Zeyun Lu

Cell: (858)207-8318 **Email**: zeyun_lu@dfci.harvard.edu **Web**: zeyunlu.github.io
Address: 375 Longwood Ave, Boston, MA. 02215 **Last Updated**: 07/25/2023

Education

PhD in Biostatistics 08/2019 - 12/2023
University of Southern California (USC) Los Angeles, CA

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

Academic Experience

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

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

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

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

Research Internship 05/2022 - 06/2024
Dana-Farber Cancer Institute Online/Boston, MA
Mentor: Dr. Alexander Gusev

Working Experience

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

- 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 R Shiny, clinical trials, and statistical analysis.

Selected Publications

2. **Improved multi-ancestry 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, Gusev A, Pirruccello J, Kachuri L, Mancuso N.

Under Review at Nature Genetics. Aprial, 2024.
DOI: https://doi.org/10.1101/2024.04.15.24305836

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. **Under Review at Nature Communication**. October, 2023.
DOI: https://www.medrxiv.org/content/10.1101/2023.10.20.23297214v1
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

Awards & Fellowship

4. **Commencement Student Speaker** 06/2024
Keck School of Medicine of the University of Southern California (PhD, DNAP, MPH, MS) Commencement Ceremony.
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.

Invited Talks

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. Invited 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

Posters & Abstracts

9. **An integrative approach with Perturb-seq, eQTL, and GWAS data identifies mediating genes in regulatory networks of complex traits**.
   - Conferences: ASHG 2023
   - Awards: Platform Talk
8. **Multi-ancestry fine-mapping of gene expression data identifies more *cis*-regulatory mechanisms**.
   - Conferences: ASHG 2023, PQG 2023, ProbGen 2023.
   - 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.
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.

Reviewership 21 times in total: Nature Genetics, Nature Communication, American Journal of Human Genetics, Molecular Psychiatry, Genome Biology, Genetics, BMC Cancer, BMC Genomics, BMC Urology, BMC PulmonaryMedicine, Scientific Reprots.

Membership American Statistical Association. American Society of Human Genetics.

Skills R (Tidyverse, Shiny), Python, Linux, plink, gcta, bedtools, LDSC, Git, Latex, CDISC, SAS, Stata.

Reference

- **Alexander Gusev, PhD**
  Title: Associate Professor
  Affiliation: Dana-Farber Cancer Institute/Harvard Medical School
  Email: alexander_gusev@dfci.harvard.edu
  Relationship: Postdoc Advisor
- **Nicholas Mancuso, PhD**
  Title: Associate Professor
  Affiliation: University of Southern California
  Email: nmancuso@usc.edu
  Relationship: PhD Advisor
- **Steven Gazal, PhD**
  Title: Assistant Professor
  Affiliation: University of Southern California
  Email: gazal@usc.edu
  Relationship: PhD Dissertation Committee Member