GUAN'AO YAN

Contact: (+1) 213-469-7747 ◇ gayan@g.ucla.edu ◇ https://www.guanaoy.com
520 Portola Plaza, Room 8911, Los Angeles, CA 90095

RESEARCH AREA

My research interests lie in developing new statistical methods for understanding real-world data. Specific research topics include

- Statistical Bioinformatics
  - Statistical methods for analyzing high-dimensional single-cell and spatial omics data
  - Using synthetic data to enhance the statistical rigor in single-cell and spatial omics data analysis
- General Statistical Methodologies: High-dimensional model inference and variable selection
- Statistics in Education: Statistical methods for enforcing education equity (reported by Forbes)

EDUCATION

University of California, Los Angeles 09/2020 - Present
Ph.D. Candidate in Statistics
Advisor: Dr. Jingyi Jessica Li

Zhejiang University 09/2017 - 03/2020
M.Sc in Probability and Mathematical Statistics
Advisor: Dr. Yi Zhang

Shandong University 09/2013 - 06/2017
B.Sc in Mathematics and Applied Mathematics
B.Ec in Economics

AWARDS

JXTX+CSHL Genome Informatics 2023 Scholarship, Cold Spring Harbor Laboratory	2023
Don Ylvisaker Award for the Best Practice of Statistics, University of California, Los Angeles	2023
Interdisciplinary Opportunity Award, NSF-Simons Center for Multiscale Cell Fate Research	2022
Most Promising Statistician Award, University of California, Los Angeles	2022
Summer Mentored Research Fellowship, University of California, Los Angeles	2021
China National Scholarship	2018
Merit Graduate Student Award, Zhejiang University	2018
Outstanding Student Award, Shandong University	2014 - 2016
Hua Loo-Keng Talent Scholarship, Chinese Academy of Sciences	2015

PUBLICATIONS & MANUSCRIPTS

†*Indicating co-first author*

G. Yan, J.J. Li and M. Biggin (2024). Question-Score Identity Detection (Q-SID): A statistical algorithm to detect collusion groups with error quantification from exam question scores. *arXiv*, 2407.07420. [Website] [Forbes article] [Podcast] (Under review at *Journal of the American Statistical Association*)

G. Yan, S. Hua and J.J. Li (2024). Categorization of 31 computational methods to detect spatially variable genes from spatially resolved transcriptomics data. *arXiv*, 2405.18779. (Under revision at *Nature Communications*)

J. Zhao, F. Lao, **G. Yan** and Y. Zhang (2024). How data heterogeneity affects innovating knowledge and information in gene identification: A statistical learning perspective. *Journal of Innovation & Knowledge*, 9-3.

G. Yan, D. Song and J.J. Li (2023). scReadSim: a single-cell RNA-seq and ATAC-seq read simulator. *Nature Communications*, 14(1), 7428. [Software] [Website]

D. Song, Q. Wang, **G. Yan**, T. Liu and J.J. Li (2023). scDesign3 generates realistic in silico data for multimodal single-cell and spatial omics. *Nature Biotechnology*, 1-6. [Software]

Z. Li, Z. M. Patel, D. Song, **G. Yan**, J. J. Li and L. Pinello (2023). Benchmarking computational methods to identify spatially variable genes and peaks. *bioRxiv*, 2023-12. (Under review at *Nature Methods*)

S. Tang, H. Wang, **G. Yan**, L. Zhang (2022). Empirical likelihood based tests for detecting the presence of significant predictors in marginal quantile regression. *Metrika*, 1-31.

S. Chen†, **G. Yan**†, W. Zhang, J. Li, R. Jiang and Z. Lin (2021). RA3 is a reference-guided approach for epigenetic characterization of single cells. *Nature Communications*, 12(1), 1-13. [Software]

J. Zhao, **G. Yan** and Y. Zhang (2021). Robust estimation and shrinkage in ultrahigh dimensional expectile regression with heavy tails and variance heterogeneity. *Statistical Papers*, 1-28.

J. Zhao†, **G. Yan**† and Y. Zhang (2019). Semiparametric expectile regression for high-dimensional heavy-tailed and heterogeneous data. *arXiv*, 1908.06431.

PATENTS

M. Biggin, J.J. Li, **G. Yan**. Systems and methods for detecting collusion in student testing using graded scores or answers for individual questions (Serial No. 17/450,984; US Patent 11,915,615 B2)

SOFTWARE

Q-SID An online anti-collusion proctoring system, 2022. [Website]

scReadSim Python package of synthetic reads simulator designed for the single-cell multiomics data, 2022. [Software]

RA3 R package of "RA3 is a reference-guided approach for epigenetic characterization of single cells", 2021. [Software]

PRESENTATIONS

Oral Presentations

- Joint Statistical Meetings, Portland, USA 08/2024
- Institute for Computational and Experimental Research in Mathematics, Providence, USA 12/2023
- Jonsson Comprehensive Cancer Center Gene Regulation Seminar, Los Angeles, USA 11/2023
- Institute for Quantitative and Computational Biosciences Research Seminar, Los Angeles, USA 12/2022
- NSF-Simons Center for Multiscale Cell Fate 5th Annual Symposium, Irvine, USA 10/2022
- The 7th International Conference on Statistics and Probability, IMS-China, Dalian, China 07/2019

Poster Presentations

- Cold Spring Harbor Laboratory Genome Informatics Conference, New York, USA 12/2023
- RECOMB/ISCB Conference on Regulatory & Systems Genomics, Los Angeles, USA 11/2023
- Chan Zuckerberg Initiative Single-Cell Biology 2023 Annual Meeting, Carlsbad, USA 11/2023
- ISMB/ECCB, Lyon, France 07/2023
- LA Bioscience Ecosystem Summit, Los Angeles, USA 05/2023

· Jonsson Comprehensive Cancer Center Retreat Poster Session, Los Angeles, USA 05/2023
· Institute for Quantitative and Computational Biosciences Poster Session, Los Angeles, USA 09/2022

TEACHING & MENTORING

Teaching Assistant

· STATS 205, Hierarchical Linear Models, UCLA Spring 2024
· STATS 203, Large Sample Theory, UCLA Winter 2024
· MATH 1001, Advanced Mathematics, Zhejiang University Fall 2019
· MATH 1001, Advanced Mathematics, Zhejiang University Fall 2018

Workshop Instructor

· Coordinator & Presenter, QCBio Workshop, UCLA May 2022
"Statistical Methods for Enhancing the Rigor in Single-cell RNA-seq Data Analysis"

Guest Lecturer

· STATS 205, Hierarchical Linear Models, UCLA Spring 2024
· BIOINFO 229, Current Topics in Bioinformatics, UCLA Winter 2024

Undergraduate Student Mentor

· Weijian Wang, Zhejiang University 12/2022 - Present
· Zhiyin Liu, Hong Kong University of Science and Technology 12/2022 - Present
· Shuo Hua, Tsinghua University 06/2022 - 12/2022

PROFESSIONAL EXPERIENCE

Graduate Student Reseracher 09/2021 – Present
Department of Statistics & Data Science
University of California, Los Angeles
Advisor: Dr. Jingyi Jessica Li

Research Scientist Intern 06/2024 – 09/2024
Data and Statistical Sciences
Product Development Department
Genentech

Research Scientist Intern 06/2023 – 09/2023
Data and Statistical Sciences
Product Development Department
Genentech

Research Assistant 06/2019 – 04/2020
Department of Statistics
The Chinese University of Hong Kong
Advisor: Dr. Zhixiang Lin

Undergraduate Researcher 06/2016 – 06/2017
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
Advisor: Dr. Jianming Xia

PROFESSIONAL SERVICE

Reviewer for Scientific Journals:
Bioinformatics (5)

Co-reviewer for Scientific Journals:
Cell (2), Nature Biotechnology (1), Nature Methods (3), Nature Communications (3), Nature Machine Intelligence (1), Genome Biology (2), Bioinformatics (2), Journal of the American Statistical Association (1), Annals of Applied Statistics (1), NAR Genomics and Bioinformatics (5), Science Bulletin (1), Statistics in Medicine (1)

Co-reviewer for Scientific Conferences:
Research in Computational Molecular Biology (5), Intelligent Systems for Molecular Biology (5)

PROFESSIONAL AFFILIATIONS

Institute of Mathematical Statistics
American Statistical Association
American Society of Human Genetics
UCLA Jonsson Comprehensive Cancer Centers
International Indian Statistical Association