Zeyun Lu

Email: zeyun_lu@dfci.harvard.edu **Web**: `zeyunlu.github.io` **Last Updated**: 07/20/2025

Summary

I am a statistical geneticist. My current research interests include: (1) developing computational causal inference methods to understand the genetic architecture of complex diseases, and (2) integrating electronic health records and AI to explore the genetic basis of treatment toxicity in cancer patients.

Education

PhD in Biostatistics 08/2019 - 12/2023
University of Southern California (USC) Los Angeles, CA
Dissertation:
Computational Causal Inference for the Genetic Architecture of Complex Traits

MS in Statistics 09/2015 - 12/2016
University of California, San Diego (UCSD) La Jolla, CA

BS in Mathematics/Economics 09/2012 - 09/2014
University of California, Los Angeles (UCLA) Los Angeles, CA

AA in Business 03/2010 - 06/2012
Shoreline Community College Shoreline, WA

Academic Experience

Research Fellow 07/2024 - Present
Dana-Farber Cancer Institute Boston, MA
Mentor: Alexander Gusev, PhD

Research Fellow 07/2024 - Present
Harvard Medical School Boston, MA
Mentor: Alexander Gusev, PhD

Postdoctoral Scholar 01/2024 - 06/2024
Keck School of Medicine of USC Los Angeles, CA
Mentor: Nicholas Mancuso, PhD

Graduate Research Assistant 08/2020 - 12/2023
Keck School of Medicine of USC Los Angeles, CA
Mentor: Nicholas Mancuso, PhD

Research Internship 09/2022 - 06/2024
Dana-Farber Cancer Institute Online/Boston, MA
Mentor: Alexander Gusev, PhD

Selected Publications

2. **Improved multiancestry fine-mapping identifies *cis*-regulatory variants underlying molecular traits and disease risk**. **Lu Z**, Wang X, Carr M, Kim A, Gazal S, Mohammadi P, Wu L, Pirruccello J, Kachuri L, Gusev A, Mancuso N. **Nature Genetics**. July, 2025.
DOI: `https://doi.org/10.1038/s41588-025-02262-7`
1. **Multi-ancestry fine-mapping improves precision to identify causal genes in transcriptome-wide association studies**. **Lu Z***, Gopalan S*, Yuan D, Conti D, Pasaniuc B, Gusev A, Mancuso N. **American Journal of Human**

Genetics. August, 2022.
DOI: https://doi.org/10.1016/j.ajhg.2022.07.002
* Equal Contribution

Other Publications

6. **Inferring causal cell types of human diseases and risk variants from candidate regulatory elements**. Kim A, Zhang Z, Legros C, **Lu Z**, de Smith A, Moore J, Mancuso N, Gazal S. **Under Review at Nature Genetics**. May, 2024.
DOI: https://doi.org/10.1101/2024.05.17.24307556
5. **Genes with differential expression across ancestries are cell-type-specific and enriched in ancestry-specific disease effects**. Wang J, Zhang Z, **Lu Z**, Mancuso N, Gazal S. **American Journal of Human Genetics**. August, 2024.
DOI: https://doi.org/10.1016/j.ajhg.2024.07.021
4. **Characterizing prostate cancer risk through multi-ancestry genome-wide discovery of 187 novel risk variants**. Wang A, ..., Haiman C. **Nature Genetics**. November, 2023.
DOI: https://doi.org/10.1038/s41588-023-01534-4
3. **Novel insight into the etiology of ischemic stroke gained by integrative transcriptome-wide association study**. Jung J, **Lu Z**, de Smith A, Mancuso N. **Human Molecular Genetics**. October, 2023.
DOI: https://doi.org/10.1093/hmg/ddad174
2. **twas_sim, a Python-based tool for simulation and power analysis of transcriptome-wide association analysis**. Wang X, **Lu Z**, Bhattacharya A, Pasaniuc B, Mancuso N. **Bioinformatics**. April, 2023.
DOI: https://doi.org/10.1093/bioinformatics/btad288
1. **Integrative genomic analyses identify susceptibility genes underlying COVID-19 hospitalization**. Pathak GA*, Singh K*, Miller-Fleming TW, Wendt FR, Ehsan N, Hou K, Johnson R, **Lu Z**, Gopalan S, Yengo L, Mohammadi P, Pasaniuc B$, Polimanti R$, Davis LK$, Mancuso N$. **Nature Communication**. July, 2021.
DOI: https://doi.org/10.1038/s41467-021-24824-z
* Equal Contribution
$ Jointly Supervision

Invited Talks

5. Presentation Talk — 05/2025
STATGEN 2025 — Minneapolis, MN

4. Zhou Lab Journal Club — 03/2024
Department of Biostatistics Journal Club — Online
University of Michigan, Ann Arbor

3. Platform Talk — 11/2023
American Society of Human Genetics (ASHG) — Washington, DC

2. Presentation Talk — 10/2023
Program in Quantitative Genomics (PQG) Conference — Boston, MA
Harvard University

1. Division of Epidemiology Journal Club 01/2023
   Vanderbilt University Medical Center Online

Judge Experience

- **Scientific Journal Reviewer**
  Served as a reviewer **96** times for **33** scientific journals, including **Nature Genetics**, **Nature Communication**, **American Journal of Human Genetics**, **Molecular Psychiatry**, Genome Biology, Genetics, BMC Cancer, BMC Genomics, BMC Medicine, BMC Urology, BMC Pulmonary Medicine, BMC Psychiatry, BMC Women's Health, Human Genomics, Human Genetics, Gene, Annals of Human Genetics, Hereditas, Immunity & Ageing, Scientific Reports, Neurosurgical Review, Journal of Ovarian Research, Naunyn-Schmiedeberg's Archives of Pharmacology, Cell Reports, Molecular Medicine, npj Cardiovascular Health, Nutrition & Metabolism, BioData Mining, Clinical Epigenetics, and Discover Oncology, Archives of Dermatological Research, Cardiovascular Diabetology – Endocrinology Reports, Molecular Neurobiology.
- **Contest Judge**
  Served as a Round 1 and Round 2 judge for the 2025 DNA Day Essay Contest, hosted by **American Society of Human Genetics**.

Grant Writing Experience

1. **Microsoft Research PhD Fellowship** (Submitted)
   Title: Computational Causal Inference to Identify Disease-related Genes and Molecular Mechanisms Leveraging Summary-based Functional Data Across Populations

Mentorship Experience

As a PhD Student

- **scFM Project**
  Mentored a fellow PhD student in developing a statistical method for fine-mapping single-cell eQTLs.
  Publication: Rui et al., in preperation.
- **twas_sim Project**
  Mentored a fellow PhD student in developing Python-based simulation software for Transcriptome-wide Association Studies.
  Publication: Wang et al., Bioinformatics, 2023.
  DOI: `https://doi.org/10.1093/bioinformatics/btad288`

Awards & Fellowship

4. **Commencement Student Speaker** 06/2024
   Keck School of Medicine Commencement Ceremony (PhD, DNAP, MPH, MS)
   University of Southern California
   Video Link: `https://www.youtube.com/live/DGYKWieETKM?si=OlQT91KsJwd1GuUW&t=2825`

3. **Third Place Winner** 12/2021
   Video Competition for Keck School of Medcine of USC graduate students to showcase their research.
   Video Link: https://youtu.be/Nc5SVhz7-aQ

2. **Outstanding Overall Performance** 08/2020
   In recognition of top PhD students in the annual screening exams.

1. **Graduate School Fellowship** 2019/2020
   In recognition of new PhD students' outstanding academic promise at Keck School of Medicine of USC.

Work Experience

Statistician 04/2017 - 06/2019
Alzheimer's Therapeutic Research Institute (ATRI) San Diego, CA
Keck School of Medicine of USC
Mentor: Rema Raman, PhD and Michael Donohue, PhD

- Participated as Biostatistician in study teams of several Alzheimer's Disease clinical trials and observational studies such as ADNI3, A4, and LEADS by designing and generating different study reports including but not limited to performing data analysis with mixed models, investigating, data anomalies, and implementing patient randomization.
- Designed, built, and administered R Shiny projects that created dynamic online clinical data reporting system and risk-based monitoring system for ADNI3.
- Mentored student intern in statistical analysis and clinical trial methodologies.

Posters & Abstracts

11. **An integrative approach with Perturb-seq, *cis*-eQTL, and GWAS data identifies mediating genes in regulatory networks of complex traits**
    - Conferences: ASHG 2023, STATGEN 2025,
    - Awards: Platform Talk (ASHG 2023), Invited Talk (STATGEN 2025)
10. **scFM: An efficient statistical fine-mapping approach for eQTLs using large-scale single-cell data**
    - Conferences: ASHG 2024
9. **Fine-mapping causal cell-types of human diseases and disease variants**
    - Conferences: ASHG 2023
8. **Multi-ancestry fine-mapping of gene expression data identifies more *cis*-regulatory mechanisms**
    - Conferences: ASHG 2023, PQG 2023, ProbGen 2023, BoG 2024.
    - Awards: Stellar Abstract Award, Invited Speaker (PQG)
7. **Residual Proteome-wide Association Study Identifies Genes for Blood-Related Traits**
    - Conferences: ASHG 2021
    - Awards: Reviewers' Choice
6. **Novel insight into the etiology of ischemic stroke gained by integrating human transcriptome-wide association study with rodent expression data**
    - Conferences: ASHG 2021.
5. **Multi-ethnic fine-mapping improves precision to identify causal genes in transcriptome-wide association studies**
    - Conferences: ProbGen 2021, ASHG 2020.
4. **Integrative analyses identify genes and their functional consequences underlying COVID-19 hospitalization**
    - Conferences: ACMG Annual Clinical Genetics Meeting 2021.

- Publications: Molecular Genetics and Metabolism 04/2021.

3. **Integrative analyses with large-scale COVID-19 GWAS identifies susceptibility genes underlying hospitalized outcomes**
   - Conferences: ASHG 2020.
2. **Brain Amyloid Burden, Sleep, and Circadian Rest/Activity Rhythms: Screening Findings from A4 and LEARN**.
   - Conferences: AAIC 2017.
   - Publications: Alzheimer's & Dementia. 11/2018.
1. **An Open-source Implementation of Data Standards for Alzheimer's Disease Clinical Trials**
   - Conferences: CTAD. 2017.

TAship

2. **Multi-omics Boot Camp of SHARPP Training**
   - Online course at Mailman School of Public Health, Columbia University.
   - Time: 06/2023, 01/2023, 06/2022, 06/2021.
1. **PM-522A: Introduction to the Theory of Statistics**
   - Graduate level course at KSOM, USC.
   - Time: Fall 2021.

Developed Software

6. **Mr. PEG**
   - A command-line Python software to perform Mendelian Randomization integrating Perturb-seq, eQTL, and GWAS summary data with high-speed inference by leveraging JAX JIT compilation.
   - `https://github.com/gusevlab/mrpeg`
5. **SuShiE**
   - A command-line Python software to perform SNP fine-mapping on molecular QTL data with high-speed inference by leveraging JAX Just In Time compilation.
   - `https://github.com/mancusolab/sushie`
4. **TWAS Simulator**
   - A Python software to perform Transcriptome-wide Association Studies (TWAS) simulations.
   - `https://github.com/mancusolab/twas_sim`
3. **MA-FOCUS**
   - A command-line Python software to perform gene-trait fine-mapping from TWAS in either single ancestry or multi-ancestry settings.
   - `https://github.com/mancusolab/ma-focus`
2. **ADNI3 and LEADS Dynamic Web Report**
   - Two Shiny apps that transform static reports to interactive web pages for clinical trails investigators' monitoring on enrollment, primary outcomes, and safety.
   - Interval apps that are not public accessible.

1. **ADNI3, A4, and LEADS**
    - Three R packages that crawl EDC API, transform data into CDISC format, and automatically build several reports (enrollment, safety etc) for clinical research.
    - Interval packages that are not public accessible.

Membership American Statistical Association. American Society of Human Genetics.

Technical Skills

- **Theoretical knowledge**: Statistics, Data analysis, Large Language Model, AI, Machine learning, Deep learning, Genomics analysis, Bayesian Statistics.
- **Programming**: R (Tidyverse, Shiny), Python (Pytorch), Linux, SAS, Stata.
- **Genetics Tools**: plink, gcta, bedtools, LDSC.
- **Others**: All of Us Platform, Git, Latex, CDISC.

References

- **Alexander Gusev, PhD**
  Associate Professor
  Dana-Farber Cancer Institute/Harvard Medical School
  Postdoc Advisor
- **Nicholas Mancuso, PhD**
  Associate Professor
  University of Southern California
  PhD and Postdoc Advisor
- **Steven Gazal, PhD**
  Assistant Professor
  University of Southern California
  PhD Dissertation Committee Member