student Teaching Award   *2021*

   Best student teaching assistant at Graduate Group of Epidemiology and Biostatistics, University of Pennsylvania

➢ Jonathan Raz Award   *2019*

   Best qualifying exam at Graduate Group of Epidemiology and Biostatistics, University of Pennsylvania

➢ 1st Prize at 2016 China Undergraduate Mathematical Contest in Modeling   *2016*

➢ 1st Prize at the 32nd National Undergraduate Physics Competition   *2015*

PUBLICATIONS

➢ Li, H. and **Zheng, H.** (2020): Statistical and Computational Methods for Analysis of Shotgun Metagenomics Sequencing Data. Statistical Analysis of Microbiome Data, edited by Somnath Datta and Subharup Guba, Springer [link].

➢ Zhan, X., Xue, L., **Zheng, H.**, Plantinga, A., Wu, M., Schaid, D., Zhao, N. and Chen, J. (2018): A small-sample kernel association test for correlated data with application to microbiome association studies. *Genetic Epidemiology.*, *42*(8), 772-782. Sep. 2018 [link].

➢ Deo, R., Dubin, R., Ren, Y., Murthy, A.C., Wang, J., **Zheng, H.**, Zheng, Z., Feldman, H., Shou, H., Coresh, J., Grams, M., Surapaneni, A., Bhat, Z., Cohen, J., Rahman, M., He, J., Saraf, S., Go, A., Kimmel, P., Vasan, R., Segal, M., Li, H. and Ganz, P. (2022+): Proteomic Risk Assessment for Incident Cardiovascular Disease among Individuals with Chronic Kidney Disease. *European Heart Journal*, under review.

➢ **Zheng, H.**, Li, S. and Li, H. (2022+): Transfer Learning in High Dimensional Linear Discriminant Analysis. Ready for submission.

- ➢ **Zheng, H.**, Li, S. and Li, H. (2022+): Transfer Learning Using Summary Statistics with Applications to Polygenetic Risk Score Prediction. Ready for submission.
- ➢ Saiman, Y., Scorletti, E., **Zheng, H.**, Mahmud, N., Liu, Q., Koo, I., Brettschneider, J., Jeon, S., Buyco, D., Lin, C., Friedman, E.S., Ughwanogho, R., Thiele, G., Toal, E., Dwyer, S., Patterson, A.D., Li, H., Wu, G.D. and Carr, R.M. (2022+): Identification of Unique Lipidomic Patters in Patients with NAFLD. Ready for submission.
- ➢ Liu, S., Ehnot, J., Jeffery, G., Jung, S., Menna, T., **Zheng, H.**, Purton, L.E. and Kurre, P.: Clonal Diversity As an Alternative Measure for Hematopoietic Reserve in Bone Marrow Failure. *American Society of Hematology 2022 Annual Meeting*. Submitted

SOFTWARE

- ➢ **Haotian Zheng**, Xiang Zhan, Anna Plantinga, Michael Wu and Ni Zhao (2017). MiRKAT: Microbiome Regression-Based Kernel Association Test. R package version 1.0. Jul. 2017. [link]

RESEARCH EXPERIENCES

University of Pennsylvania | Dissertation Research   **Philadelphia, PA**

*Advisor: **Dr. Hongzhe Li**, Professor, Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania*   *Sept 2018 - Present*

Theoretical works:

- ● **Theoretical and methodological research with applications in High Dimensional Genetics and Genomics Studies**
- ➢ Developed novel transfer learning methods for linear discriminant analysis (TransLDA) and for linear regression using summary statistics (TLASS) that effectively utilize information from auxiliary data.
- ➢ Proved transfer learning methods have better estimation and classification rule/polygenic risk score (PRS) prediction for the target study.
- ➢ Evaluated the proposed methods using simulations and multiple datasets including gut microbiome data for colorectal cancer classification and genotypes data for blood related phenotypes PRS prediction.
- ➢ Developed novel high-dimensional ridge regression models where we only observe proxy data (marginal statistics and external sample covariance matrix). Evaluated the method using simulations with real genotypes data.

Applied projects:

- ● **Proteomics association analysis and risk prediction for cardiovascular disease in CRIC cohort**
- ➢ Conducted elastic net for Cox regression and identified protein markers that are associated with incident cardio-vascular disease (CVD) in Chronic Renal Insufficiency Cohort (CRIC) study cohort.
- ➢ Built a risk prediction model using Cox proportional hazards model with protein markers for incident CVD, evaluated with Kaplan-Meier survival curves.
- ● **PenNSAM metabolomics study of NAFLD**
- ➢ Performed cluster analysis for large-scale metabolomics data using sparse hierarchical clustering, t-Distributed Stochastic Neighbor Embedding (t-SNE) and multiple dimension scaling (MDS) plots.
- ➢ Conducted Lasso logistic regression used R program glmnet to identify the lipid profile and to build the predictive model for NAFLD.
- ➢ Conducted GWAS using PLINK to find genetic relations between lipidomics and genotypes, aiming to find

the mediation causality between genotypes and NAFLD.

Collaborative works:

➢ **Food and Resulting Microbiota and Metabolite (FARMM) study**: Developed a statistical model to classify longitudinal alterations in the abundance of fecal metabolites to specify metabolic functions contributed by the gut microbiome. Detected that the trajectory of metabolite alteration was very different for one diet group relative to other groups.

➢ **PennCHOP IGRAM study**: Performed association studies between human microbiome and longitudinal Body Mass Index (BMI) among infants and built a predictive model for the BMI using Random Forrest.

➢ **iFaM Plasma NMR Targeted metabolites study**: Conducted the analysis of relationship between longitudinal metabolites of different sources and diet for mice using Linear Mixed Models, considering living environment as a random effect.

Johns Hopkins University | Summer Intern Research Assistant   **Baltimore, MD**
Advisor: Dr. Ni Zhao   *Jun - Sept 2017*

➢ Built an R package *MiRKAT*: Microbiome Regression-Based Kernel Association Test and published it on the Comprehensive R Archive Network (CRAN).

➢ Developed a novel method of association test for correlated data with application to microbiome association studies, especially small sample longitudinal data.

Tsinghua University | Research Assistant   **Beijing, China**
Advisor: Drs. Ke Deng, Ying Yang and Jun S. Liu   *Oct 2016 - Jun 2018*

➢ Applied the Dirichlet mixture model for clustering droplet-based scRNA-Seq data (DIMM-SC) as a topic model to single-cell clustering problems

➢ Developed a transition model of unstructured Chinese medical texts in the Electronic Medical Record (EMR) with the Hidden Markov Model (HMM)

ORAL PRESENTATIONS

➢ Transfer Learning in High Dimensional Linear Discriminant Analysis

Joint Statistical Meetings 2022 | Washington, DC

ENAR 2022 spring meeting | Houston, TX

TEACHING EXPERIENCE

University of Pennsylvania | Teaching Assistant   **Philadelphia, PA**
Department of Biostatistics, Epidemiology, and Informatics
BSTA 622: Statistical Inference II   Fall 2020 (Best TA Award), Fall 2021

PROFESSIONAL MEMBERSHIPS

➢ American Statistical Association
➢ Eastern North American Region of the International Biometric Society (ENAR IBS)

LEADERSHIP

➢ Vice chair, Tsinghua University Bridge Association
➢ Division leader, Student union at Tsinghua University

SKILLS

- Programming: R, Python, C/C++, SAS, Stata, Matlab, Mathematica, Q-Basic
- Toolsets: PLINK, R shiny, Qt, SPSS, Excel, SQL
- Operating Systems: Unix, Microsoft Windows, macOS
- Documentations: LaTeX, R Markdown, Microsoft Office
- Languages: Chinese (native), English (proficient), Spanish (basic)