# 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 promoting 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

# **GRANTS & AWARDS**

Don Ylvisaker Award for the Best Practice of Statistics, University of California, Los Angeles 2023
Interdisciplinary Opportunity Award (\$10,000), NSF-Simons CMCF 2022
Most Promising Statistician Award, University of California, Los Angeles 2022
Summer Mentored Research Fellowship (\$6,000), 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 Sciences2015

# **PUBLICATIONS & MANUSCRIPTS**

#### $\dagger$ Indicating co-first author

**<u>G. Yan</u>**, 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. (Under review at *Journal of the American Statistical Association*) [Website] [Forbes article] [Podcast]

**<u>G. Yan</u>**, S. Hua and J.J. Li (2025). Categorization of 34 computational methods to detect spatially variable genes from spatially resolved transcriptomics data. *Nature Communications*, 16, 1141.

J. Zhao, F. Lao, <u>**G. Yan**</u> 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.

**<u>G. Yan</u>**, 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, <u>**G. Yan**</u>, 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, <u>**G. Yan**</u>, 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, <u>**G. Yan**</u>, L. Zhang (2022). Empirical likelihood based tests for detecting the presence of significant predictors in marginal quantile regression. *Metrika*, 1-31.

S. Chen<sup>†</sup>, <u>G. Yan<sup>†</sup></u>, 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, <u>G. Yan</u> 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<sup>†</sup>, **G. Yan<sup>†</sup>** and Y. Zhang (2019). Semiparametric expectile regression for high-dimensional heavy-tailed and heterogeneous data. arXiv, 1908.06431. (In press at Applied Mathematics-A Journal of Chinese Universities)

# PATENTS

M. Biggin, J.J. Li, <u>G. Yan</u>. 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)

# SOFTWARES

**Q-SID** An online anti-collusion proctoring system [Website]

**scReadSim** Python package of synthetic read simulator designed for the single-cell multiomics data [Software]

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

# PRESENTATIONS

#### **Oral Presentations**

•	NHGRI Genome Tech Dev Working Group, Jackson Laboratory, USA	1/2025
•	Department of Mathematical Sciences, New Jersey Institute of Technology, USA	12/2024
•	Department of Statistics and Data Sciences, University of California, Los Angeles, USA	11/2024
•	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
•	Los Angeles 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
· Statistical Science with Applications to Epidemiology, ElevatePro	Summer 2021
· MATH 1001, Advanced Mathematics, Zhejiang University	Fall 2019
$\cdot$ MATH 1001, Advanced Mathematics, Zhejiang University	Fall 2018
Workshop Instructor	
<ul> <li>Presenter, Jonsson Comprehensive Cancer Center Workshop, UCLA</li> <li>"Categorization of 34 Computational Methods to Detect Spatially Variable Generational Spatially Resolved Transcriptomics Data"</li> </ul>	Dec 2024 enes from
· Coordinator & Presenter, QCBio Workshop, UCLA "Statistical Methods for Enhancing the Rigor in Single-cell RNA-seq Data Ar	May 2022 malysis"
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	$\frac{12}{2022}$ - Present
$\cdot$ Shuo Hua, Tsinghua University	06/2022 - 12/2022
PROFESSIONAL EXPERIENCE	
<b>Graduate Student Reseracher</b> Department of Statistics & Data Science University of California, Los Angeles Advisor: Dr. Jingyi Jessica Li	09/2021 – Present
Research Scientist Intern Data and Statistical Sciences Product Development Department Genentech, Inc. <i>"Enhancing Statistical Practice in Recurrent Event Endpoint Data Analysis"</i>	06/2024 - 09/2024
<b>Research Scientist Intern</b> Data and Statistical Sciences Product Development Department Genentech, Inc. <i>"Dynamic Monitoring of Ongoing Clinical Trials for Early Decision Making"</i>	06/2023 - 09/2023

Department of Statistics The Chinese University of Hong Kong Advisor: Dr. Zhixiang Lin	
<b>Undergraduate Researcher</b> Academy of Mathematics and Systems Science Chinese Academy of Sciences Advisor: Dr. Jianming Xia	06/2016 - 06/2017
PROFESSIONAL SERVICES	
Conference and Session Organization	
<ul> <li>Session Organizer, STATGEN 2025 Conference</li> <li>"Spatially Variable Gene Detection: Advances and Applications in Spatial Transcriptomics data"</li> </ul>	May 2025
Statistical Consulting Services	
$\cdot$ Jonsson Comprehensive Cancer Center	2022 - 2024
ASSISTANCE WITH GRANT PROPOSALS	
Single-Cell Biology Data Insights Grant       07/01         Chan-Zuckerberg Initiative       07/01	1/2022 - 12/31/2023 \$200,000
"Enhancing Rigor and Reliability of Single-Cell Data Science" (PI: Dr. Jingyi Jes	ssica Li)

06/2019 - 04/2020

#### R35 GM140888

**Research Assistant** 

06/01/2021 - 05/31/2026NIH / NIGMS MIRA for established investigators \$1,848,665 "Statistical Methods for Elucidating Regulatory Mechanisms and Functional Impacts of Transcriptome Variation at Population and Single-cell Scales" (PI: Dr. Jingyi Jessica Li)

#### **REVIEWING ACTIVITIES**

# of papers in parentheses

#### **Reviewer for Scientific Journals**

Nature Communications (1), Bioinformatics (5)

# **Co-reviewer for Scientific Journals**

Cell (2), Nature Biotechnology (1), Nature Methods (4), Nature Communications (3), Nature Machine Intelligence (1), Genome Biology (2), Genome Research (1), Bioinformatics (2), Journal of the American Statistical Association (1), Annals of Applied Statistics (1), NAR Genomics and Bioinformatics (8), Science Bulletin (1), Statistics in Medicine (1)

#### **Co-reviewer for Scientific Conferences**

Research in Computational Molecular Biology (5), Intelligent Systems for Molecular Biology (5), National Council on Measurement in Education (8)

### **PROFESSIONAL AFFILIATIONS**

Institute of Mathematical Statistics	2023 - Present
American Statistical Association	2022 - Present
American Society of Human Genetics	2022 - 2023
UCLA Jonsson Comprehensive Cancer Centers	2022 - Present
International Indian Statistical Association	2022 - 2023