

Xiran Liu

Email: xiran_liu1@brown.edu | Phone: (+1)314-421-9681
GitHub: github.com/xr-cc | Website: xr-cc.github.io

EDUCATION AND TRAINING

Postdoc	Data Science Institute & Center for Computational Molecular Biology, Brown University	Oct 2023-
<i>Advisor: Dr. Sohini Ramachandran</i>		
Ph.D.	Computational and Mathematical Engineering, Stanford University	Sep 2023
<i>Dissertation: "Computational Methods and Mathematical Measures for Population Relationships"</i>		
<i>Advisor: Dr. Noah Rosenberg</i>		
<i>Thesis Committee: Dr. Marcus Feldman, Dr. Julia Palacios</i>		
M.S.	Computational and Mathematical Engineering, Stanford University	Sep 2020
B.S.	Systems Science and Engineering, Washington University, St. Louis, <i>Summa Cum Laude</i>	May 2018
B.S.	Computer Science, Washington University, St. Louis, <i>Summa Cum Laude</i>	May 2018

AWARDS AND FELLOWSHIPS

Rising Stars in Data Science Workshop 2025 Cohort, hosted by Stanford University in collaboration with the University of California, San Diego, and the University of Chicago; \$1,000 travel award (2025)

Gene Golub Dissertation Award, Institute for Computational and Mathematical Engineering, Stanford University; awarded to 1-3 recipients annually (2023)

CEHG Predoctoral Fellowship, Center for Computational, Evolutionary and Human Genomics, Stanford University; one-quarter fellowship including \$17,173 total support (2022–2023)

The Enlight Foundation Graduate Fellowship, Stanford University; departmental fellowship supporting first-year doctoral students (2018–2019)

Dean's List, Outstanding Sophomore, Junior, and Senior Awards, and Undergraduate Research Scholarship, Washington University (2014–2018)

RESEARCH INTERESTS

Data science; machine learning; computational biology; development of scalable computational frameworks to extract patterns, model complex relationships, and interpret biological variation; advancement of learning-based approaches for scientific discovery; data-driven cross-domain methodological innovation; data visualization.

PUBLICATIONS

* denotes corresponding author(s).

Peer-Reviewed Journal Articles

Liu X*, Crawford L, Ramachandran S (2025). ML-MAGES enables multivariate genetic association analyses with genes and effect size shrinkage. *Genome Res. (accepted)*

Liu X*, Ahsan Z, Rosenberg NA (2025). Using mathematical constraints to explain narrow ranges for allele-sharing dissimilarities. *Theor. Popul. Biol. (in press)*

Liu X*, Kopelman N, Rosenberg NA* (2024). Clumppling: cluster matching and permutation program with integer linear programming. *Bioinformatics* 40(1).

Liu X*, Ahsan Z, Martheswaran T, Rosenberg NA (2023). When is the allele-sharing dissimilarity between two populations exceeded by the allele-sharing dissimilarity of a population with itself?. *Stat. Appl. Genet. Mol. Biol.* 22(1).

Liu X, Kopelman N, Rosenberg NA* (2022). A Dirichlet model of alignment cost in mixed-membership unsupervised clustering. *J. Comput. Graph. Stat.* 1-35.

Liu X, Rosenberg NA*, Greenbaum G* (2022). Extracting hierarchical features of cultural variation using network-based clustering. *Evol. Hum. Sci.* 4.

Parikh VN, Ioannidis AG, Jimenez-Morales D, Gorzynski JE, De Jong HN, Liu X, Roque J, Cepeda-Espinoza VP, *et al.* (2022) Deconvoluting complex correlates of COVID-19 severity with local ancestry inference and viral phylodynamics: Results of a multi-omic pandemic tracking strategy. *Nat. Comm.* 13(1), 1-10.

Liu X, Feldman MW* (2021). Effects of cultural transmission of surnaming decisions on the sex ratio at birth. *Theor. Popul. Biol.* 141, 44-53.

Ye Z, Price RL, Liu X, *et al.* (2020). Diffusion histology imaging detects and classifies glioblastoma pathology missed by conventional magnetic resonance imaging. *Clin. Cancer Res.* 26(20):5388-5399.

Conference Paper

Liu X*, Crawford L, Ramachandran S (2025). ML-MAGES: A machine learning framework for multivariate genetic association analyses with genes and effect size shrinkage. In: Sankararaman, S. (eds), *Research in Computational Molecular Biology (RECOMB) 2025*. LNCS, Vol. 15647.

Works in Progress

Liu X, Singh R, Ramachandran S. Clustering alignment for single-cell clustering analysis facilitates model comparison and characterization of clustering-informative genes contributing to cellular heterogeneity. (*in prep*)

Ahsan Z, Liu X, Rosenberg NA. Combinatorics of a dissimilarity measure for pairs of draws from discrete probability vectors on finite sets of objects. (*under review; preprint on arXiv*)

Liu X, Greenbaum G, Rosenberg NA. Can migration be inferred from pairwise F_{ST} ? (*in prep*)

Application Notes

Guo A, Ramachandran S*, Liu X*. *KAlignedoscope*: an interactive visualization tool for aligned clustering results. (*under review*)

Liu X, Rosenberg NA, Ramachandran S. *Clumppling 2.0*: A Clustering Alignment Program for Population Structure Analyses. (*in prep*)

CONFERENCE PRESENTATIONS

Oral Presentations

“Clustering alignment for single cell analyses: streamlining model comparison and revealing informative genes.” *Genome Informatics 2025*. Cold Spring Harbor Laboratory, NY, Nov 2025

“ML-MAGES: A machine learning framework for multivariate genetic association analyses with genes and effect size shrinkage.” *The Research in Computational Molecular Biology (RECOMB) 2025 Conference*. Seoul, Korea, Apr 2025.

“Clumppling: a new method for aligning replicate solutions in population structure analysis.” *Evolution 2023 Conference*. Albuquerque, NM, Jun 2023.

“A Dirichlet model of alignment cost in mixed-membership clustering results of ancestry inference.” *Bay Area Population Genomics Conference*. UC Berkeley, Berkeley, CA, Oct 2022.

Poster Presentations

“Clustering alignment for single cell analysis facilitates model comparison and characterization of informative genes contributing to cellular heterogeneity.” *The American Society of Human Genetics 2025 Annual Meeting (ASHG 2025)*. Boston, MA, Oct 2025.

“ML-MAGES: a computationally efficient tool for multivariate genetic association analyses with effect size shrinkage via machine learning.” *ASHG 2024*. Denver, CO, Nov 2024. (**Reviewers’ Choice Abstract - top 10%.**)

“Mathematical properties of allele-sharing dissimilarities.” *The Allied Genetics Conference (TAGC) 2024*. Metro Washington, DC, Mar 2024.

“Modeling alignment cost in mixed-membership unsupervised genetic clustering.” *Population, Evolutionary, and Quantitative Genetics Conference (PEQG) 2022*. Pacific Grove, CA, May 2022.

“Optimizing the quality-cost trade-off of human annotation for labeling webpages to train webpage classifiers.” (*Sucar E**, **Liu X***, *Khillan S**, *Li Z**) *Nebraska Conference for Undergraduate Women in Mathematics (NCUWM)*. Lincoln, NE, Feb 2017.

“Optimizing the quality-cost trade-off of human annotation for labeling webpages to train webpage classifiers.” (*Khillan S**, *Sucar E**, **Liu X***, *Li Z**) *AMS Contributed Paper Session on Undergraduate Research at Joint Mathematics Meetings*. Atlanta, GA, Feb 2017. (*Poster accepted; unable to attend.*)

RESEARCH AND PROJECT EXPERIENCE

Selected Research Projects

Integrative Analysis of Single-Cell Omics Clustering and Feature Identification, Postdoctoral Researcher

Dr. Sohini Ramachandran, Data Science Institute, Brown University 2024-

- Develop a framework inspired by population-genetics clustering to align hard and mixed-membership clustering results from single cell and spatial transcriptomics data.
- Enable model comparison, consistency evaluation, and identification of clustering-informative genes.
- Demonstrate the framework on benchmark single-cell data and breast cancer spatial transcriptomics data.

Machine Learning for Multivariate Genetic Association Analysis, Postdoctoral Researcher

Dr. Sohini Ramachandran, Data Science Institute, Brown University 2023-25

- Developed *ML-MAGES*, a computationally efficient framework leveraging supervised neural networks and variational inference for mixture models to achieve effect-size shrinkage and multi-trait effect clustering from genome-wide association summary statistics.
- Applied the method to UK Biobank data to perform gene-level analyses and visualized shared versus trait-specific genetic associations.

Optimization-based Methods for Alignment of Population Structure Analyses, Graduate Student Researcher

Dr. Noah Rosenberg, Department of Biology, Stanford University 2021-23

- Investigated the alignment problem of multiple clustering results in population structure analysis.
- Devised a theoretical measure quantifying individual contributions to clustering misalignment and demonstrated its application to ancestry inference using HGDP microsatellite data.
- Designed *Clumpling*, an optimization-based framework for efficiently aligning multiple clustering results with improved quality and runtime, released as a Python package.

Ancestry-Specific Genetic Association Study of COVID-19 Susceptibility, Graduate Student Researcher

Dr. Alexander Ioannidis, Biomedical Data Science & ICME, Stanford University 2020-21

- Contributed to an institution-wide multi-omic effort to track and study the COVID-19 pandemic.
- Analyzed genome sequence, risk factor, and demographic data collected from Stanford Hospital.
- Identified chromosomal segments with potential ancestry-specific associations to infection severity.

Other Relevant Projects

Improving Fundraiser Efficacy Using Large Language Models, Data Scientist Intern

PayPal Global Data Science Group, San Jose, CA

2022

- Built tree-based models on text-derived features and applied explainable AI techniques to interpret results and generate actionable writing suggestions for fundraiser posts.
- Trained transformer-based large language models for text classification and guided text generation to suggest higher-quality headlines.
- Delivered a baseline model improving one-third of fundraisers and an advanced model achieving improvements in over half of the cases.

Latent Embeddings of Phylogenetic Trees via Variational Autoencoders, Graduate Student Researcher

Dr. Julia Palacios, Department of Statistics, Stanford University

2021-22

- Built a multi-task supervised neural network to visualize matrix encodings of phylogenetic trees in a low-dimensional latent space.
- Applied the model to distinguish branching patterns and demographic histories.

Spatial Analysis of Relationships Between Environmental and Socioeconomic Factors, Team Project

Data Science for All (DS4A): Women's Summit, Correlation One

2020

- Performed geographically weighted regression and spatial cross-correlation analysis to analyze spatial relationships between public health risks, socioeconomic disparities, and seismic hazards across Bay Area.
- Visualized correlation between various factors and provided data-driven insights for local policy making.

TEACHING AND MENTORING EXPERIENCE

Undergraduate Research Mentor, Ramachandran Lab, *Brown University*

2025-

- Co-supervise an undergraduate student in summer research program on advancing interactive visualization for clustering alignment in population genetics.
- Guide the student in project design, methodological thinking, communication, and scientific writing.
- Encourage independent learning and professional growth, leading toward a manuscript in preparation.

Instructor, Data Science, Computation, and Visualization (DSCoV) Workshop, *Brown University*

2025

- Prepare materials and deliver a workshop on topic modeling and its applications within and beyond NLP.

Instructor, Advanced MATLAB for Scientific Computing, *Stanford University*

Winter 2022, 2023

- A 4-week short course for students of all levels interested in advanced MATLAB features and toolboxes.
- Designed course syllables, prepared materials, and delivered lectures both over Zoom and in class.
- Open-sourced teaching materials on MATLAB File Exchange with >1.3K downloads (as of Oct 2025).

Graduate Teaching Assistant, *Stanford University*

Linear Algebra with Application to Engineering Computations (CME 200/ME 300A)

Dr. Parviz Moin, Department of Mechanical Engineering, Stanford University

Fall 2021

Dr. Margot Gerritsen, ICME, Stanford University

Fall 2020

Overall effectiveness rating in course evaluation: average of 4.31/5 from 161/173 responses over 2 offerings

- A lecture-style course for approximately 90 graduate students in engineering.
- Held office hours, led recitation sessions, designed exercise problems and graded problem sets and exams.

Ordinary Differential Equations for Engineers (CME 102)

2020-22

Dr. Hung Le, ICME, Stanford University

Overall effectiveness rating in course evaluation: average of 4.26/5 from 97/133 responses over 3 offerings

- A lecture-style course for approximately 50 undergraduate students in engineering.
- Held office hours, led recitation sessions, designed exercise problems and graded problem sets and exams.

Summer Workshop on Intro to Machine Learning, *ICME*

2021

Peer Mentor , <i>School of Engineering, Washington University in St. Louis</i>	2015-17
Undergraduate Teaching Assistant , <i>School of Engineering, Washington University in St. Louis</i>	2015-17
Engineering Freshman Seminar (EN 120, Course Assistant)	
Algorithms and Data Structure (CSE 241/247)	
Introduction to Electrical Engineering (ESE 103)	
Signals and Systems (ESE 351)	

LEADERSHIP AND SERVICE

Ad hoc Reviewer: <i>PNAS, Theor. Popul. Biol.</i>	
Reviewer (GSA Journals Peer Review Training Program), <i>Genetics</i>	2021-23
Ambassador and Organizing Committee Member, <i>Women in Data Science (WiDS) at Brown University</i>	2024-
<ul style="list-style-type: none"> • Co-organized the Women in Data Science (WiDS) Providence 2024 Conference. • Planned the event agenda and invited speakers for talks and panels. • Brought over 100 students and community members for a day of panels, seminars, and networking events. 	
International Chair, <i>Brown Postdoc Council</i>	2024-
<ul style="list-style-type: none"> • Organized social and networking events for postdoctoral researchers. 	
Reviewer, 2024 <i>SACNAS NDiSTEM Conference</i>	2024
Member, <i>ICME Student Action Group, Stanford University</i>	2021-23
<ul style="list-style-type: none"> • Proposed improvements for the ICME program, discussed proposals with faculty and staff, and monitored the progress of proposal implementation. • Resolved concerns in the general student body and organized student community events. 	
Member and Secretary, <i>Alpha Omega Epsilon Sorority Beta Xi Chapter, Washington University</i>	2016-18
Program Assistant, <i>International Summer Study Programs, Washington University</i>	2015, 2017
Volunteer, <i>Campus Y, Washington University</i>	2015-16

SKILLS

Programming Languages: Python, R, MATLAB, C++, MySQL, Mathematica, JavaScript, HTML
 Software, Packages, and Tools: Git, Slurm, PyTorch, TensorFlow, Hugging Face, PLINK, BEAST 2
 Languages: Chinese, English, German (basic)