Zilin Li, Ph.D.

CONTACT INFORMATION

Address:
5268 Renmin Street
Changchun, Jilin 130024

Phone: +1(857)272-5060
Email: lizl@nenu.edu.cn
https://zilinli1988.github.io/

RESEARCH INTERESTS

Methodology:
Statistical genetics, high-dimensional statistics

Applications:
Large scale genetic and genomic data, whole-genome sequencing data

ACADEMIC APPOINTMENTS

Northeast Normal University
Professor, School of Mathematics and Statistics 2023-present

Indiana University School of Medicine
Assistant Professor, Department of Biostatistics and Health Data Science 2022-2023

Harvard T.H.Chan School of Public Health

Research Scientist, Department of Biostatistics 2021-2022

Research Associate, Department of Biostatistics 2019-2021

Research Fellow, Department of Biostatistics 2016-2019

- Advisor: Professor Xihong Lin

EDUCATION

Tsinghua University
Ph.D. in Statistics 2011-2016

- Advisor: Professor Xihong Lin
- Thesis: Statistical methods for large-scale sequencing data

Tsinghua University
B.S. in Mathematics 2007-2011

EDITORIAL ACTIVITIES

Guest Editor, **Frontiers in Genetics**, Research Topics in Statistical Genetics and Methodology, 2021-2022

HONORS AND AWARDS

Elected Member, International Statistical Institute 2023

Institute of Mathematical Statistics New Researcher Travel Award, Institute of Mathematical Statistics 2021

ASHG "Reviewers Choice" Poster Award, American Society of Human Genetics 2020 Virtual Meeting 2020

NHLBI Biodata Catalyst Fellow, NIH/NHLBI 2020

PQG Travel Award, Harvard T.H. Chan School of Public Health 2017

ASA Student Paper Award, Statistics in Genomics and Genetics Section of the American Statistical Association 2017

Beijing Outstanding Graduates, Beijing 2016

Graduate Paper Award, Second Joint Biostatistics Symposium 2012

Excellent Undergraduate Paper, Tsinghua University 2011

JOURNAL PUBLICATIONS

19. Li, X., Quick, C., Zhou, H., Gaynor, S., Liu, Y., Chen, H., Selvaraj, M., Sun, R., Dey, R., Arnett D., Bielak, L., Bis, J., Blangero, J., Boerwinkle, E., Bowden, D., Brody, J., Cade, B., Correa, A., Cupples, L.A., Curran, J., De Vries, P., Duggirala, R., Freedman, B., Goring, H., Guo, X., Haessler, J., Kalyani, R., Kooperberg, C., Kral, B., Lange, L., Manichaikul, A., Martin, L., McGarvey, S., Mitchell, B., Montasser, M., Morrison, A., Naseri, T., O'Connell, J., Palmer, N., Peyser, P., Psaty, B., Raffield, L., Redline, S., Reiner, A., Reupena, M., Rice, K., Rich, S., Sitlani, C., Smith, J., Taylor, K., Vasan, R., Wilson, J., Willer, C., Yanek, L., Zhao, W., NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed Lipids Working Group, Rotter, J., Natarajan, P., Peloso, G., **Li, Z.**#, Lin, X.# (2023) Powerful and resource-efficient meta-analysis of rare variant association tests in large whole genome sequencing studies at scale. *Nature Genetics*, **55**, 155-164. (# indicates joint correspondence.)

    – Charles J. Epstein Research Semifinalist Award (ASHG, 2020).

18. **Li, Z.***#, Li, X.*, Zhou, H., Gaynor, S.M., Selvaraj, M., Arapoglou, T., Quick, C., Liu, Y., Chen, H., Sun, R., Dey, R., Arnett D., Auer, P., Bielak, L., Bis, J., Blackwell, T., Blangero, J., Boerwinkle, E., Bowden, D., Brody, J., Cade, B., Conomos, M., Correa, A., Cupples, L.A., Curran, J., De Vries, P., Duggirala, R., Franceschini, N., Freedman, B., Goring, H., Guo, X., Kalyani, R., Kooperberg, C., Kral, B., Lange, L., Lin, B., Manichaikul, A., Manning, A., Martin, L., Mathias, R., Mitchell, B., Montasser, M., Morrison, A., Naseri, T., O'Connell, J., Palmer, N., Peyser, P., Psaty, B., Raffield, L., Redline, S., Reiner, A., Reupena, M., Rice, K., Rich, S., Smith, J., Taylor, K., Taub, M., Vasan, R., Week, D., Wilson, J., Yanek, L., Zhao, W., NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed Lipids Working Group, Rotter, J., Willer, C., Natarajan, P., Peloso, G., Lin, X.# (2022) A framework for detecting noncoding rare variant associations of large-scale whole-genome sequencing studies. *Nature Methods*, **19**, 1599-1611. (* indicates equal contribution, # indicates joint correspondence.)

    – "Reviewers' choice" poster award (ASHG, 2020).

17. **Li, Z.**, Liu, Y., Lin X. (2022). Simultaneous detection of signal regions using quadratic scan statistics with applications in whole genome association studies. *Journal of American Statistical Association*, **117(538)**, 823-834.
    - Winner of ASA Student Paper Award (SSGG Section, 2017).
    - Winner of IMS New Researcher Travel Award (2021).
16. Li, X.*, **Li, Z.***, Zhou, H., Gaynor, M., Liu, Y., Chen, H., *et al.* (2020). Dynamic incorporation of multiple in-silico functional annotations empowers rare variant association analysis of large whole genome sequencing studies at scale. *Nature Genetics*, **52**, 969-983. (* indicates equal contribution.)
15. **Li, Z.**, Li, X., Liu, Y., Shen, J., Chen, H., Morrison, A.C., Boerwinkle, E., Lin X. (2019). Dynamic scan procedure for detecting rare-variant association regions in whole-genome sequencing studies. *The American Journal of Human Genetics*, **104(5)**, 802-814.
14. **Li, Z.**, Wang, S., Lin, X. (2012). Variable selection and estimation in generalized linear models with the seamless L_0 penalty. *The Canadian Journal of Statistics*, **40(4)**, 745-769.
    - Winner of Best Graduate Paper Award (Second Joint Biostatistics Symposium, 2012).
13. Feofanova, E., Brown, M., Alkis, T., Manuel, A., Li, X., Tahir, U., **Li, Z.**, *et al.* (2023). Whole-Genome Sequencing Analysis of Human Metabolome in Multi-Ethnic Populations. *Nature Communications*, **14**, 3111.
12. Huang, X., Yao, M., Tian, P., Wong, J., **Li, Z.**, Liu, Z., Zhao, J. (2023). Genome-wide cross-trait analysis and Mendelian randomization reveal a shared genetic etiology and causality between COVID-19 and venous thromboembolism. *Communication Biology*, **6**, 441.
11. Zhou, H., Arapoglou, T., Li, X., **Li, Z.**, *et al.* (2023). FAVOR: functional annotation of variants online resource and annotator for variation across the human genome. *Nucleic Acids Research*, **51(D1)**, D1300-D1311.
10. Liu, Y., **Li, Z.**, Lin, X. (2022). A Minimax Optimal ridge-type set test for global hypothesis with applications in whole genome sequencing association studies. *Journal of American Statistical Association*, **117(538)**, 897-908.
9. Selvaraj, M., Li, X., **Li, Z.**, *et al.* (2022). Whole genome sequence analysis of blood lipid levels in >66,000 individuals. *Nature Communications*, **13**, 5995.
8. Gaynor, S., Westerman, K., Ackovic, L., Li, X., **Li, Z.**, Manning, A., Philippakis, A., Lin, X. (2022), STAAR Workflow: A cloud-based workflow for scalable and reproducible rare variant analysis. *Bioinformatics*, **38(11)**, 3116-3117.

7. Li, X., Yung, G., Zhou, H., Sun, R., **Li, Z.**, Liu, Y., Ionita-Laza,I., Lin, X. (2022), A Multi-dimensional Integrative Scoring Framework for Predicting Functional Regions in the Human Genome. *The American Journal of Human Genetics*, **109(3)**, 446-456.
6. Sun, R., Xu, M., Li, X., Gaynor, S., Zhou, H., **Li, Z.**, *et al.* (2021). Integration of multi-omic annotation data to prioritize and characterize inflammation and immune-related risk variants in squamous cell lung cancer. *Genetic Epidemiology*, **45(1)**, 99-114.
5. Liu, Y., Chen, S., **Li, Z.**, Morrison, A., Boerwinkle, E., Lin, X. (2019). ACAT: A fast and powerful p Value combination method for rare-variant analysis in sequencing studies. *The American Journal of Human Genetics*, **104(3)**, 410-421.
4. Chen, H., Huffman, J., Brody, J., Wang C., Lee, S., **Li, Z.**, *et al.* (2019). Efficient variant set mixed model association tests for continuous and binary traits in large-scale whole-genome sequencing studies. *The American Journal of Human Genetics*, **104(2)**, 260-274.
3. Xu, Miao., Yao, Y., Chen, H., Zhang, S., Cao, S., Zhang, Z., Luo, B., Liu, Z., **Li, Z.**, *et al.* (2019). Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. *Nature Genetics*, **51**, 1131-1136.
2. Chen, H., Wang, C., Conomos, M., Adrienne, M., **Li, Z.**, Sofer, T., Szpiro, A., Chen, W., Brehm, J., Celedon, J., Redline S, Papanicolaou, G., Thornton, T., Laurie, C., Rice, K., Lin, X. (2016). Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *The American Journal of Human Genetics*, **98(4)**, 653-666.
1. Lin, X., Lee, S., Wu, M. C., Wang, C., Chen, H., **Li, Z.**, Lin, X. (2016). Test for rare variants by environment interactions in sequencing association studies. *Biometrics*, **72(1)**, 156-164.

GRANTS

BioData Catalyst Fellowship 03/25/2020 - 06/19/2021
NIH/NHLBI $69,733
A powerful and resource-efficient rare variant meta-analysis workflow for large-scale multi-ethnic sequencing association studies using summary statistics and functional annotations
Role: Principal Investigator

TEACHING EXPERIENCE

Department of Mathematical Sciences, Tsinghua University
Teaching Assistant

- Linear Algebra Fall 2011
- Experiment Design and Data Processing Spring 2012 and Fall 2012
- Inference Spring 2013

Summer School, Tsinghua University
Instructor
- Statistical Methods Summer 2013

TALKS AND PRESENTATIONS

Invited Talks

- "STAARpipeline: an all-in-one rare-variant tool for biobank-scale whole-genome sequencing data", SJTU-Yale Biostatistics Seminar, SJTU, Shanghai, June 2023
- "STAARpipeline: an all-in-one rare-variant tool for biobank-scale whole-genome sequencing data", ICSA 2023 Applied Statistics Symposium, University of Michigan, Ann Arbor, June 2023
- "STAARpipeline: an all-in-one rare-variant analysis tool for biobank-scale whole-genome sequencing data", 10th International Purdue Symposium on Statistics, West Lafayette, June 2023
- "STAARpipeline: an all-in-one rare-variant analysis tool for biobank-scale whole-genome sequencing data", Statistics and Probability Seminar Series, Department of Mathematics, School of Science and Engineering, Tulane University, New Orleans, April 2023
- "STAARpipeline: an all-in-one rare-variant analysis tool for biobank-scale whole-genome sequencing data", Machine Learning Seminar Series, Data Science Initiative, College of Science and Engineering, University of Minnesota, Minneapolis, April 2023
- "STAARpipeline: a comprehensive framework for flexible and scalable rare variant association analysis using whole-genome sequencing data and annotation information", Statistics Seminars, Department of Mathematical Sciences, Indiana University-Purdue University Indianapolis, January 2023 (virtual)
- "STAARpipeline: an all-in-one rare variant tool for biobank-scale whole-genome sequencing data", Plant and Animal Genome 30, Integration of Functional Genomics Data in Genetic and Genomic Research Workshop, San Diego, January 2023
- "STAARpipeline: an all-in-one rare-variant tool for biobank-scale whole-genome sequencing data", 23andme StatGen and CompBio Seminar Series, December 2022 (virtual)
- "Why do we do burden testing", IGVF (Impact of Genomic Variation on Function Consortium) CAMP Methods Workshop, November 2022 (virtual)
- "A powerful and resource-efficient pipeline for association analysis of large-scale whole-genome sequencing studies", 2022 International Chinese Statistical Association (ICSA) China Conference, July 2022 (virtual)
- "A powerful and resource-efficient pipeline for association analysis of large-scale whole-genome sequencing studies", 5th Eastern Asia Chapter

of the International Society for Bayesian Analysis Annal Conference, November 2021 (virtual)

- "Scalable integrative statistical inference for whole-genome sequencing association studies," Lecture of Statistics, Renmin University of China, October 2020 (virtual)
- "Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale," Data Science Seminars, Shanghai Jiao Tong University, October 2020 (virtual)
- "Rare variant association analysis pipeline with application to Freeze 8 TOPMed lipids WGS data of 65,000 individuals," 2020 GSP-TOPMed Analysis Workshop, New York City, NY, February 2020
- "Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale," TOPMed series of FOCuS 2019 seminars, Boston University, Boston, MA, October 2019
- "Simultaneous detection of signal regions with applications in genome-wide association studies," Northeastern Normal University, Changchun, China, September 2017
- "Detecting signal regions in whole genome association studies," Biostatistics in the Era of Big Data 2016, Beijing, China, May 2016
- "Variable selection and estimation in generalized linear models with the Seamless L_0 penalty," Second Joint Biostatistics Symposium, Beijing, China, July 2012

Conference Presentations

- Platform Presentation: "Whole Genome Sequencing Analyses of 45,090 Individuals Reveal Rare Coding and Noncoding Variants Associated with Kidney Function," ASHG, Los Angeles, CA, October 2022
- Invited Poster: "Simultaneous Detection of Signal Regions Using Quadratic Scan Statistics with Applications to Whole Genome Association Studies," JSM, Washington, DC, August, 2022
- Poster: "A powerful and resource-efficient pipeline for association analysis of large-scale whole-genome sequencing studies," ASHG 2021 Virtual Meeting, October 2021 (virtual)
- Poster: "A framework for detecting non-coding rare variant associations in large whole genome sequencing studies at scale, with application to 30,138 TOPMed participants for lipid traits" ASHG 2020 Virtual Meeting, October 2020 (virtual)
- Platform Presentation: "Whole genome sequence association analysis of body mass index in 45,159 TOPMed participants using STAAR," ASHG, Houston, TX, October 2019
- Invited Poster: "Dynamic incorporation of multiple in-silico functional annotations empowers rare variant association analysis of large whole

genome sequencing studies at scale," Annual Genome Sequencing Program Meeting, Bethesda, MD, March 2019

- Contributed talk: "Tuning parameter selection for prediction in high-dimensional ridge regression," ENAR, Atlanta, GA, March 2018
- Poster: "Detection of signal regions in whole genome genotyping and sequencing association studies using scan statistics," ASHG, Orlando, FL, October 2017
- Contributed talk: "Detection of signal regions in whole genome association studies," JSM, Baltimore, MD, August, 2017
- Contributed talk: "Detecting signal regions in whole-genome association studies," JSM, Seattle, WA, August, 2015
- Poster: "Variable selection and estimation in generalized linear models with the seamless L_0 penalty," JSM, Boston, MA, August, 2014

PROFESSIONAL SERVICE

Journal Reviewer

- American Journal of Human Genetics
- Bioinformatics
- Biometrics
- Biostatistics
- BMC Bioinformatics
- Frontier in Genetics
- Genetic Epidemiology
- Genome Research
- Journal of Computational and Graphical Statistics
- Journal of Computational Biology
- Journal of Statistical Planning and Inference
- Journal of the American Statistical Association
- Human Genetics
- Nature Communications
- PLoS Computational Biology
- PLoS Genetics
- PLoS One
- Quantitative Biology
- SSM - Population Health

Conference Reviewer

- American Society of Human Genetics 2023 Meeting
- ICSA 2023 Applied Statistics Symposium Student Paper Competition

Invited Session Organizer

- EAC-ISBA 2023
- ENAR 2023
- NESS 2023

Session Chair

- EAC-ISBA 2023
- ENAR 2023

COMPUTER SKILLS

R, Python, Matlab, SAS, Linux

PROFESSIONAL MEMBERSHIPS

American Society of Human Genetics
American Statistical Association
Eastern North American Region, International Biometric Society
Institute of Mathematical Statistics
International Statistical Institute