

YINQI ZHAO

(929) 383-3149 | yinqiz@usc.edu | Los Angeles, CA

EDUCATION

- University of Southern California**, Los Angeles, CA Sept. 2018 - Dec. 2022 (Expected)
Ph.D. candidate in Biostatistics, GPA 3.93/4.00 (*Fellowship*, 2018)
- Columbia University**, New York, NY Sept. 2016 - May 2018
M.S. in Biostatistics (Theory and Method), GPA 4.00/4.00
- China Agricultural University**, Beijing, China Sept. 2012 - June 2016
B.S. in Biotechnology, GPA 3.56/4.00 (*Outstanding Undergraduate Student*, 2016)

RESEARCH EXPERIENCE & INTERNSHIP

- University of Southern California**, Los Angeles, CA Sept. 2019 - Present
Research Assistant
- **1. Clustering Analysis Integrating Omics data**
 - Developed a clustering model (*LUCID*) to uncover the latent clusters by integrating multi-view data and extended *LUCID* to incorporate missingness in omics data.
 - Proposed an integrated variable selection framework for *LUCID*, based on L_1 penalty.
 - Published and maintained an R package, *LUCIDus* on CRAN (current version is 2.1.5 with $\sim 17,000$ downloads).
 - **2. Applied Data Analysis**
 - Applied *LUCID* model to evaluate how prenatal exposure to PFAS associated with increased susceptibility to liver injury in children; created molecular signature for children at high risk of liver injury by using serum metabolomics data.
 - Conducted predictive modeling for Mediterranean diet adherence (*KIDMED*) and ultra-processed food consumption by using urinary metabolic profiling; evaluated model performance through Receiver-Operating Characteristic (ROC) analysis.
- Genentech**, San Francisco, CA (remotely) May 2021 - Aug. 2021
Summer Intern, Product Development & Data Science
- **1. Independent Read Data Quality Check**
 - Collaborated with biostatisticians, clinical scientists and software engineers across different functional groups and developed a pipeline to check quality of Independent Read data.
 - Developed a modularized R Shiny app to implement the proposed pipeline; the app was adopted by 2 clinical trial study teams.
 - **2. Automatic Clinical Trial Reporting**
 - Enhanced EnableRF R package to automatically generate PowerPoint slides for clinical trial report, featuring user-specified content format and automatic pagination for long tables.
- Mount Sinai**, New York, NY May 2017 - May 2018
Research Assistant
- Worked on trans-eQTL identification using data from GTEx; developed a pipeline to analyze gene-SNP pairs among 3 tissues, which decreased the total time of analysis for a single tissue (number of computations larger than 4×10^{10}) from more than 3 days to around 8 hours.

PROGRAMMING

R • R Shiny • \LaTeX • Python • TensorFlow • Keras • C++

SELECTED PRESENTATIONS

- Zhao, Y.**, et al., “*LUCID*, An Integrative Clustering Model For Multi Omics Data”, selected as platform presentation at *IGES* (2022)
- Zhao, Y.**, et al., “Integrative Clustering Analysis For Omics Data With Missingness”, *IGES* (2021)
- Zhao, Y.**, et al., “Latent Unknown Clustering With Integrated Omics Data (*LUCID*)”, *ExpSome Data Challenge* (2021)