

Yujia Li

Department of Biostatistics, 130 De Soto Street, Pittsburgh, PA 15261
4124828320 yul178@pitt.edu <https://yujiali1994.github.io/>

Summary of Qualifications

- 1: Proficient in **R** and familiar with SAS, Stata and UNIX.
- 2: Methodological research experience about various statistical learning topics for more than four years.
- 3: Strong communication skills with non-statistical researchers and strong writing skills.
- 4: Hands-on experience of analyzing datasets by various statistical and machine learning methods using R, such as MCMC, Sparse K-means, Random forest, Generalized linear models, and t-SNE.
- 5: Both theoretical and application knowledge of statistical analysis for various types of data, such as longitudinal data, high dimensional multi-omics data, and time-to-event data.

Education

University of Pittsburgh, Pittsburgh, PA

Expected 04-07/2022

Ph.D. Candidate in Biostatistics (GPA: 3.99/4.0)

- Advisor: Prof. George Tseng, ScD.
- Research interest: **Statistical Machine Learning, Bayesian Statistics, High-Dimensional Data, Analysis of multi-omics data.**
- **2019 ENAR distinguished student paper award.**

Xiamen University, Xiamen, China

Aug 2013 to June 2017

B.S. in Mathematics and Applied Mathematics (Major GPA: 3.97/4.0; Rank: 1/71)

Minor in Mathematical Finance

Work Experience

Statistician Intern (May 2020 to August 2020)

Abbvie Inc, Chicago, Illinois.

Supervisor: Yunxia Sui.

Worked on missing-data handling in confirmatory trials.

- Designed and implemented simulations to systematically evaluate the performance of different multiple imputation (MI) methods (MCMC, Monotone and FCS) based on SAS PROC MI, under MAR assumption.
- Evaluated the performance of tipping point analysis and pattern mixture model under MNAR assumption.

Graduate Student Researcher (Sep 2017 to present)

University of Pittsburgh Medical Center (UPMC), Pittsburgh, PA

Principal investigator: Dr. Steffi Oesterreich, Hung Jung Park and Liza Konnikova.

Provide statistical support for multiple investigators from different institutions.

- Analyze various high-throughput multi-omics datasets, including metabolomics and RNA-Seq data, starting from preprocessing raw data using command-line tools to downstream analysis such as differentially expressed (DE) analysis, pathway enrichment analysis, dimension reduction, clustering analysis, and network analysis.
- Involved in multiple breast cancer projects, such as meta-analytic outcome-guided clustering in breast cancer research, analyzing clinicopathological characterization and clinical outcomes using 20-year cancer registry data from UPMC, transcriptomic landscape with tumor immune infiltration, etc.
- Prepare peer-reviewed publications and posters.

Teaching Experience

Teaching Fellow/Instructor (Jan 2021 to May 2021)

BIOST 2094-Advanced R Computing,

Teaching Assistant (Aug 2019 to Oct 2019)

BIOST 2079-Introductory Statistical Learning for Health Science

Publication

^ indicate co-first authors.

Yujia Li, Xiangrui Zeng, Chien-Wei Lin, George C. Tseng (2021). “Simultaneous estimation of cluster number and feature sparsity in high-dimensional cluster analysis”.

Biometrics (In press, <https://doi.org/10.1111/biom.13449>). The earlier version won 2019 ENAR distinguished student paper awards.

Yujia Li[^], Tanbin Rahman[^], Tianzhou Ma, Lu Tang, George Tseng (2021). “A Sparse Negative Binomial Mixture Model for Clustering RNA-seq Count Data”, Biostatistics. Accepted (<https://doi.org/10.1093/biostatistics/kxab025>).

Yujia Li[^], Jessica Toothaker[^], Collin McCourt, Lael Werner, Scott Snapper, Dror Shouval, Omry Koran, Sameer Agnihorti, George Tseng, Liza Konnikova. (2020) “In Utero Human Intestine Harbors Unique Metabolome Including Bacterial Metabolites”. JCI Insight. (<https://insight.jci.org/articles/view/138751>)

Tanbin Rahman, Hsin-En Huang, **Yujia Li**, An-shun Tai, Wen-Ping Hsieh, George C. Tseng. (2021+) “A sparse negative binomial classifier with covariate adjustment for RNA-seq data”. Annals of Applied Statistics. Accepted.

Yujia Li and George C. Tseng (2021+). “Association study between gene expression and multiple phenotypes in omics applications of complex diseases”. Manuscript in preparation.

Azadeh Nasrazadani[^], Yujia Li[^], Yusi Fang, Jennifer M. Atkinson, Joanna S. Lee, Priscilla F. McAuliffe, George Tseng, Adrian V. Lee, Peter C. Lucas, Steffi Oesterreich, Norman Wolmark. (2021+) “Comprehensive Clinicopathologic Characterization of Mixed Invasive Ductal Lobular Carcinoma: A Meta-Analysis”. Manuscript in preparation.

Abstract Accepted by Conferences

Azadeh Nasrazadani, Yujia Li, et al. Abstract PS7-15: Mixed invasive ductal lobular carcinomas (mDLC) are clinically more similar to invasive lobular carcinoma (ILC) than to invasive ductal carcinoma (IDC)[J]. SABCS, 2021. DOI:10.1158/1538-7445.SABCS20-PS7-15

Azadeh Nasrazadani, Yujia Li, et al. “Mixed Invasive Ductal-Lobular Carcinoma: Clinicopathological Characterization and Clinical Outcomes.”, S. Journal of Clinical Oncology, 2020. DOI:10.1200/JCO.2020.38.15_suppl.1085

Azadeh Nasrazadani, Jennifer M Atkinson, Yujia Li, et al. Abstract P2-16-26: Mixed invasive ductal and lobular carcinoma (IDC/L) behaves similarly to invasive lobular carcinoma (ILC) with regard to neoadjuvant chemotherapy response and metastatic dissemination[J]. SABCS, 2020. DOI:10.1158/1538-7445.SABCS19-P2-16-26

Azadeh Nasrazadani, Yujia Li, et al. “Mixed Invasive Ductal-Lobular Carcinoma: Clinicopathological Characterization and Clinical Outcomes.”, S. Journal of Clinical Oncology 2019. DOI:10.1200/JCO.2019.37.15_suppl.e12531

Awards

ENAR Distinguished Student Paper Award	2019
Outstanding Graduate of Xiamen University	2017
Meritorious Winner in Mathematical Contest in Modeling	2016
The First Prize in China Undergraduate Mathematical Contest in Modeling.	2015
Academic Excellence Scholarship-First Prize	2015
Academic Excellence Scholarship-First Prize	2014
Third Prize in Jing Yun Mathematical Contest in Xiamen University	2014

Conference Presentation

“A Sparse Negative Binomial Mixture Model for Clustering RNA-seq Count Data”, speed presentation on JSM, August 2021.

“In Utero Human Intestine Harbors Unique Metabolome Including Bacterial Metabolites”. Oral Presentation at Dean’s Day Competition. April 2021.

“A Sparse Negative Binomial Mixture Model for Clustering RNA-seq Count Data”, Oral presentation on Biostatistics Research Day at University of Pittsburgh, March 2021

“Simultaneous Estimation of Number of Clusters and Feature Sparsity in Clustering High-Dimensional Data”, Oral presentation on ENAR, Mar 25th, 2019