

# BORAN GAO

✉ borang@umich.edu 🌐 <https://borangao.github.io/>

📍 4331 SPH II, 1415 Washington Heights, Ann Arbor, MI 📞 (734)-546-2396

## EDUCATION

---

### Ph.D. in Biostatistics

University of Michigan, Ann Arbor, MI

Advisor: Dr. Xiang Zhou

Dissertation Title: Statistical Methods and Computational Tools for Genetics and Genomics Data

Jun 2024

GPA: 4.0/4.0

### M.S. in Biostatistics, MPH in Epidemiology

University of Michigan, Ann Arbor, MI

Sep 2014 - Apr 2017

GPA: 4.0/4.0

### Bachelor of Medicine, Bachelor of Surgery

Soochow University, Suzhou, China

Sep 2006 - Jul 2011

GPA: 3.2/4.0

## ACADEMIC EXPERIENCE

---

**Assistant Professor**, Department of Statistics, Purdue University

Aug 2024 - Present

**Assistant Professor**, Department of Biology, Purdue University

Aug 2024 - Present

## RESEARCH INTEREST

---

**Methods:** Bayesian variable selection, mixed effects models, data integration, high-dimensional data analysis, and machine learning

**Applications:** Heritability and genetic correlation estimation, fine-mapping, colocalization, polygenic risk score prediction, and multi-omics integration

**Data types:** Genome-wide association study (GWAS), bulk RNA-seq, spatially resolved transcriptomics (SRT), perturb-seq, proteomic, and electronic health records (EHR) data

## PUBLICATIONS

---

### Peer-reviewed Publications

- Boran Gao** and Xiang Zhou (2024). MESuSiE enables scalable and powerful multi-ancestry fine-mapping of causal variants in genome-wide association studies. *Nature Genetics*
- Boran Gao**, Can Yang, Jin Liu, and Xiang Zhou (2021). Accurate genetic and environmental covariance estimation with composite likelihood in genome-wide association studies. *PLOS Genetics*, 17(1), e1009293.
- Lu Wang, **Boran Gao**, Yue Fan, Fuzhong Xue, and Xiang Zhou (2021). Mendelian randomization under the omnigenic architecture. *Briefings in Bioinformatics*, 22(6), bbab322.

### Manuscript under Review

- Zheng Li, Boran Gao, Xiang Zhou. VINTAGE: A unified framework integrating gene expression mapping studies with genome-wide association studies for detecting and deciphering gene-trait associations. Under revision in *Nature Genetics*.
- Daiwei Zhang, **Boran Gao**, Qidi Feng, Ani Manichaikul, Gina M. Peloso, Russell P. Tracy, Peter Durda, Kent D. Taylor, Yongmei Liu, W. Craig Johnson, Stacey Gabriel, Namrata Gupta, Joshua D. Smith, Francois Aguet, Kristin G. Ardlie, Thomas W. Blackwell, Robert E. Gerszten, Stephen S. Rich, Jerome I. Rotter, Laura J. Scott\*\*, Xiang Zhou\*\*, and Seunggeun Lee\*\* (2023+). Proteome-wide association studies for blood lipids and comparison with transcriptome-wide association studies. Under revision in *HGG Advances*. \*\* co-corresponding authorship.

## Manuscripts in Preparation

1. Qidi Feng\*, **Boran Gao\***, Daiwei Zhang, Laura J. Scott\*\*, Seunggeun Lee\*\*, and Xiang Zhou\*\*. Transcriptome-wide association studies in lipid traits in diverse populations. \* Co-first authorship, \*\* co-corresponding authorship.
2. **Boran Gao**, Zheng Li, and Xiang Zhou. Logica: A powerful approach for estimating local genetic correlations across ancestries via GWAS summary statistics.
3. **Boran Gao**, Peijun Wu, Lulu Shang, and Xiang Zhou. MECO enables powerful multi-ancestry colocalization analysis.
4. Robert Langefeld, **Boran Gao**, and Xiang Zhou. pyGEMMA: A user-friendly Python implementation of linear mixed model for genome-wide association studies.

## RESEARCH EXPERIENCE

---

### Graduate Student Research Assistant

*Jan 2018 - Present*

Department of Biostatistics, University of Michigan

Advisor: Dr. Xiang Zhou

### *Genetic and Environmental Correlation Estimation in GWAS*

*Jan 2018 - Oct 2020*

- Statistical Modeling
  - Developed a scalable method for genetic and environmental correlation estimation, GECKO
  - Utilized a bivariate linear mixed model to capture both genetic and environmental correlations
  - Derived and implemented a composite likelihood based optimization algorithm
- Application
  - Identified genetic and environmental correlated traits and enhanced understanding of trait co-evolution
  - Experience with five large scale GWASs and gene set enrichment analysis

### *Multi-Ancestry Fine-Mapping*

*Jan 2021 - Oct 2022*

- Statistical Modeling
  - Developed a scalable and powerful method for multi-ancestry fine-mapping, MESuSiE
  - Used a multivariate sum of single effects model for causal variant detection and employed a mixture of multivariate normal to model both shared and ancestry-specific effects
  - Derived and implemented the multivariate iterative bayesian stepwise selection algorithm
- Application
  - Identified shared and ancestry-specific causal SNPs, facilitated understanding of disease etiology across diverse ancestries, realized the benefits of GWAS for underrepresented populations
  - Experience with multi-ancestry large-scale GWASs and function enrichment analysis

### *Multi-ancestry Colocalization*

*Jun 2022 - Present*

- Statistical Modeling
  - Developed a scalable and powerful method for multi-ancestry colocalization, MECO
  - Employed multivariate normal mixture assumption to model colocalization signals, extended the MESuSiE algorithm towards multi-ancestry colocalization

- Application
  - Identified colocalization signals across ancestries, provided insights into shared biological pathways between molecular phenotype and complex traits across ancestries
  - Bridged GWAS with bulk RNA-seq studies

***Local Genetic Correlation Estimation across Ancestries***

*Nov 2022 - Present*

- Statistical Modeling
  - Developed a scalable and powerful method for local genetic correlation estimation, Logica
  - Utilized a bivariate linear mixed model for genetic correlation estimation and testing
  - Derived and implemented the parameter expansion expectation–maximization algorithm for scalable estimation
- Application
  - Estimated and detected local genetic correlations, uncovered shared genetic architecture across ancestries
  - Experience with multi-ancestry large-scale GWASs

**Graduate Student Research Assistant**

*Sep 2019 - Apr 2021*

Department of Biostatistics, University of Michigan

Advisor: Dr. Xiang Zhou, Dr. Seunggeun Lee, Dr. Laura Scott

***eQTL Mapping and TWAS in the Multi-Ethnic Study of Atherosclerosis***

- Conducted multiple ancestry expression quantitative trait locus (eQTL) mapping study, and compared genetic architecture underlying expression variation
- Utilized PrediXcan to train gene expression predictors and performed transcriptome-wide association study (TWAS) across multiple ancestries
- Developed highly efficient and scalable eQTL detection and TWAS analysis pipeline

**SELECTED PRESENTATIONS**

- 
- Department of Epidemiology, UT Health Houston *April 2024*  
MESuSiE enables scalable and powerful multi-ancestry fine-mapping in GWASs
  - PGC Cross-population Group, Broad Institute of MIT and Harvard *March 2024*  
MESuSiE enables scalable and powerful multi-ancestry fine-mapping in GWASs
  - Department of Statistics, Purdue University *March 2024*  
MESuSiE enables scalable and powerful multi-ancestry fine-mapping in GWASs
  - Department of Statistics, Texas A&M *Feb 2024*  
MESuSiE enables scalable and powerful multi-ancestry fine-mapping in GWASs
  - Department of Biostatistics, Ohio State University *Feb 2024*  
MESuSiE enables scalable and powerful multi-ancestry fine-mapping in GWASs
  - Department of Biostatistics, University of Alabama, Birmingham *Feb 2024*  
MESuSiE enables scalable and powerful multi-ancestry fine-mapping in GWASs
  - Division of Epidemiology, Vanderbilt University Medical Center *Jan 2024*  
MESuSiE enables scalable and powerful multi-ancestry fine-mapping in GWASs

- Department of Biostatistics, Huazhong University of Science and Technology *Jan 2024*  
MESuSiE enables scalable and powerful multi-ancestry fine-mapping in GWASs
- Platform talk, American Society of Human Genetics (ASHG), Los Angeles, CA *Oct 2022*  
MESuSiE: a novel method for discovering shared and unique putative causal variants by fine-mapping across diverse ancestries
- Center for Statistical Genetics, Department of Biostatistics, University of Michigan *2019-2021*  
MESuSiE: a novel method for discovering shared and unique putative causal variants by fine-mapping across diverse ancestries  
GECKO: Accurate genetic and environmental covariance estimation with composite likelihood in genome-wide association studies

## PROFESSIONAL EXPERIENCE

---

**Intern Statistical Genetics, 23andMe, CA** *Jun 2022 – Aug 2022*

- Developed an R package for multi-ancestry fine-mapping of 23andMe data
- Accounted for varying number of SNPs across ancestries, assumed causal SNPs are shared

**Medical Consultant, Gongren's Hospital, Tangshan, China** *Mar 2012 – Aug 2013*

- Assisted with various administrative tasks, arranged rotations for resident physicians
- Managed the resolution of medical disputes

## TEACHING AND MENTORING EXPERIENCE

---

**Graduate Student Instructor (GSI)**

- EPID 503: Strategies and Uses of Epidemiology *Jan 2016 - Apr 2016*  
Advisor: Dr. Aubree Gordon, Department of Epidemiology, University of Michigan.
- BIOSTAT 523: Statistical Methods for Epidemiology *Sep 2016 - Dec 2016*  
Advisor: Dr. Susan Murray, Department of Biostatistics, University of Michigan

**Research Mentor for:**

- Peijun Wu, Biostatistics Ph.D. student, University of Michigan *Jun 2023 - Present*
- Robert Langefeld, Biostatistics M.S. student, University of Michigan *Mar 2023 - Present*
- Zheng Li, Biostatistics Ph.D. student, University of Michigan *Nov 2022 - Present*

## SELECTED HONORS AND AWARDS

---

- **Rackham One-Term Dissertation Fellowship**, University of Michigan *2023*
- **Excellence in Research Award**, Department of Biostatistics, University of Michigan *2023*
- **Charles J. Epstein Trainee Award**, Predoctoral semifinalist, ASHG *2022*
- **Rackham Travel Award**, University of Michigan *2022*
- **Outstanding GSI**, Department of Biostatistics, University of Michigan *2017*

## LEADERSHIP AND SERVICES

---

**Diversity, Equity and Inclusion Committee** of Biostatistics *Sep 2019 - July 2021*

## SOFTWARE

---

- MESuSiE: Multiple ancEstry fine-mapping via Sum of the Single Effect model  
<https://github.com/borangao/MESuSiE>
- GECKO: Genetic and Environmental covariance estimation via Composite liKelihood Optimization  
<https://github.com/borangao/GECKO>

## **TECHNICAL SKILLS**

---

- Programming: R, Python, C++, Shell, Git, L<sup>A</sup>T<sub>E</sub>X, SAS
- Operating Systems: Mac OS X, Linux, Windows.

## **PROFESSIONAL SERVICE**

---

Journal Reviewer: Nature Genetics, American Journal of Human Genetics, Briefing in Bioinformatics, Journal of Orthopaedic Surgery and Research, Frontiers in Immunology, BMC Bioinformatics

## **ACADEMIC MEMBERSHIP**

---

American Statistical Association, American Society of Human Genetics

## REFERENCE LIST

---

1. **Xiang Zhou**

Professor, Department of Biostatistics, University of Michigan, Ann Arbor, MI

✉ xzhousph@umich.edu

☎ (734)-764-5722

Use interfolio: send.XiangZhou.499CABCFA2@interfoliodossier.com

2. **Seunggeun Lee**

Professor, Graduate School of Data Science, Seoul National University, Seoul, Republic of Korea

Adjunct Professor, Department of Biostatistics, University of Michigan, Ann Arbor, MI

✉ lee7801@snu.ac.kr

☎ (734)-764-7884

Use interfolio: send.SeunggeunLee.1BF3339B54@interfoliodossier.com

3. **Zhongshang Yuan**

Professor, Department of Biostatistics, Shandong University, Jinan, Shandong, P.R.China

✉ yuanzhongshang@sdu.edu.cn

☎ +86 150 6909 5790

Use interfolio: send.ZhongshangYuan.1E82559C2D@interfoliodossier.com