
**Empower an end-to-end scalable and interpretable data science ecosystem by
integrating statistics, AI, and Genomics and Health**

Xihong Lin

**Department of Biostatistics and Department of Statistics
Harvard University**

September 26, 2024

1:30PM-2:50PM

Clark S361

Abstract:

The data science ecosystem encompasses data fairness, scalable statistical and ML/AI methods and tools, interpretable data analysis, and trustworthy decision-making. Rapid advancements in AI have revolutionized data utilization and enabled machines to learn from data more effectively. Statistics, as the science of learning from data while accounting for uncertainty, plays a pivotal role in addressing complex real-world problems and facilitating trustworthy decision-making. In this talk, I will discuss the challenges and opportunities involved in building an end-to-end scalable and interpretable data science ecosystem that integrates statistics, AI, and genomics and health. I will illustrate key points using the analysis of large-scale whole genome sequencing data, electronic health records, and biobanks. This talk aims to ignite proactive and thought-provoking discussions, foster collaboration, and cultivate open-minded approaches to advance scientific discovery.

Reading list:

1. Liu, Y., Liu, Z., and **Lin, X.** (2024). Ensemble testing for global null hypothesis. *Journal of the Royal Statistical Society, Series B*, 86(2), pp.461-486.
2. Chen, T., Zhang, H., Mazumder, R. and Lin, X., 2024. Fast and scalable ensemble learning method for versatile polygenic risk prediction. *Proceedings of the National Academy of Sciences*, 121(33), p.e2403210121.
3. Li, X., 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(9), pp.969-983.

4. McCaw ZR, Gao J, Lin X, Gronsbell J (2024). Synthetic surrogates improve power for genome-wide association studies of partially missing phenotypes in population biobanks. *Nature Genetics*. 56, pages 1527–1536.