Luli S Zou

zou@g.harvard.edu • lulizou.github.io • Google Scholar

EDUCATION

Sept. 2018-May 2023 **Harvard University**   Boston, MA
(expected) *Doctor of Philosophy, Biostatistics*
Dissertation: Statistical and computational methods for spatial genomics
NSF GRFP Fellow

Sept. 2012-May 2016 **University of Wisconsin - Madison**   Madison, WI
Bachelor of Science, Genetics, Statistics
With distinction (Genetics) and honors (Statistics)

RESEARCH AND WORK EXPERIENCE

Jan. 2019-Present   *Graduate Research Fellow*   Boston, MA
Harvard University, Department of Biostatistics; Broad Institute; Dana Farber Cancer Institute
Two primary projects: 1) Designed a probabilistic framework for deconvolution of chromatin interaction signal in 3D genome data. Demonstrated that deconvolution localizes noisy RNA-DNA interaction signal and improves power for differential loop analysis. 2) Designed a beta-binomial generalized additive model with smoothing splines for maximum likelihood estimation and hypothesis testing for 2D spatial allele-specific expression (ASE) patterns in spatial transcriptomics (R package on GitHub: spASE). Demonstrated that some ASE spatial effects are attributable to cell type alone; discovered cell type-specific ASE explainable by differential transcription factor binding.
Advisors: Rafael A. Irizarry, Martin J. Aryee.

June 2022-Present   *Summer Intern*   South San Francisco, CA
Genentech Research & Development, Oncology Bioinformatics
Implemented a stochastic gradient descent algorithm for non-convex optimization in Python and JAX for estimation of sparse gene regulatory network edge weights (GRNs) from single-cell RNA-seq and single-cell ATAC-seq. Designed simulations to test the improvement of the method over existing GRN methods. Project continued after the summer as a collaboration.
Managers: Xiaosai Yao, Jack Kamm

July 2016-July 2018   *Post-baccalaureate Intramural Training Award Program Trainee*   Bethesda, MD
National Institutes of Health, National Human Genome Research Institute
Implemented random forest and convolutional neural network methods to impute whole-genome DNA methylation values; developed a gradient boosting method publicly available as an R package (BoostMe). Tested for significant differences in methylation among tissues associated with type 2 diabetes incorporating patient phenotype data.
Advisor: Francis S. Collins.

Aug. 2013-May 2016   *Undergraduate Research Assistant*   Madison, WI
University of Wisconsin - Madison, Dept. of Medicine, Gastroenterology and Hepatology
Completed an honors thesis (Statistics), "Evidence for Early Mutation Accumulation in Human Colorectal Polyps Using Approximate Bayesian Computation"
Designed and implemented an approximate Bayesian computation simulation of colorectal tumor growth to infer the mutational timeline of human colorectal polyps in C++ using targeted sequencing data as input. Applied the method to show that mutational accumulation happens early in the development of colorectal polyps.
Advisors: Richard B. Halberg, Michael A. Newton, Bret R. Larget.

PUBLICATIONS

As first author:
1. **Zou LS**, Cable DM, Jachowicz JW, Irizarry RA, Aryee MJ. Probabilistic deconvolution of chromatin interaction signal in 3D genome data. *In preparation*. (manuscript available upon request)

2. **Zou LS**, Barrera I, Zhao T, Cable DM, Murray E, Aryee MJ, Irizarry RA. Detection of allele-specific expression in spatial transcriptomics with spASE. *In revision.* https://doi.org/10.1101/2021.12.01.470861
3. **Zou LS**, Erdos MR, Taylor DL, Chines PS, Varshney A, Parker SCJ, Collins FS, Didion JP. BoostMe accurately predicts DNA methylation values in whole-genome bisulfite sequencing of multiple human tissues. *BMC Genomics* 19, 390 (2018). https://doi.org/10.1186/s12864-018-4766-y

As a contributing author:

4. Cable DM, Murray E, Shamugam V, Zhang S, **Zou LS**, Diao M, Chen H, Macosko EZ, Irizarry RA, Chen F. Cell type-specific inference of differential expression in spatial transcriptomics. *Nature Methods* 19, 1076-1087 (2022).
5. Tak YE, Horng JE, Perry NT, Schultz HT, Iyer S, Yao Q, **Zou LS**, Aryee MJ, Joung JK. Augmenting and directing long-range CRISPR-mediated activation in human cells. *Nature Methods* 18, 1075-1081 (2021). https://doi.org/10.1038/s41592-021-01224-1
6. Cable DM, Murary E, **Zou LS**, Goeva A, Macosko EZ, Chen F, Irizarry RA. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature Biotechnology* (2021). https://doi.org/10.1038/s41587-021-00830-w
7. Johnstone SE, Reyes A, Qi Y, Adriaens C, Hegazi E, Pelka K, Chen JH, **Zou LS**, Drier Y, Hecht V, Shoresh N, Selig MK, Lareau CA, Iyer S, Nguyen SC, Joyce EF, Hacohen N, Irizarry RA, Zhang B, Aryee MJ, Bernstein BE. Large-scale topological changes restrain malignant progression in colorectal cancer. *Cell* 182(6) 1474-1489.e23 (2020). https://doi.org/10.1016/j.cell.2020.07.030
8. Rai V, Quang DX, Erdos MR, Cusanovich DA, Daza RM, Narisu N, **Zou LS**, Didion JP, Guan Y, Shendure J, Parker SCJ, Collins FS. Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell type-specific type 2 diabetes regulatory signatures. *Molecular Metabolism* 32 109-121 (2020). https://doi.org/10.1016/j.molmet.2019.12.006
9. Sievers CK, **Zou LS**, Pickhardt PJ, Matkowskyj KA, Albrecht DM, Clipson L, Bacher JW, Pooler BD, Moawad FJ, Bash BD, Reichelderfer M, Vo T, Newton MA, Larget BR, Halberg RB. Subclonal diversity arises early even in small colorectal tumours and contributes to differential growth fates. *Gut* (2016). https://doi.org/10.1136/gutjnl-2016-312232

TECHNICAL SKILLS

Computation
High proficiency**:** R, Python, R Shiny, UNIX shell/bash; high-performance computing cluster
Some proficiency: jax, Keras, PyTorch, C++
Bioinformatics
Methylation (WGBS, EPIC), 3D genome organization (Hi-C, HiChIP, RD-SPRITE), transcriptomics (spatial, single-cell), ATAC-seq (bulk, single-cell), alignment (bwa, bowtie, STAR), samtools, bedtools

RELEVANT GRADUATE COURSEWORK

At Harvard T.H. Chan School of Public Health:
      BIOSTAT 238 Principles and Advanced Topics in Clinical Trials
      EPI 511   Advanced Population and Medical Genetics
      BIOSTAT 245 Analysis of Multivariate and Longitudinal Data
      BIOSTAT 249 Bayesian Methodology
      BIOSTAT 235 Advanced Regression and Statistical Learning
      STAT 221   Monte Carlo Methods & other computational tools for statistical learning
      BIOSTAT 230 Probability I
      BIOSTAT 232 Methods I
      BIOSTAT 231 Statistical Inference I
At MIT:
      MIT 6.867   Machine Learning

TEACHING EXPERIENCE

Harvard T.H. Chan School of Public Health
Fall 2022   BST 260: Introduction to Data Science (R), taught 1 lecture on R Shiny
Fall 2021   BST 260: Introduction to Data Science (R), TA rating 4.8/5.0 (9 responses),
Fall 2020   BST 227: Introduction to Statistical Genetics, TA rating 4.7/5.0 (13 responses)

Fall 2020	BST 273: Introduction to Programming (Python), TA rating 4.9/5.0 (13 responses)	
Spring 2020	BST 281: Genomic Data Manipulation, TA (no rating due to pandemic)	
Fall 2019	BST 273: Introduction to Programming (Python), TA rating 5.0/5.0 (15 responses)	

HONORS AND AWARDS

April 2018	NSF Graduate Research Fellowship	
Oct. 2017	NHGRI Symposium Best Traditional Scientific Poster Award	Bethesda, MD
May 2017, 2018	NIH Post-baccalaureate Poster Day Outstanding Poster Award	Bethesda, MD
April 2015	Hilldale Undergraduate/Faculty Research Fellowship	Madison, WI
April 2015	Second Place - Midwest Undergraduate Data Analytics Competition	Winona, MN
June 2014-July 2014	Summer Institute for Training in Biostatistics	Madison, WI
Aug. 2012-May 2016	Academic Excellence Scholarship	Madison, WI
Aug. 2012	AT&T Foundation Scholarship	Madison, WI
Aug. 2012	Kemper K Knapp Scholarship	Madison, WI
May 2012	National Merit Scholar	Madison, WI

PRESENTATIONS

	Probabilistic deconvolution of chromatin interaction signal in 3D genome data	
Sep 2022	Epigenetics and Chromatin (poster presentation)	Cold Spring Harbor, NY
	Analysis of allele-specific expression in single cell and spatial transcriptomics	
Oct 2022	University of Sydney Statistical Bioinformatics Seminar (remote, invited talk)	Sydney, Australia
May 2022	Biology of Genomes (poster presentation)	Cold Spring Harbor, NY
June 2022	Broad Institute Primer on Medical and Population Genetics (invited talk)	Cambridge,MA
	Differentially methylated regions in whole-genome bisulfite sequencing of type 2 diabetes-relevant human tissues	
May 2018	NIH Post-baccalaureate Poster Day (poster presentation)	Bethesda, MD
	DNA methylation imputation using machine and deep learning	
Dec. 2017	NIH Postbac Elevator Pitches (oral presentation)	Bethesda, MD
Oct. 2017	NHGRI Scientific Symposium (poster presentation)	Bethesda, MD
May 2017	NIH Post-baccalaureate Poster Day (poster presentation)	Bethesda, MD
April 2017	NHGRI Friday Floor Forum (oral presentation)	Bethesda, MD
	Evidence for early mutation accumulation in human colorectal polyps using approximate Bayesian computation	
April 2016	Statistics Honors Poster Session (poster presentation)	Madison, WI
April 2016	Senior Honors Thesis Symposium (oral presentation)	Madison, WI
April 2016	Undergraduate Research Symposium (poster presentation)	Madison, WI

SCIENTIFIC SERVICE

| Summer 2019, 2021 | **StatStart** | Boston, MA |

Harvard T.H. Chan School of Public Health
Volunteer Instructor (2019), *Mentor* (2021)
Taught a 2-day course on introduction to data visualization in R (2019) and mentored students from underserved backgrounds through a pose prediction machine learning project (2021).

| Summer 2019-2021 | **Summer Program in Biostatistics and Computational Biology** | Boston, MA |

Harvard T.H. Chan School of Public Health
Mentor
Guided undergraduate students from underserved backgrounds through a 1-month summer research project with a biostatistics faculty member.

| Nov. 2018-Feb. 2019 | **Health Professions Recruitment & Exposure Program (HPREP)** | Boston, MA |

Mentor
Meeting weekly with a Boston area high school student from an underserved background to work on academic, personal, and professional improvement.

Sept. 2017-May 2018 **College Bound**   Washington, D.C.
                            Mentor
                            Meeting weekly with a D.C. area student from an underserved background to work on academic development and discuss personal and professional goals related to college.

Sept. 2016-May 2018 **STEM in your "Hood"**   Washington, D.C.
                            Volunteer STEM tutor (Sept. 2016 – Jan. 2017)
                            Postbac Service Committee College Bound volunteer coordinator (Jan. 2017 – May 2018)
                            Tutoring underserved D.C. area students grades 5-12 in STEM subjects biweekly through College Bound. Organizing groups of Postbacs to tutor students every Monday, Wednesday, and Thursday during the D.C. public school year.