Jianle Sun

Tel：(+86)15983653157 E-mail：sjl-2017@sjtu.edu.cn Homepage: https://sjl-sjtu.github.io/

Education

Shanghai Jiao Tong University   Sep. 2021-Jun. 2024 (expected)
- **Master of Science (Research-based) in Biostatistics & Bioinformatics**
- **Advisors:** Associate Prof. Yue Zhang & Prof. Zhangsheng Yu
- **GPA:** major 3.83/4.0, total 3.90/4.0
- **Related courses:** Fundamental Mathematical Statistics (A+), Biology Mathematics I (A+), Applied Stochastic Processes (A+), Causal Inference Methods in Data Science (A+), Data Mining (A+), Introduction to Computational Biology (A+), Machine Learning (A), Statistical Learning (A), Medical Bioinformatics (A), Neural Network Theory and Application (A-)
- **Thesis:** Statistical Methods for Genetic Association and Causal Inference with Genomic and Single-cell Transcriptomic Data

Shanghai Jiao Tong University   Sep. 2017- Jun. 2021
- **Bachelor of Science in Biotechnology**
- **GPA** 3.94/4.3 (rank 1/25)
- **Related courses:**
  - **Quantitative:** Calculus I & II (A-), Linear Algebra (A+), Probability & Statistics (A), Thinking & Approach of Programming (A), Biostatistical Models (A+), Biostatistics & Mathematical Modeling (A), Stochastic Simulation Methods & Its Applications (A), Business Analytics & Data Mining (A)
  - **Science/Biology-related:** Physics I & II, Inorganic & Analytic Chemistry, Organic Chemistry, Physical Chemistry, Biochemistry I & II, Genetics, Cell Biology, Molecular Biology, Microbiology, Anatomy & Physiology, Ecology & Evolution, Immunology, Developmental Biology
- **Thesis:** Bayesian Network-based Mendelian Randomization for Causal Inference (A)

Research Interest

Statistical Genetics and Genomics, Causal Inference & Causal Learning, Computational Methods in single-cell omics, Genetic Epidemiology, Bayesian Methods

Publications

- **Sun, J.**, Lyu, R., Deng, L., Li, Q., Zhao, Y., & Zhang, Y. (2022). SMetABF: A rapid algorithm for Bayesian GWAS meta-analysis with a large number of studies included. *PLOS Computational Biology*, *18*(3), e1009948. https://doi.org/10.1371/journal.pcbi.1009948
- **Sun, J.**, Zhou, J., Gong, Y., Pang, C., Ma, Y., Zhao, J., Yu, Z., & Zhang, Y. (2024). Bayesian network-based Mendelian randomization for variant prioritization and phenotypic causal inference. *Human Genetics*. https://doi.org/10.1007/s00439-024-02640-x
- **Sun, J.**, Deng, L., Li, Q., Zhou, J., & Zhang, Y. Dynamic relations between longitudinal morphological, behavioral, and emotional indicators and cognitive impairment: evidence from the Chinese Longitudinal Healthy Longevity Survey. *Under Review*
- Lyu, R., **Sun, J.**, Xu, D., Jiang, Q., Wei, C., & Zhang, Y. (2021). GESLM algorithm for detecting causal SNPs in GWAS with multiple phenotypes. *Briefings in Bioinformatics*, *22*(6), bbab276. https://doi.org/10.1093/bib/bbab276
- Zhou, Y., Fa, B., Wei, T., **Sun, J.**, Yu, Z., & Zhang, Y. (2021). Elastic Correlation Adjusted Regression (ECAR) scores for high dimensional variable importance measuring. *Scientific Reports*, *11*(1), 1-12. https://doi.org/10.1038/s41598-021-02706-0
  Gong, Y., Xu, J., Wu, M., Gao, R., **Sun, J.**, Yu, Z., & Zhang, Y. (2024). Single-cell biclustering for cell-specific transcriptomic perturbation detection in AD progression. *Cell Reports Methods*, 100742. https://doi.org/10.1016/j.crmeth.2024.100742
- Liang, C., Bai, W., Qiao, L., Ren, Y., **Sun, J.**, Ye, P., Yan, H., Ma, X., Zuo, W., & Ouyang, W. Rethinking the BERT-like pretraining for DNA sequences. *ICML 2024 in submission*. https://arxiv.org/abs/2310.07644

Projects

- **Bayesian GWAS meta-analysis methodology and its application to Parkinson's disease**
  - Improve the subset selection process in ABF-based Bayesian GWAS meta-analysis by introducing MCMC and Shotgun Stochastic Search algorithm

- Analyze genetic patterns of Parkinson's disease and other autoimmune disorders through cross-trait Bayesian meta-analysis
- Implement proposed algorithms as an R package and an R shiny online software

- **Bayesian network-based Mendelian randomization for causal inference**
  - Propose the random graph forest, comprised of a series of Bayesian network structure learning processes, to prioritize candidate variants and select proper genetic instruments, and obtained pleiotropy-robust causal estimates through Bayesian Mendelian randomization with a shrinkage prior
  - Apply proposed model in UK Biobank to infer the causality between erythrocyte-related hematologic parameters and blood pressures, and between lymphocyte and psychiatric disorders like schizophrenia

- **Epidemiological analysis on cognitive impairment using Chinese Longitudinal Healthy Longevity Survey (CLHLS) datasets**
  - Identify risk factors of aging cognitive impairment with data of participants' physical indicators, activities of daily life, emotional status, and cognitive abilities from 8 waves of CLHLS datasets
  - Establish dynamic predictive models for cognitive impairment using Bayesian joint model and dynamic-DeepHit, a deep learning-based survival model
  - Assess longitudinal indicators' effect on cognitive decline by controlling confounders using marginal structural Cox model.

- **Dynamic Mendelian randomization with genomics and single-cell transcriptomics**
  - Extend Mendelian randomization to single-cell resolution by integrating eQTL and single-cell transcriptomics
  - Employ functional data analysis (like functional principal analysis) and g-estimation of longitudinal confounding to model dynamic effect in single-cell trajectory

- **Deep learning-based prospective stroke predictions using dynamic longitudinal physiological records**
  - Label longitudinal physiological records according to symptom descriptions in each follow-up
  - Employ Cox model with lasso to select crucial risk factors, RNN module to capture the longitudinal features, and random forest for the final classification on stroke

Internship

- **Shanghai AI laboratory | Research Intern on AI for Science**   Jul. 2023 – current
  - Develop pretrained BERT-based DNA large language model
  - Conduct downstream tasks, including identifying genomic functional elements (promoters, enhancers, etc.), predicting epigenomic & transcriptomic features, and assessing functional effects of variants

- **Baiyi Health | Biodata Analyst**   Aug. 2022 - Jan. 2023
  - Establish biomarker-based biological age models for eight systems/organs, and establishing classification model for thyroid disorders
  - Participate in the trial design and data analysis for the performance of Eye Parameter Evaluator

Honors

- Graduation with honor: College Graduate Excellence Award of Shanghai, 2021 (5%)
- National Encouragement Scholarship
- Fan Hsu-chi Scholarship
- Tang You Shuqi Named Scholarship
- Sino Biological Named Scholarship
- Gold award and nominated for best new application project in the International Genetically Engineered Machine (iGEM) Competition, 2019

Skills

- **Language:** Chinese, English (IELTS 7.5), Japanese (Junior), learning experience in ancient languages like Latin
- **Computer:** Python, R, MATLAB, SPSS, SQL, LaTeX, PyTorch, Linux