Zilin Li, Ph.D.

CONTACT INFORMATION

Address:
5268 Renmin Street
Changchun, Jilin 130024

Phone: +86 18946539790
Email: lizl@nenu.edu.cn
https://zilinli1988.github.io/

RESEARCH INTERESTS

Methodology:

Statistical genetics, high-dimensional statistics, AI for Health

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, **Mathematics**, Special Issue: Statistical Methods and Applications in Genetics and Genomics, 2023-2024

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

HONORS AND AWARDS

Promising Scientist Award of 2023 DAMO Academy Young Fellow,
Alibaba DAMO Academy 2023

Elected Member,
International Statistical Institute 2023

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

SELECTED PUBLICATIONS

*** indicates equal contribution, # indicates joint correspondence**

1. Li, X.*#, Wood, A.*#, Yuan, Y., Zhang, M., Huang, Y., Hawkes, G., Beaumout, R., Weedon, M., Li, W., Li, X., Lin, X.#, **Li, Z.#** (2025) Streamlining large-scale genomic data management: Insights from the UK Biobank whole-genome sequencing data. *Cell Genomics*, **5**, 101009.
2. Zhang, S., Zhang, M., Yuan, Y., **Li, Z.#**, Li, X.#, Li, X.# (2025) Health risks and genetic architecture of objectively measured multidimensional sleep health. *Nature Communications*, **16**, 7026.
3. Li, X., Chen, H., Selvaraj, M.M., Van Buren, E., Zhou, H., Wang, Y., Sun, R., McCaw, Z.R., Yu, Z., Jiang, M., DiCorpo, D., Gaynor, S.M., Dey, R., Arnett, D.K., Benjamin, E.J., Bis, J.C., Blangero, J., Boerwinkle, E., Bowden, D.W., Brody, J.A., Cade, B.E., Carson, A.P., Carlson, J.C., Chami, N., Chen, Y., Curran, J.E., de Vries, P., Fornage, M., Franceschini, N., Freedman, B.I., Gu, C., Heard-Costa, N., He, J., Hou, L., Hung, Y., Irvin, M.R., Kaplan, R., Kardia, S., Kelly, T., Konigsberg, I., Kooperberg, C., Kral, B., Li, C., Li, Y., Lin, H., Liu, C., Loos, R., Mahaney, M., Martin, L., Mathias, R., Mitchell, B., Montasser, M., Morrison, A., Naseri, T., North, K., Palmer, N., Peyser, P., Psaty, B., Redline, S., Reiner, A., Rich, S., Sitlani, C., Smith, J., Taylor, K., Tiwari, H., Vasan, R., Viali, S., Wang, Z., Wessel, J., Yanek, L., Yu, B., Dupuis, J., Meigs, J., Auer, P., Raffield, L., Manning, A., Rice, K., Rotter, J., Peloso, G., Natarajan, P., **Li, Z.#**, Liu, Z. #, Lin, X. # (2025). A statistical framework for multi-trait rare variant analysis in large-scale whole-genome sequencing studies. *Nature Computational Science*, **5**, 125–143.
4. Hawkes, G.*, Beaumont, R.N.*, **Li, Z.***, Mandla, R.*, Li, X.*, *et al.* (2024) Whole-genome sequencing in 333,100 individuals reveals rare non-coding single variant and aggregate associations with height. *Nature Communications*, **15**, 8549.
5. Wang, X.*, Zhang, Q.*, Chen, X., Huang, Y., Zhang, W., Liao, L., Zhang, X., Huang, B.,Huang, Y., Ye, Y., Song, M., Lao, J., Chen, J., Feng, X., Long, X., Liu, Z., Zhu, W., Yu, L., Fan, C., Tang, D., Zhong, T., Fang, M., Li, C., Niu, C., Huang, L., Lin, B., Hua, X., Jin, X., **Li, Z.#**, Xu, X.# (2024) Whole-genome Sequencing Association Analysis of Quantitative Platelet Traits in A Large Cohort of β-thalassemia. *Genomics, Proteomics & Bioinformatics*, qzae065.
6. 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.)

7. **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.)
8. **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.
9. 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.
10. **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.

GRANTS

General Program 01/01/2026-12/31/2028
National Natural Science Foundation of China ¥440,000
Statistical Methods for Meta-analysis of High-Dimensional Genet-

ics Data
Role: Principal Investigator

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 Statistics, Northeast Normal University
Instructor

- Multivariate Regression — Spring 2025
- Statistical Computing — Fall 2024, Fall 2025
- Advanced Topics in Statistics — Fall 2024, Fall 2025

Summer School, Tsinghua University
Instructor

- Statistical Methods — Summer 2013

TALKS AND PRESENTATIONS

Invited Talks

- "A statistical framework for powerful multi-trait rare variant analysis in large-scale whole-genome sequencing studies"
  - Biostatistics Seminar, Peking University, Beijing, October 2024
  - 2024 Annual Meeting of the Chinese Society for Medical Mathematics, Jiaxing, November 2024
- "Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies",
  - GPB Omics & Bioinformatics Frontiers Symposium 2024, August 2024
  - ML Stat Meeting 2024, Changchun, August 2024
  - Annual Meeting of the Chinese Biostatistics Society, Guangzhou, July 2024
  - EcoStat 2024, Beijing, July 2024
  - Joint Conference on Statistics and Data Science in China 2024 (JCSDS 2024), Kunming, July 2024
  - East China Normal University Statistics Seminar, Shanghai, June 2024
  - Center for Statistical Science of Tsinghua, Tsinghua University, Beijing, April 2024

  - School of Statistics and Data Science, Nankai University, Tianjin, March 2024

- "STAARpipeline: an all-in-one rare-variant tool for biobank-scale whole-genome sequencing data",
  - 2023 Annual Meeting of the Chinese Society for Medical Mathematics, Dalian, December 2023
  - Department of Mathematical Sciences, Tsinghua University, Beijing, December 2023
  - School of Public Health, Fujian Medical School, Fuzhou, November 2023
  - School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, October 2023
  - School of Public Health, Nanjing Medical School, Nanjing, October 2023
  - Lin Gang Laboratory, Shanghai, September 2023
  - Center for Evolutionary Biology, Fudan University, Shanghai, September 2023
  - School of Statistics, Southwestern University of Finance and Economics, Chengdu, September 2023
  - School of Public Health, Zhejiang University, Hangzhou, August 2023
  - The 9th International Forum on Statistics (RUC IFS 2023), Beijing, July 2023
  - Joint Conference on Statistics and Data Science in China (JCSDS 2023), Beijing, July 2023
  - 2023 12th ICSA International Conference, Hongkong, July 2023
  - 2023 9th International Statistical Genetics and Genomics Forum, Jinan, July 2023
  - SJTU-Yale Biostatistics Seminar, Shanghai Jiao Tong University, Shanghai, June 2023
  - ICSA 2023 Applied Statistics Symposium, University of Michigan, Ann Arbor, June 2023
  - 10th International Purdue Symposium on Statistics, West Lafayette, June 2023
  - Statistics and Probability Seminar Series, Department of Mathematics, School of Science and Engineering, Tulane University, New Orleans, April 2023
  - Machine Learning Seminar Series, Data Science Initiative, College of Science and Engineering, University of Minnesota, Minneapolis, April 2023
  - Statistics Seminars, Department of Mathematical Sciences, Indiana University-Purdue University Indianapolis, January 2023 (virtual)

  - Plant and Animal Genome 30, Integration of Functional Genomics Data in Genetic and Genomic Research Workshop, San Diego, January 2023
  - 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

- Advanced Science
- American Journal of Human Genetics
- Bioinformatics
- Biometrics
- Biostatistics
- BMC Bioinformatics
- Cell Genomics
- Frontier in Genetics
- Genetic Epidemiology
- Genome Research
- Genomics, Proteomics & Bioinformatics

- Journal of Computational and Graphical Statistics
- Journal of Computational Biology
- Journal of Statistical Planning and Inference
- Journal of the American Statistical Association
- Human Genetics
- Human Genetics and Genomics Advances
- Nature Communications
- Nature Genetics
- 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

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