Martin Jinye Zhang

Phone: (650) 924-5691 | Email: martinzh@andrew.cmu.edu | Website: https://mzhanglab.github.io/

RESEARCH INTERESTS

I focus on integrative analysis of large-scale molecular data such as GWAS and single-cell data. The biological endeavor is on mechanistic interpretation of variants discovered in GWAS. The computational challenges involve analyses of high-dimensional, multimodal, and very-large-scale data. Specific topics include:

Genetics: GWAS, scRNA-seq, scATAC-seq, eQTL/pQTL, heritability estimation, disease-critical cellular contexts, Mendelian randomization, scRNA-seq experimental design.

Machine learning and statistics: multiple hypotheses testing, Monte Carlo methods, empirical Bayes, causal inference, algorithm acceleration, multi-armed bandits.

ACADEMIC APPOINTMENTS

Assistant Professor, Ray and Stephanie Lane Computational Biology Department, School of Computer Science, Carnegie Mellon University   8/2023 - present

Research Associate, Department of Epidemiology, Harvard University   7/2022 – 8/2023
Advisor: Alkes Price

Postdoctoral Researcher, Department of Epidemiology, Harvard University   9/2019 – 7/2022
Advisor: Alkes Price

EDUCATION

Ph.D. Electrical Engineering, Stanford University   9/2014 – 9/2019
Advisors: David Tse and James Zou

M.S. Electrical Engineering, Stanford University   9/2014 – 7/2017

B.Eng. Electrical Engineering, Tsinghua University   9/2010 – 7/2014

HONORS AND AWARDS

Major Awards
- Charles J. Epstein Trainee Awards Postdoctoral Semifinalist, ASHG   2021
- 2020 Top 50 Life and Biological Sciences Articles, Nature Communications, "Determining sequencing depth in a single-cell RNA-seq experiment"   2020
- Best Paper Award, RECOMB, "AdaFDR: a fast, powerful and covariate-adaptive approach to multiple hypothesis testing"   2019
- Stanford Graduate Fellowship (SGF), Stanford University   2015
- Numerical Technologies Award in Electrical Engineering, Stanford University   2015

- Outstanding Undergraduate Thesis, Tsinghua University, "Speech diarization based on the determinantal point processes" 2014

Additional Recognition
- Reviewers' Choice Award, ASHG, "Estimating SNP-pair effect correlations across functional annotations" 2023
- Reviewers' Choice Award, ASHG, "Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data" 2021
- Reviewers' Choice Award, ASHG, "Transcriptome-wide association studies and fine-mapping at cell-type resolution" 2021
- Travel Award, PQG (Program in Quantitative Genomics, Harvard School of Public Health) 2020
- Travel Award, RECOMB 2019
- Travel Award, NIPS 2017
- Ranked 2/79 in the Electrical Engineering Ph.D. qualifying exam, Stanford University 2015

RESEARCH GRANTS

1. The Shurl and Kay Curci Foundation Award 3/2024 – 3/2026
   "Understanding cell type-specific disease mechanisms using GWAS and single-cell data"
   (PI: Zhang; $500,000)

PUBLICATIONS

*equal contribution, †co-corresponding author, [name1, name2, …] alphabetical order
Full list at https://scholar.google.com/citations?user=zjr6n-QAAAAJ&hl=en

Selected main publications (first/senior author)

1. Mo Tiwari, Ryan Kang*, Donghyun Lee*, Sebastian Thrun, Chris Piech, Ilan Shomorony†, **Martin Jinye Zhang**†. "BanditPAM++: Faster k-medoids Clustering", *NeurIPS*, 2023.
2. **Martin Jinye Zhang***†, Kangcheng Hou*†, Kushal K. Dey, Karthik A. Jagadeesh, Kathryn Weinand, Saori Sakaue, Aris Taychameekiatchai, Poorvi Rao, Angela Oliveira Pisco, James Zou, Bruce Wang, Michael Gandal, Soumya Raychaudhuri, Bogdan Pasaniuc†, Alkes L. Price†. "Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data", *Nature Genetics*, 2022. Abstract selected as a postdoctoral semifinalist for the ASHG 2021 Charles J. Epstein Trainee Awards.
3. Mo Tiwari, Ryan Kang*, Je-Yong Lee*, Sebastian Thrun, Chris Piech, Ilan Shomorony†, **Martin Jinye Zhang**†. "MABSplit: Faster Forest Training Using Multi-Armed Bandits", *NeurIPS*, 2022.
4. **Martin Jinye Zhang**†, Angela Oliveira Pisco†, Spyros Darmanis, James Zou†. "Mouse aging cell atlas analysis reveals global and cell type-specific aging signatures", *eLife*, 2021.
5. **Martin Jinye Zhang***, Vasilis Ntranos*, and David Tse. "Determining sequencing depth in a single-cell RNA-seq experiment", *Nature Communications*, 2020. Top 50 Life and Biological Sciences Articles.
6. **Martin J. Zhang**, James Zou, and David Tse. "Adaptive Monte Carlo Multiple Testing via Multi-armed Bandits", *ICML*, 2019.
7. **Martin J. Zhang**, Fei Xia, and James Zou. "Fast and covariate-adaptive method amplifies detection power in large-scale multiple hypothesis testing", *Nature Communications*, 2019. Preliminary version won RECOMB 2019 Best Paper Award, out of 175 submissions.
8. [Abubakar Abid*, **Martin J. Zhang***], Vivek K. Bagaria, and James Zou, "Exploring patterns enriched in a dataset with contrastive principal component analysis", *Nature Communications*, 2018.
9. [Vivek Bagaria*, Govinda Kamath*, Vasilis Ntranos*, **Martin J. Zhang***], and David Tse, "Medoids in Almost Linear Time via Multi-armed Bandits", *AISTATS*, 2018.
10. [Fei Xia*, **Martin J. Zhang***], James Zou†, and David Tse†, "NeuralFDR: Learning Discovery Thresholds from Hypothesis Features", *NIPS*, 2017.

11. **Martin J. Zhang**, and Zhijian Ou, "Block-wise MAP Inference for the Determinantal Point Processes with Application to Change Point Detection", *SSP*, 2016.

12. **Jinye Zhang**, Laming Chen, Petros T. Boufounos, and Yuantao Gu, "On the Theoretical Analysis of Cross Validation in Compressive Sensing", *ICASSP*, 2014.

Selected other publications

1. Xilin Jiang, **Martin Jinye Zhang**[*], Yidong Zhang[*], Micheal Inouye, Chris Holmes, Alkes L. Price[†], Gil McVean[†]. "Age-dependent topic modeling of comorbidities in UK Biobank identifies disease subtypes with differential genetic risk". *Nature Genetics*, 2022. Role: co-mentor.

2. Li Gao, Ryutaro Kuraji, **Martin Jinye Zhang**, April Martinez, Allan Radaic, Pachiyappan Kamarajan, Charles Le, Ling Zhan, Changchang Ye, Hélène Rangé, M Reza Sailani, Yvonne L Kapila. "Nisin probiotic prevents inflammatory bone loss while promoting reparative proliferation and a healthy microbiome", *npj Biofilms and Microbiomes*, 2022.

3. Xihao Li, Godwin Yung, Hufeng Zhou, Ryan Sun, Zilin Li, Kangcheng Hou, **Martin Jinye Zhang**, Yaowu Liu, Theodore Arapoglou, Chen Wang, Iuliana Ionita-Laza, Xihong Lin. "A multi-dimensional integrative scoring framework for predicting functional variants in the human genome", *The American Journal of Human Genetics*, 2022.

4. Antonio Ginart, **Martin Jinye Zhang**, James Zou. "MLDemon: Deployment Monitoring for Machine Learning Systems", *AISTATS*, 2021. Role: co-mentor.

5. **The Tabula Muris Consortium**. "A single-cell transcriptomic atlas characterizes ageing tissues in the mouse", *Nature*, 2020.

6. Mo Tiwari, **Martin Jinye Zhang**, James Mayclin, Sebastian Thrun, Chris Piech, Ilan Shomorony. "Bandit-PAM: Almost Linear Time k-Medoids Clustering via Multi-Armed Bandits", *NeurIPS*, 2020. Role: co-mentor.

7. Li Gao, Misun Kang, **Martin Jinye Zhang**, M. Reza Sailani, Ryutaro Kuraji, April Martinez, Changchang Ye, Pachiyappan Kamarajan, Charles Le, Ling Zhan, Hélène Rangé, Sunita P. Ho, Yvonne L. Kapila. "Polymicrobial periodontal disease triggers a wide radius of effect and unique virome", *npj Biofilms and Microbiomes*, 2020.

8. Wenyu Zhou[*], M. Reza Sailani[*], Kévin Contrepois[*], Yanjiao Zhou[*], Sara Ahadi[*], Shana Leopold, **Martin J. Zhang**, Varsha Rao, Monika Avina, Tejaswini Mishra, Jethro Johnson, Brittany Lee-McMullen, Songjie Chen, Ahmed A Metwally, Thi Dong Binh Tran, Hoan Nguyen, Xin Zhou, Brandon Albright, Bo-Young Hong, Lauren Petersen, Eddy Bautista, Blake Hanson, Lei Chen, Daniel Spakowicz, Amir Bahmani, Denis Salins, Benjamin Leopold, Melanie Ashland, Orit Dagan-Rosenfeld, Shannon Rego, Patricia Limcaoco, Elizabeth Colbert, Candice Allister, Dalia Perelman, Colleen Craig, Eric Wei, Hassan Chaib, Daniel Hornburg, Jessilyn Dunn, Liang Liang, Sophia Miryam Schüssler-Fiorenza Rose, Kim Kukurba, Brian Piening, Hannes Rost, David Tse, Tracey McLaughlin, Erica Sodergren, George M. Weinstock[†], Michael Snyder[†], "Longitudinal multi-omics of host-microbe dynamics in prediabetes", *Nature*, 2019.

Preprints / papers in preparation

1. **Martin Jinye Zhang**[†], Arun Durvasula[*†], Colby Chiang[*†], Evan M. Koch, Benjamin J. Strober, Huwenbo Shi, Alison R. Barton, Samuel S. Kim, Omer Weissbrod, Po-Ru Loh, Steven Gazal, Shamil Sunyaev[†], Alkes L. Price[†]. "Pervasive correlations between causal disease effects of proximal SNPs vary with functional annotations and implicate stabilizing selection", *medRxiv*, 2023, *in revision in Nature Genetics*.

2. Benjamin J Strober, **Martin Jinye Zhang**, Tiffany Amariuta, Jordan Rossen, Alkes L. Price. "Fine-mapping causal tissues and genes at disease-associated loci", *medRxiv*, 2023, *in revision in Nature Genetics*.

3. Mo Tiwari, Ryan Kang[*], Je-Yong Lee[*], Luke Lee, Chris Piech, Sebastian Thrun, Ilan Shomorony[†], **Martin Jinye Zhang**[†]. "Faster Maximum Inner Product Search in High Dimensions". *arXiv*, 2022.

4. Huwenbo Shi[*], **Martin Jinye Zhang**[*], Alkes L. Price. "Cell-type transcriptome-wide association studies and fine-mapping via deconvolution using single-cell RNA-seq". *ASHG* 2022 platform talk.

RESEARCH SUPERVISION AND MENTORING

Ph.D. Student
- Ana Prieto (CMU Biological Sciences)   Fall 2023 - Present
- Alistair Turcan (CMU CPCB)   Fall 2023 - Present

M.S. Student
- Yuxin Lu (CMU MSAS)   Spring 2024 - Present
- David Sun (CMU)   Fall 2023 - Present
- Johnson Huang (CMU)   Fall 2023 - Present
- Grace Du (CMU MSCB)   Fall 2023 - Present

Undergraduate Student
- Maxwell Soh (CMU)   Spring 2024 - Present

INVITED TALKS

- Keynote Presentation (one hour), UCSD Causality Workshop, UCSD   4/4/2024
- Technical session, MCBIOS, Emory   3/23/2024
- "Estimating SNP-pair effect correlations across functional annotations", CPCB seminar, CMU   9/8/2023
- "Computational Methods to Decipher Human Disease Using GWAS and Functional Genomics", Department of Computer Science, UCI   3/8/2023
- "Computational Methods to Decipher Human Disease Using GWAS and Functional Genomics", School of Computer Science, CMU   2/16/2023
- "Computational Methods to Decipher Human Disease Using GWAS and Functional Genomics", Department of Genetics Symposium I, Yale   2/13/2023
- "Computational Methods to Decipher Human Disease Using GWAS and Functional Genomics", Special Seminar, UIUC   2/2/2023
- "Computational Methods to Decipher Human Disease Using GWAS and Functional Genomics", Special Biostatistics Seminars, UNC   1/19/2023

TEACHING

Carnegie Mellon University
- Co-instructor, 02-620, Machine Learning for Scientists   Spring 2024

Stanford University
- TA, EE 278, Introduction to Statistical Signal Processing   Spring 2017

PROFESSIONAL SERVICE

1. Frontiers in Genetics topic editor for the research topic "molecular and genetic heterogeneity of human diseases", 2022.
2. Journal paper reviewer for *Nature Genetics (1 paper, 2022), Cell Genomics (1 paper, 2023), Nature Communications* (6 papers, 2019-23), *Bioinformatics* (3 papers, 2019-22), *Journal of Advanced Research* (1 paper, 2022), *IEEE Transactions on Signal Processing* (1 paper, 2022), *Journal of Machine Learning Research* (1 paper, 2021), *Annals of Applied Statistics* (1 paper, 2021), *BMC Biology* (1 paper, 2020), *Biometrics* (1 paper, 2020), *Scientific Reports* (1 paper, 2020), and *Journal of Genetics and Genomics* (1 paper, 2018).

<div align="right">Martin Jinye Zhang</div>

3. Conference PC member for *GLBIO (2024) ISMB (2024), RECOMB (2024), ISMB/ECCB (2023) AAAI Student Program (2024).* Conference paper reviewer for *ICML* (2020-22), *NeurIPS* (2016, 2019-22), *IJCAI* (2021-22), *RECOMB* (2023-24), *ISMB* (2022-24), *ICLR* (2021).
4. Organizer of Information Systems Laboratory Colloquium, EE, Stanford, 2015-19.

Martin Jinye Zhang