

ZHONGYI (JAMES) GUO

Email: guozy@stanford.edu

Phone: (+1) 607-262-3415

Location: Palo Alto, CA (flexible to relocate)

LinkedIn: [linkedin.com/in/zhongyi-guo/](https://www.linkedin.com/in/zhongyi-guo/)

GitHub: github.com/JG1ANDONLY

Website: jg1andonly.github.io

RESEARCH INTEREST

machine learning, deep learning, single-cell, computational genomics, multi-omics, bioinformatics

EDUCATION

MS Epidemiology and Clinical Research, Stanford University

(Expected) 06/2025

GPA: 3.94/4.00 (as of Spring 2024)

BS Biological Sciences and Biometry & Statistics (double major), Cornell University

05/2023

Concentrations: Computational Biology and Statistical Genetics

GPA: 3.57/4.30, Honors: *cum laude*, Dean's List

SELECTED RESEARCH EXPERIENCE

Graduate Student Researcher, the Kundaje Lab, Stanford University

04/2024 – Present

Mentored by Prof. Anshul Kundaje and Dr. Selin Jessa

Title: Deep Learning and Single-Cell Multiome Analysis of Psychiatric Disorder Regulatory Genomics

Description: This project investigates the pathogenesises of bipolar disorder (BPD) and schizophrenia (SCZ) using *PsychENCODE* scRNA-seq and scATAC-seq data.

- Pre-processed the single-cell multiome dataset using quality control, dimension reduction, annotation with reference, cell clustering, doublet removal, and marker gene detection with *Seurat* and *Signac*.
- Conducted differential gene expression (DGE) and peak accessibility (DA) analyses (pseudobulk: DESeq2; single-cell: Wilcoxon rank-sum test and MAST) across cell types by sex and disease.
- Visualized creatively and validated differential patterns for ~300 detected genes and peaks by cell type, aiding collaborators in integrating these genes into the Xenium spatial transcriptomics panel.
- Presented my result interpretations to diverse professors monthly, refining next-stage strategies.
- Training cell-type-specific *ChromBPNet* models to predict chromatin accessibility, assess mutation effects, and identify those causally linked to BPD and SCZ.

Graduate Student Researcher, the Graff Lab, UCSF

10/2023 – Present

Co-mentored by Prof. Rebecca Graff (at UCSF) and Prof. John Witte (at Stanford)

Title: Investigating Black-White Racial Disparities of Prostate Cancer via LC-MS Metabolomics

Description: This project examines metabolomic differences in Black-White prostate cancer disparities.

- Discovered key biological pathways using chemical similarity enrichment analysis with three methods: sub-pathway information, correlation modules, and predicted Medical Subject Headings classes.
- Found corroborative metabolites using PCA, partial least squares discriminant analysis, random forest, support vector machine, logistic regression, and pathway analysis using *MetaboAnalystR*.
- Identified aging-associated metabolites through differential abundance analysis with *limma*, following network construction and module detection via weighted gene co-expression network analysis.
- **Leading the analyses and manuscript preparation for submission in early January 2025.**

PUBLICATION

[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** and **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](#)

‡ indicates co-first authorship.

SELECTED CLASS PROJECT

Using Deep Learning to Study Sickle Cell Anemia Ischemic Stroke DGE 09/2024 – Present
Project Leader, *Course: Deep Learning, by Prof. Andrew Ng and Prof. Kian Katanforoosh at Stanford*
Description: This project predicts DGE patterns in sickle cell anemia patients with ischemic stroke using putative upstream promoter sequences, supporting early stroke prediction and precision medicine.

- Conducted DGE analysis on bulk RNA-seq data using DESeq2 as the ground truth.
- Trained a 2D convolutional neural network (CNN) using Tensorflow with mini-batch, max pooling, and ReLU activation on one-hot encoded promoter sequences, and fine tuned using regularization.
- Training a generative adversarial network (GAN) to generate promoter sequences, improving the 2D CNN performance and addressing label imbalance caused by fewer differentially expressed genes.

Global Disparities in Life Expectancy: Development Indicators and Energy 11/2024 – Present
Project Leader, *Course: Data Visualization, advised by Prof. Maneesh Agrawala at Stanford*
Description: Designing effective visualizations is crucial for conveying research interpretations. In this project, we created two visualizations to interactively illustrate how world development indicators and energy influence life expectancy across the globe.

- Created a global map illustrating life expectancy by country using a red-to-blue heat map, featuring tooltips displaying key development indicators on hover and a slider to adjust the heat map by year.
- Currently developing an interactive streamgraph transition plot for each development indicator.

Integrating Single-Cell Multi-omics Data to Investigate Alzheimer's Disease 03/2024 – 06/2024
Project Leader, *Course: Translational Bioinformatics, advised by Prof. Sylvia Plevritis at Stanford*
Description: This conceptual project focused on practicing NIH-style proposal writing with specific aims, without conducting actual analyses, but included a detailed statistical analysis plan and backup plan.

- Reviewed literature intensively on cell-type-specific multi-omics integration, and conducted preliminary analysis of multi-omics datasets with MOFA, targeting on microglial cells in human brains.
- Drafted the *Specific Aims* section of a conceptual NIH grant proposal and presented the statistical analysis plan to ~30 professors and students.

α -helix or not? 11/2021 – 12/2021

Final Independent Project, *Course: Biomedical Data Science, advised by Prof. Haiyuan Yu at Cornell*
Description: In this project, I built three classifiers to predict if a give protein is α -helix or not.

- Performed feature engineering on the training set by averaging 5 neighboring amino acids and removed redundant features measured by correlation coefficient.
- Trained binary classifiers (Logistic Regression, Decision Tree, and Random Forest) using Sklearn to predict α -helix or not using features derived from primary structures of proteins.

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.

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, Intro. 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 applicants, explaining the curriculum and research resources.

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, JavaScript (D3 and Vega-Lite)

Markup/Styling Languages: HTML, CSS, LaTeX

Integrated Development Environments (IDEs): Jupyter Notebook, RStudio, Overleaf, SAS Studio

Tools: UNIX (CLI on HPC clusters), Terminal, SSH, Conda, Git, GitHub, AWS EC2, CUDA

Packages: Tensorflow, Keras, pandas, numpy, Seurat, Signac, biomaRt, tidyverse, DESeq2, ...

CERTIFICATION

• SAS Certified Specialist: Base Programming Using SAS 9.4	10/2023
• Coursera: Introduction to Deep Learning & Neural Networks with Keras	08/2024
• Coursera: Deep Learning Specialization from DeepLearning.AI	11/2024