MIN-ZHI JIANG

https://github.com/zjgbz mjiang26@jhu.edu

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

EDUCATION

University of North Carolina at Chapel Hill, Chapel Hill, NC — Aug 2016 - Dec 2022
Ph.D. in Materials Science

New York University, New York, NY — Sep 2013 - Aug 2016
M.S. in Physics

Nanjing University - Kuang Yaming Honors School, Nanjing, Jiangsu, China — Sep 2008 - Jun 2012
B.S. in Physics

WORKING EXPERIENCE

Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics — Jun 2023 - Present
Postdoctoral Fellow

University of North Carolina at Chapel Hill, Department of Genetics — Dec 2022 - May 2023
Postdoctoral Fellow

RESEARCH EXPERIENCE

Developing Pipeline for Analyzing Next Generation Cut & Tag Technology — Jun 2023 - Present
- Developing peak calling method for next generation Cut & Tag technology which is able to measure multiple targets simultaneously.
- Analyzing the signals of epigenomic biomarker pairs in different genomic bins.
- Exploring the efficacy of single-cell next generation Cut & Tag data for cell-type identification.

Proteomic Associations with Chronic Kidney Disease (CKD) — Sep 2021 - Present
- Evaluating the association between estimated glomerular filtration rate (eGFR), CKD and protein abundance.
- Detecting modules of proteins utilizing WGCNA (Langfelder, et al., 2008) and further exploring the pathways enriched in the modules significantly associated with eGFR and CKD.

Effects of α-globin Copy Number Variation (CNV) on Complex Traits — Nov 2018 - Dec 2020, Oct 2022 - May 2023
- Exploring if α-globin copy number variation modifies the effects of the β-globin variant rs334 (sickle cell disease defining variant) or an independent single nucleotide variant in the MCS-R2 enhancer region on complex traits associated with sickle cell disease or trait such as stroke, kidney disease, and blood cell traits in Black and Hispanic/Latino participants utilizing generalized linear mixed models.
- Called α-globin copy number variation utilizing Genome STRiP.
- Developed whole genome sequencing based CNV imputation pipeline utilizing Minimac4 and BEAGLE.

Cohort-Agnostic Genotype Imputation Quality Calibration — Sep 2022 - May 2023
- Developing semi-supervised learning based cohort-agnostic method for calibrating genotype imputation quality.
- Exploring the potential of multi-task learning frameworks to enhance performance of our calibration model.

Whole Genome Sequence Analysis of Inflammation Traits — Mar 2020 - Apr 2024
- Found *MMP9* locus significantly associated with Matrix Metallopeptidase 9 (MMP9) trait.
- Revealed 22 distinct and putatively novel signals across 8 previously identified loci for 6 inflammation traits.

Multi-Omics Integration Analysis — Sep 2018 - Nov 2022
- Improved the Sparse Multiple Canonical Correlation Analysis (Witten, et al., 2009) method by enhancing the orthogonality of the canonical variables.
- Developed the Supervised Sparse Multiple Canonical Correlation Analysis method.
- Found pronounced effect of blood cell counts on protein abundance in two independent cohort studies.
- Evaluated the advantages of our method over principal component analysis on explaining the variation of outcomes when assay-specific batch effects are present.

PUBLICATIONS

Working Papers

2. Mapping Combinatorial Epigenetic Events Using Hi-Plex CUT&Tag
   Y. Liao*, W. Zhou*, **M. Jiang***, K. Yu*, Z. Wang, T. Drummond, M. Ding, Y. Zhao, I. Pino, S. Taverna#, H. Ji#, H. Zhu#, ready for submission to *Cell*

1. Proteomics associations with CKD in aging women: the Women's Health Initiative
   M. Jiang, K. Reynolds, B. M. Lin, A. P. Reiner, C. Kooperberg, M. J Lamonte, S. W. Smoller, H. Kramer, N. Franceschini, ready for submission to *The American Journal of Human Genetics*

Refereed Journal Articles

13. scMBERT: A Pre-Trained Deep Learning Model for Single-Cell Multiomic Data Representation and Prediction
X. Chen, K. Yu, **M. Jiang**, C. Xiao, Z. Fu, W. Zhou *Proceedings of the AAAI Conference on Artificial Intelligence*: Vol. 39. No. 28. 2025.

12. A statistical framework for multi-trait rare variant analysis in large-scale whole-genome sequencing studies
X. Li, H. Chen, M. S. Selvaraj, [and 80 others, including **M. Jiang**], *Nature Computational Science* (2025): 1-19.

11. Whole genome sequencing based analysis of inflammation biomarkers in the Trans-Omics for Precision Medicine (TOPMed) consortium
M. Jiang*, S. M. Gaynor*, X. Li, et al., *Human Molecular Genetics* 33.16 (2024): 1429-1441.

10. MagicalRsq-X: A cross-cohort transferable genotype imputation quality metric
Q. Sun, Y. Yang, J. Rosen, [and 24 others, including **M. Jiang**], *The American Journal of Human Genetics* 111.5 (2024): 990-995.

9. Genome-wide study investigating effector genes and polygenic prediction for kidney function in persons with ancestry from Africa and the Americas
O. Hughes, A. Bentley, C. Breeze, [and 71 others, including **M. Jiang**], *Cell genomics* 4.1 (2024)

8. Functional characterization of Alzheimer's disease genetic variants in microglia
X. Yang, J. Wen, H. Yang, [and 25 others, including **M. Jiang**], *Nature Genetics* 55.10 (2023), 1735-1744

7. Canonical correlation analysis for multi-omics: Application to cross-cohort analysis
M. Jiang, F. Aguet, K. Ardlie, J. Chen, et al., *PLOS Genetics* 19, no. 5 (2023): e1010517.

6. Cell composition inference and identification of layer-specific spatial transcriptional profiles with POLARIS
J. Chen, T. Luo, **M. Jiang**, et al., *Science Advances*, 9(9), eadd9818.

5. Whole genome sequencing identifies common and rare structural variants contributing to hematologic traits in the NHLBI TOPMed program
M. M. Wheeler, A. M. Stilp, S. Rao, [and 52 others, including **M. Jiang**], *Nature Communications* 13, no. 1 (2022): 1-18.

4. MagicalRsq: Machine-learning-based genotype imputation quality calibration
Q. Sun, Y. Yang, J. D. Rosen, **M. Jiang**, et al., *The American Journal of Human Genetics* 109.11 (2022): 1986-1997.

3. A comprehensive comparison on cell-type composition inference for spatial transcriptomics data
J. Chen, W. Liu, T. Luo, Z. Yu, **M. Jiang**, et al., *Briefings in Bioinformatics* 23, no. 4(2022): bbac245.

2. A systematic evaluation of Hi-C data enhancement methods for enhancing PLAC-seq and HiChIP data
L. Huang, Y. Yang, G. Li, **M. Jiang**, et al., *Briefings in Bioinformatics* 23, no. 3(2022): bbac145.

1. Adjusting prediction of ozone concentration based on CMAQ model and machine learning methods in Sichuan-Chongqing region, China
H. Lu, M. Xie, X. Liu, B. Liu, **M. Jiang**, et al., *Atmospheric Pollution Research*, 12(6), 101066.

CONFERENCE PRESENTATIONS

5. Proteomics study of kidney traits in aging women: the Women's Health Initiative
M. Jiang, K. Reynolds, B. M. Lin, A. P. Reiner, C. Kooperberg, M. J Lamonte, S. W. Smoller, H. Kramer, N. Franceschini
TOPMed 2023 Annual Meeting, Lightning Talks

4. Whole Genome Sequencing (WGS) based Association Study of 21 Inflammation Biomarkers in up to 38,473 Multi-Ethnic Individuals Identifies Novel Signals
M. Jiang, S. M. Gaynor, Y. Li, L. M. Raffield, P. L. Auer, TOPMed Inflammation Working Group, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium
ASHG 2021, Session 034 - Genetic variation in the context of immune traits, **Oral Presentation**

3. Multi-Omics Data Integration with Sparse Multiple Canonical Correlation Analysis in the Multi-Ethnic Study of Atherosclerosis (MESA) Study
M. Jiang, L. M. Raffield, [and 19 others], Y. Li, TOPMed MESA Multi-omics Working Group
ASHG 2020, Session 002 - Computational Methods for Association Studies, **Oral Presentation**

2. Leveraging deep learning methods developed for Hi-C data to enhance resolution of HiChIP/PLAC-seq data
L. Huang, **M. Jiang**, G. Li, A. Abnousi, J. D. Rosen, Y. Yang, M. Hu, Y. Li
ASHG 2020, Session 201 - Bioinformatics and Computational Approaches, Poster Presentation

1. Calling and Imputation of the Common α-globin Copy Number Variant with Whole Genome Sequencing Data in TOPMed and Association with Hematologic and Other Clinical Phenotypes
M. Jiang, Y. Su, T.W. Blackwell, A. Correa, G. Abeçasis, A.P. Reiner, Y. Li, L.M. Raffield
ASHG 2019, Session - Cardiovascular Phenotypes, Poster Presentation

OTHER INVITED PRESENTATIONS

4. Whole Genome Sequencing (WGS) based Association Study of 21 Inflammation Biomarkers in up to 38,473 Multi-Ethnic Individuals Identifies Novel Signals
   TOPMed Inflammation Working Group Community Showcase, Nov 2021
   NHLBI BioData Catalyst Platform Community Showcase, Nov 2021
3. Multi-Omics Data Integration with Sparse Multiple Canonical Correlation Analysis in the Multi-Ethnic Study of Atherosclerosis (MESA) Study
   Invited Talk: Statistical Genetics Seminar Series, Boston University, Dec 2020
   Invited Talk: Framingham Heart Study - OMICS Conference Series (FOCuS) Presentation, Dec 2020
2. Exploratory Analysis of MESA Multi-omics Data via Sparse Multi-CCA (Canonical Correlation Analysis)
   TOPMed Multi-Omics Working Group Community Showcase, Jul 2020
1. Sparse Multi-CCA in Action: Preliminary Results From MESA
   Invited Talk: Human Functional Regulatory Genomics Meeting, Feb 2020

TEACHING EXPERIENCE

Guest Lecture of PH.140.860 Current Topics in Biostatistics Research, JHU Feb 2025

- Topic:*"Advances in Epigenomic Technologies and Challenges in Data Analysis"*

Guest Lecture of BCB 723.001 TOPICS IN STATISTICAL GENETICS, UNC-CH Feb 2022

- Topic:*"Canonical Correlation Analysis, Derivatives, and its Application: A Glimpse of Multi-Omics Integration Analysis"*

Graduate Teaching Assistant, UNC-CH Jun 2016 - Dec 2016

- PHYS 114 General Physics for non-physics major, led workshop as *Teaching Assistant*

Graduate Teaching Assistant, NYU Sep 2014 - Dec 2014

- PHYS-UA 120 Undergraduate Dynamics for physics major, led recitation session as *Instructor*

TECHNICAL SKILLS

Bioninformatics Tools: NHLBI BioData Catalyst Platform (with API control), Genome STRiP, Minimac4, BEAGLE
AI Agent: openai-agents-python, LangChain
Programming Languages: Python, R, MatLab, C, Bash/Zsh Scripting
Frameworks and Libraries: PyTorch, Numpy, Matplotlib, Scipy, Cython, scikit-learn, Pandas, Seurat, Scanpy
Supporting Skills: Git, LaTeX, Jupyter Lab, Microsoft Office