

Zhongyi (James) Guo

(+1) 607-262-3415 • guozy@stanford.edu • Palo Alto, CA 94304 •

Github: <https://github.com/JG1ANDONLY> • LinkedIn: <https://www.linkedin.com/in/zhongyi-guo>

Personal Website: <https://jg1andonly.github.io>

EDUCATION

M.S., Stanford University, Palo Alto, CA
Epidemiology and Clinical Research

Sep. 2023 – (Expected) Jun. 2025

B.S., Cornell University, Ithaca, NY

May 2023

Double Major: **Biological Sciences** (Computational Biology) and **Biometry & Statistics** (Statistical Genetics)

Honors: **CUM LAUDE (GPA: 3.57/4.30)**, Dean's List

PUBLICATION

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*)

SKILLS

Programming: R, SAS, Python, Java, Swift, SQL, HTML, CSS, JavaScript, UNIX/Linux, LaTeX

Developer Tools: Git, GitHub, Terminal, Jupyter Notebook, Quarto, RStudio, Eclipse, Overleaf, & Xcode

Core Skills: Genome-wide Association Study (GWAS), Statistical Testing & Modeling, Machine Learning, Data Science, Data Structure, iOS App Development, UNIX Commands & Shell Scripting, Website Design

RESEARCH EXPERIENCES

Undergraduate Research Assistant at Basu Lab, Ithaca, NY

Sep. 2022 – Dec. 2022

P.I.: Dr. Sumanta Basu, Department of Statistics and Data Science, Cornell University

- Applied Time Series Analysis in R to study time course and consequences of raised nocturnal blood pressure on cardiovascular disease among HIV-infected vs. uninfected adults in developing countries

Undergraduate Research Assistant at Lan Lab, Beijing, China

Apr. 2021 – Jul. 2021

P.I.: Dr. Xun Lan, School of Medicine, Tsinghua University; Supervisor: Dr. Lihui Wang

- Designed a High Throughput Screening (HTS) to study the interaction between high-affinity TCR and pMHC (major histocompatibility complex) in human and used k-means clustering to create prototypes
- Found the optimal set of mutation locations in yeast to inhibit phosphoprotein functions by performing CRISPR/Cas9 every week and then analyzed the returned sequences of purified plasmids extracted

PROJECT EXPERIENCES

R package: hurdatPro (unpublished class project)

May 2023

Co-authored with He Zhang, Advised by Dr. Guinness, Dept. of Statistics and Data Science, Cornell University

Github Repository: <https://github.com/JG1ANDONLY/hurdatPro.git>

- Created an R package in tar.gz format to analyze HURDAT Atlantic basin tropical cyclone activities
- Cleaned data, wrote functions for storm plotting (track, position, size) and detecting landfall in US, computed accumulated cyclone energy of storms, and wrote unit tests for each function using testthat

Causal Effect of Type II Diabetes on Prostate Cancer in East Asian Population

May 2022 – Dec. 2022

Advisor: Anqi Wang, Dept. of Preventive Medicine & Public Health Sciences, Univ. of Southern California

Github Repository: https://github.com/JG1ANDONLY/causal_effect_DM2_PrCa

- Performed two-sample Mendelian randomization with the inverse variance weighted method while using MR Egger and weighted median methods as sensitivity analysis on genetic-level data
- Identified proxy SNPs in linkage disequilibrium ($r^2 > 0.8$) and obtained OR = 0.76, 95% CI = [0.76, 0.89], P -value = 2.26×10^{-6} and similar results in sensitivity analysis
- Concluded that Type II diabetes has a negative causal effect on prostate cancer using genetic evidence

CVD Mortality Rate Analysis: Pre- & Post-COVID

Jun. 2022 – May 2023

Advisors: Dr. Sumanta Basu & Dr. Sreyoshi Das, Department of Statistics & Data Science, Cornell University

- Performed Time Series Analysis and detected CVD seasonality in cyclic pattern and regionally different mortality rate trend and corresponding three-month-average trend by state and by sex
- Used t-test, Wilcoxon test & permutation test for robustness, and hypothesis testing & post hoc tests with Tukey for each confounder respectively (age, sex, region, season)
- Led group teamwork and reported programming and result interpretation weekly

GWAS Study: Analysis of Lymphoblastoid Cell Lines (LCL) mRNA Levels

Apr. 2022 – May 2022

Class Project: BTRY 4830 Quantitative Genomics and Genetics

GitHub Repository: https://github.com/JG1ANDONLY/LCL_mRNA_GWAS

- Analyzed genotype & phenotype data and tested whether population and sex as two covariates could influence the GWAS result and cause differences in LCL mRNA level expressions by employing 2 different strategies for comparison: excluding both covariates & including both covariates
- Identified significant SNPs from Manhattan & QQ plots and phenotypes by 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

May 2022

Final Project: BTRY 4830 Quantitative Genomics and Genetics

GitHub Repository: https://github.com/JG1ANDONLY/citrulline_chronic_kidney_disease

- 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 Manhattan plot with 2 covariates included and interpreted the influence of linkage disequilibrium on the result

Weather Data Analysis in Ithaca, NY from 2021.01 to 2022.04

Mar. 2022 – Apr. 2022

Class Project: INFO 1998 Introduction to Machine Learning

GitHub Repository: https://github.com/JG1ANDONLY/Weather_Data_Analysis_in_Ithaca_NY

- Built a Logistic Regression model and a K-Nearest Neighbors (KNN) model to forecast snow in Ithaca, NY, based on daily temperature range using train-test split after data cleaning & EDA
- Reached the model accuracy at 0.809 for the Logistic Regression and 0.786 for the KNN with $k = 10$ and plotted confusion matrices for two models' tuning & validating and error analysis

α -helix or not?

Dec. 2021

Class Project: BTRY 4381 Biomedical Data Mining and Modeling

GitHub Repository: https://github.com/JG1ANDONLY/alpha_helix_or_not

- Trained a binary classifier to predict α -helix or not using features derived from the primary structure
- Performed feature engineering on the training set by averaging each feature of each observation's 4 neighboring amino acids and kept non-redundant distinctive features among 57 amino acid scales
- Built Logistic Regression, Decision Tree Regressor, and Random Forest models in Python
- Tuned maximum number of iterations using random search method to optimize Logistic Reg. model
- Conducted cross-validation and reached the model accuracy at 0.625, measured by AUROC

Salaries in Big Techs

Sep. 2021 – Dec. 2021

Class Project: INFO 2950 Introduction to Data Science

GitHub Repository: https://github.com/JG1ANDONLY/Salaries_in_Big_Techs

- Built a multiple linear regression model to predict total yearly salaries based on employee features, including years of experience, gender, race, education, etc. for tech companies in US and overseas
- Detected gender and education inequity among different races in the US technology companies by modeling the relationship between [gender, race] and education levels using logistic regression
- Established three equations to predict total yearly salary based on user inputs for users to optimize their incomes in the United States or overseas

BOOKED

Apr. 2022 – May 2022

Hack Challenge: CS 1998 Introduction to iOS Development

Demo Video: <https://www.youtube.com/watch?v=4wOO3DZTrUA>

GitHub Repository: <https://github.com/thuypham03/cu-libraries>

- Programmatically developed a Cornell library study room booking system iOS app by integrating UIKit, AutoLayout, Navigation, UITableView & UICollectionView, MVC, Delegation, and Animation
- Implemented GET all libraries and available rooms, POST new reservation(s), UPDATE reservation history & DELETE reservation(s) that interact with backend API using Alamofire
- Implemented UI with designers and collaborated with backend teammates for backend requests
- Nominated as Honorable Mention for **Best UI** for Hack Challenge Spring 2022

TEACHING EXPERIENCES

Cornell Bowers C•IS (College of Computing and Information Science)

- **Beta Tester & Teaching Assistant**, Introduction to Data Science Jan. 2023 – May 2023
 - Graded homework, held office hours, beta tested assignments, and communicated
- **Grader**, Probability Models and Inference Aug. 2022 – Dec. 2022

Teaching Assistant, Department of Molecular Biology and Genetics, Cornell University

- Laboratory in Genetics and Genomics (BIOMG 2801) Jan. 2021 – May 2021
 - Created and stabilized knockout mutations on target gene of fruit flies using CRISPR/Cas9
 - Analyzed mutations vs. wildtype on UCSC Genome Browser using bioinformatics skills, assisted with designing & cloning primers with sgRNA, and piloted students to locate sgRNA transgene

Teaching Assistant (Summer), JNC Study Abroad Platform

Jul. 2022 – Aug. 2022

- Introductory Biology (GNAEE07) & Fundamentals of Physics I (GNAEE06)
 - Contributed to exams, graded homework & exams, and hosted discussions & office hours

INDUSTRY EXPERIENCES

Match Group, Inc., Mobile (iOS) Development Intern (Remote)

Jun. 2022 – Aug. 2022

- Replaced singletons using dependency injections in Match & Stir codebases using POP
- Developed a new feature named “Enhanced Interests” using SwiftUI that users can choose tags comprised of text and emoji under many categories to better deliver their interests to other people
- Replaced part of UIKit code with SwiftUI in watchOS and abridged AB Test to improve performance

Tencent Ltd., Data Analyst Project Intern (Remote)

Jul. 2021 – Sep. 2021

- Web-scraped product details from e-commerce platforms & extracted data from databases using SQL
- Performed linear regression in Sklearn to detect customer preference patterns on different categories of products and practiced machine learning to simulate models to forecast the trend in item sales
- Visualized and presented my analysis results to marketing department for marketing strategies

EXTRACURRICULAR ACTIVITIES

Community HealthEd, Education Branch – Scientific Review Editor

Mar. 2022 – May 2023

- Focused primarily on maternal health, prenatal health, neurological & psychiatric health materials
- Visited each scientific paper/website cited in articles to validate the accuracy of the cited information
- Removed technical jargon from each article while retaining the meaning to make articles written clearly and concisely in plain language accessible to the general public as newsletters
- Cooperated efficiently with the authors, copy editors, and community outreach coordinators