Zhongyi (James) Guo

Email: guozy@stanford.edu

Phone: (+1) 607-262-3415

Location: Palo Alto, CA, US

LinkedIn: www.linkedin.com/in/zhongyi-guo/

GitHub: github.com/JG1ANDONLY

Personal Website: jg1andonly.github.io

EDUCATION

M.S. Epidemiology and Clinical Research, Stanford University   (Expected) 06/2025
GPA: 3.94/4.00 (by Spring 2024)

B.S. Biometry & Statistics *and* **Biological Sciences,** Cornell University   05/2023
Concentrations: Computational Biology *and* Statistical Genetics
GPA: 3.57/4.30, Honors: c*um laude*, Dean's List

RESEARCH EXPERIENCE

Graduate Student Researcher, the Kundaje Lab, Stanford   04/2024 – Present
Mentored by Prof. Anshul Kundaje and Dr. Selin Jessa
Project: Understanding Regulatory Genomics in Psychiatric Disorders Using ChromBPnet
- This project aimed to study regulatory genomics of bipolar and schizophrenia at the single-cell level using transcriptomics and chromatin accessibility data, utilizing ChromBPNet, a convolutional neural network (CNN) model that predicts chromatin accessibility at a base-pair resolution.
- Pre-processed the multiome dataset (scRNA-seq and scATAC-seq) using quality control, dimension reduction, annotation with reference, cell clustering, doublet removal, and marker gene detection.
- Performed differential expression analyses (DESeq2, Wilcoxon rank-sum test, MAST) across all cell types within brain regions by sex and disease to integrate into a spatial transcriptomics gene panel.
- Train cell-type specific ChromBPNet models to predict chromatin accessibility in active regulatory regions for each cell type. Will evaluate the effects of genetic variants in regulatory regions of the genome associated with schizophrenia and bipolar using these trained models.

Graduate Student Researcher, the Graff Lab, UCSF (remote)   10/2023 – Present
Mentored by Prof. Rebecca Graff
Project: Exploring Black-White Racial Disparities of Prostate Cancer via LC-MS Metabolomics
- This project aimed to study metabolic differences contributing to Black-White racial disparities in prostate cancer using Liquid Chromatography-Mass Spectrometry (LC-MS) metabolomics using R.
- Discovered acyl carnitine derivatives and dicarboxylates as key contributors through chemical similarity enrichment analysis (ChemRICH) by designing and implementing three methods: sub-pathway information, correlation modules, and predicted Medical Subject Headings (MeSH) classes.
- Found corroborative contributors using Principle Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), Random Forest, Support Vector Machine (SVM), quantitative set enrichment, and pathway analysis using *MetaboAnalystR*.
- Identified up-regulated carnitine derivatives and dicarboxylic acids influenced by aging through differential expression analysis using *limma*, after successful network construction and module detection using WGCNA based on pairwise correlations, soft-thresholding, and topological overlap.
- Replicated and validated t-test results (including fold changes of mean concentrations, p-values, and FDR q-values) reported by Metabolon, Inc.

PUBLICATIONS

[1] Lu, J., **Guo, Z.,** & Rehkopf, D. H. (2024). *Towards Causal Interpretation of Sexual Orientation in Regression Analysis: Applications and Challenges.* arXiv preprint arXiv:2405.02322. preprint

[2] **Poster Session Presenter/First Author,** *Causal effect of type II diabetes on prostate cancer in the East Asian population: A two-sample Mendelian randomization study*, AACR Special Conference: Aging and Cancer, 2022. published

CLASS PROJECT EXPERIENCE

Sequence Design for Dopaminergic Neurons in the Mammalian Midbrains   03/2024 – 06/2024
Deep Learning in Genomics and Biomedicine, advised by Prof. Anshul Kundaje and Prof. James Zou
- Generated synthetic enhancers for dopaminergic (DA) neurons to develop potential gene therapy solutions for Parkinson's disease, utilizing sequence evolution (Ledidi), motif implantation, and Generative Adversarial Network (GAN) techniques in collaboration.
- Maximized chromatin accessibility in DA neurons and minimized it in glutamatergic cerebellar neurons (GB Glut, neuronal) and astrocytes (non-neuronal) using Ledidi and motif implantation.

Integration of Single-cell Multi-omics Data for Alzheimer's Disease   03/2024 – 06/2024
Translational Bioinformatics, advised by Prof. Sylvia Plevritis
- Conducted an extensive literature review on multi-omics integration at a cell-type-specific level, and conducted preliminary integration and analysis of epigenomic, transcriptomic, and proteomic datasets using MOFA+, targeting on microglial cells in human brains.
- Composed the Specific Aims section of a theoretical NIH grant proposal in collaboration.

Does Oscar Winning/Nomination Affect Longevity? A Survival Analysis.   03/2024 – 05/2024
Intermediate Biostatistics: Regression, Prediction, Survival Analysis, advised by Prof. Kristin Sainani
- Concluded from statistical evidence that winning or being nominated for an Academy Award does not affect longevity, comparing winners, nominees, and controls (never nominated) using R.
- Constructed three Cox regression models, validated linearity of all variables using Martingale residual plots, and assessed proportional hazard assumptions via log-log and Schoenfeld residuals, including interaction term testing, through collaboration.

GWAS Study: Analysis of Lymphoblastoid Cell Lines (LCL) mRNA Levels   04/2022 – 05/2022
Quantitative Genomics and Genetics, advised by Prof. Jason Mezey
- Analyzed genotype and phenotype data and tested whether population and sex as two covariates could influence the GWAS result and cause differences in LCL mRNA level expressions.
- Identified significant SNPs from Manhattan and QQ plots and causal polymorphisms.
- Concluded that population and sex as two covariates do not impact the GWAS result significantly.

GWAS Study: Analysis of Citrulline Levels and Chronic Kidney Disease   05/2022
Quantitative Genomics and Genetics, advised by Prof. Jason Mezey
- Performed GWAS analysis on citrulline levels and chronic kidney disease data using two PCs obtained from PCA as covariates on genotype data, and Bonferroni correction to reduce Type I error.
- Identified 2 significant SNPs from the Manhattan plots with 2 covariates included and interpreted the influence of linkage disequilibrium based on the result.

TEACHING EXPERIENCE

Beta Tester and Teaching Assistant, Introduction to Data Science, Cornell   01/2023 – 05/2023
• Led discussions, assessed homework and exams, held office hours, and proofread solutions.

Grader, Probability Models and Inference, Cornell   08/2022 – 12/2022
• Assessed homework and exams.

Teaching Assistant, Laboratory in Genetics and Genomics, Cornell   01/2021 – 05/2021
• Created and stabilized knockout mutations on target gene of fruit flies using CRISPR/Cas9.
• Assisted with designing and cloning primers with sgRNA and guided 20 students in analyzing
  mutations vs. wildtype on the UCSC Genome Browser and in locating sgRNA transgenes.

Teaching Assistant, Introductory Biology and Physics, JNC Study Abroad Platform 07/2022 – 08/2022
• Led discussions, assessed homework and exams, held office hours, and proofread solutions.

SERVICE

Epidemiology M.S. Student Ambassador, Stanford University   05/2024 – Present
• Respond to inquiries from incoming M.S. students, explaining the curriculum and study environment.

Social Media Content Editor *and* Model, Chinese Rainbow Network (CRN)   03/2024 – Present
• Edited newsletter and promotional content on sexual and mental health resources for the Chinese
  LGBTQ+ community, with a focus on individuals exploring their identity, integrating fashion insights.
• Modeled to promote the CRN Pride Parade 2024 in San Francisco.

Scientific Review Editor (Education Branch), Community HealthEd   03/2022 – 05/2023
• Validated the accuracy of cited information in weekly articles by reviewing each referenced paper and
  website, with a primary focus on neurological, psychiatric, and prenatal health materials.
• Converted scientific jargon to plain language in weekly articles to ensure accessibility to the general
  public as a newsletter, retaining precise and accurate meanings while making it clear and concise.

TECHNICAL SUMMARY

Programming Languages: R, Python, Java, Swift
Markup/Styling Languages: HTML, CSS, LaTeX
Integrated Development Environments (IDEs): Jupyter Notebook, RStudio, Overleaf, SAS Studio
Tools: UNIX (on HPC cluster), Terminal, Secure Shell (SSH), Conda Environment, Git, GitHub

CERTIFICATIONS

SAS
• SAS Certified Specialist: Base Programming Using SAS 9.4   10/2023

Coursera Badges
• Introduction to Deep Learning & Neural Networks with Keras   08/2024