Zhengwei Song

Tel: (332) 256-6895 | Email: zs2539@cumc.columbia.edu | Personal Website

EDUCATION

Columbia University Mailman School of Public Health, New York, NY, USA Sep 2022 - May 2024 (Expected)

Master of Science in Biostatistics, GPA: 4.03/4

- Relevant Coursework: (Advanced) Probability, Biostatistical Methods, Epidemiology, Data Sciences, Survival Analysis, Randomized Clinical Trial, Data Mining, Advanced Statistical Computing

University of Manchester, Manchester, UK Sep 2019 - Jul 2021

Bachelor of Science in Mathematics and Statistics, GPA: 3.65/4, 1st honor in class

- Relevant Coursework: Real Analysis, Statistical Inference, Markov Chain, Martingales

Shandong University, Jinan, China Sep 2017 - Jun 2021

Bachelor of Science in Mathematics, GPA: 3.62/4, top 30% in class

- Relevant Coursework: Advanced (Linear) Algebra, Calculus, Geometry

RESEARCH EXPERIENCE

Annie Lee Lab, Columbia University Department of Neurology Mar 2023 – Present

Aim: To broaden gene discovery for Alzheimer's disease (AD) by incorporating additional cardiovascular and cerebrovascular risk factors (CVRFs) and examining the multi-omics profiles of these genes to unravel their mechanisms and pathways

Multi-omics integration via similarity network fusion to detect molecular subtypes of aging, for effective treatments with cerebrovascular factors

- Utilized Random Forest for missing data imputation to ensure the similarity networks were based on comprehensive and representative data
- Integrated RNA-seq (1092 samples, 18629 genes), proteomics (400 samples, 8817 proteins), and DNA Methylation (704 samples, 420132 CpG sites) data by applying Similarity Network Fusion, identifying molecular subtypes of Alzheimer's patients via spectral clustering to comprehensively understand AD's heterogeneity across patient subtypes
- Tested associations betweem cerebrovascular factors (e.g. infarctions in 45 brain regions) and subtypes, extracted 5 statistically and significantly infarct regions to provide idea on more personalized and effective treatments with cerebrovascular factors

Causal Mediation Analysis and Trait Analysis on CVRFs interacted, AD risk altered genes, to quantify the involvement of gene expression in brain pathology in AD

- Adjusted technical variables (e.g. batch) identified by a forward selection approach using the voom/limma pipeline in RNA-seq data to ensure analysis results were not biased due to variations across different technical variables
- Used adaptive gene-environment interaction (aGE) test to test for the genes that interacted with CVRFs to alter AD risk
- Conducted causal mediation and trait analysis to elucidate (causal) relationships between 45 cerebral infarctions, AD, Amyloid-β, tau, and candidate expression (RNA-seq or Proteomics) of aGE tested genes, to identify whether the candidate expression resulted from amyloid or phosphorylated tau deposition or the reverse
- Draft manuscript on mediation analysis which is under review for publication

Underlying Pathology of Pure Vascular Cognitive Impairment

- Divided participants as vascular, neurodegenerative and mixed group according to pathological data to identify participants without neurodegenerative pathologies, according to clinician's evaluation
- In vascular group, tested associations of brain pathologies and cognitive decline and extracted macroinfarcts, specifically in the white matter, was the main cerebrovascular disease pathology associated with cognitive decline

PUBLICATIONS

1. (+ contributed equally) Annie J. Lee, Zuoqiao Cui+, **Zhengwei Song**+, Dolly Reyes-Dumeyer, Philip L. De Jager, David A. Bennett, Julie A. Schneider, Vilas Menon, Yanling Wang, Rafael A. Lantigua, Martin Medrano, Diones Rivera, Ivonne Z. Jiménez-Velázquez, Walter A. Kukull, Adam M. Brickman, Jennifer J Manly, Giuseppe Tosto, Caghan Kizil, Lindsay A. Farrer, Jesse Mez, Jaeyoon Chung, Badri N. Vardarajan, Richard Mayeux. Genome-wide gene-based study in multi-ethnic cohorts identifies genes that interact with vascular risk factors in Alzheimer's Disease. In preparation for *JAMA Neurology*

2. **Zhengwei Song**, Annie J. Lee. Integrated Transcriptomics and Epigenomics Analysis Identifies Molecular Subtypes of Alzheimer's Disease. In preparation for *Neurology*

PRESENTATIONS

Annie J. Lee, Zuoqiao Cui, **Zhengwei Song**, Dolly Reyes-Dumeyer, Philip L. De Jager, David A. Bennett, Julie A. Schneider, Vilas Menon, Yanling Wang, Rafael A. Lantigua, Martin Medrano, Diones Rivera, Ivonne Z. Jiménez-Velázquez, Walter A. Kukull, Adam M. Brickman, Jennifer J Manly, Giuseppe Tosto, Caghan Kizil, Lindsay A. Farrer, Jesse Mez, Jaeyoon Chung, Badri N. Vardarajan, Richard Mayeux. Genome-wide gene-based study in multi-ethnic cohorts identifies genes that interact with vascular risk factors in Alzheimer's Disease. Oral Presentation to be delivered at International Conference on Alzheimer's & Parkinson's Diseases and related neurological disorders (AD/PD), 2024, Lisbon, Portugal

PROFESSIONAL & WORK EXPERIENCE

Biostatistics Intern, Medical Scientific Affairs Dept, Roche Diagnostics (Shanghai, China) Apr 2022 – Sep 2022

- Developed statistical methods for analyzing clinical trial data, including the development of novel approaches to address specific research questions and issues with existing methodologies
- Collaborated with 10 medical team members and provided statistical support in phase 4 clinical trials by co-developing analytical plans, performing analyses, interpreting results, and summarizing findings into concise reports that are understandable to non-statisticians
- Co-developed an R package (jmpost) of linear mixed effects models for the tumor size over time by Bayesian inference using Hamiltonian Monte Carlo method

Data Analyst Intern, Information System Dept, Sina Weibo (Beijing, China) Oct 2021 – Apr 2022

- Scraped & wrangled user data, and created visualization (user portraits) for rankings in the entertainment operations
- Presented final statistics for several popular TV series, variety shows, and documentaries, to provide data support for social media influencers and internal operations
- Maintained Hive SQL and data warehouse services

PROJECT EXPERIENCE

Black-Scholes Pricing Model Data Simulation by Multilevel Monte-Carlo Method 2020 –2021
Prof. Jianliang Chen, Shandong University School of Mathematics

- Executed Monte-Carlo simulations for path-dependent option pricing, using Weiner process models
- Developed R scripts for Asian Option pricing, performed asset path averaging, payoff calculation, and variance reduction in multilevel Monte Carlo methods

Edible Tableware based on Finite Element Analysis 2019 –2020
Prof. Song Yu, Shandong University School of Mathematics

- Led a team of five with diverse academic backgrounds and secured full funding (around $900)
- Designed and produced a chopstick-like mold by SolidWorks software according to finite element analysis theories
- Connected and partnered with local restaurants and bars for testing mechanical characteristics

TEACHING & MENTORING EXPERIENCE

Teaching Assistant, Columbia University 2023

Prof. Todd Ogden, Introduction to Mathematical Statistics

- Hold weekly office hours, grade weekly homework, attend the instructor's flipped class for Q&A to 25 students

Prof. Molei Liu, Biostatistical Methods I

- Hold triweekly office hours, grade monthly homework for 20-30 students

HONORS & AWARDS

AAAI 2022 Security AI Challenger VIII Award (44 out of 3692) 2022
Enactus Annual Outstanding Individual (Top 10%) 2018
Shandong University Third-level Academic Scholarship (Award, GPA top 30% in class) 2018

SKILLS

Software: R (tidyverse, caret, survival, lme4, gee, httr, bioconductor, etc.), SAS, Shiny, SQL, C, MATLAB, Unix, Microsoft Office, AutoCAD

Tests: t, z, ANOVA, chi-squared, Fisher's exact, McNemar's, Log-rank, sign, Wilcoxon signed-rank & rank-sum, etc.

Modeling / Machine Learning: linear, generalized linear (logistic, Poisson), weighted least squares, mixed effect, GEE, survival (Cox, Stratified PH, AFT), decision tree, random forest, boosting, K-NN, cubic splines, local regression, GAM, MARS, LDA, QDA, NB, SVM, clustering (K-means, Hierarchical, spectral, etc.), PCA, LASSO, Elastic net, Ridge, PCR, PLS, cross validations

Simulation & Optimization: Data generation, Newton-Raphson, EM, bootstrapping, Monte Carlo (MCMC, HMC, MLMC)

COMMUNITY SERVICE

Treasurer at Enactus Shandong University 2018 – 2019
Secretary at Shandong University Association of International Exchange 2017 – 2018