



Jin Jin
Department of Biostatistics
University of Pennsylvania

PennPRS: a centralized cloud computing platform for efficient polygenic risk score training

Wednesday, September 10th, 2025

12:10-1:30 PM

SEC 117

Zoom Meeting: Meeting ID: 983 2864 8232

Password: 665123

<https://rutgers.zoom.us/j/98328648232?pwd=gcHifQ9OouKZnUY0EHLe8sjNXaV1oj.1>

Light refreshments will be served in Hill 452, 11:40am

Abstract:

Polygenic risk scores (PRS) are becoming increasingly vital for risk prediction and stratification in precision medicine. However, PRS model training presents significant challenges for broader adoption of PRS, including limited access to computational resources, difficulties in implementing advanced PRS methods, and availability and privacy concerns over individual-level genetic data. Cloud computing provides a promising solution with centralized computing and data resources. Here we introduce PennPRS (<https://pennprs.org>), a scalable cloud computing platform for online PRS model training in precision medicine. We developed novel pseudo-training algorithms for multiple PRS methods and ensemble approaches, enabling model training without requiring individual-level data. These methods were rigorously validated through extensive simulations and large-scale real data analyses involving over 6,000 phenotypes across various data sources. PennPRS supports online single- and multi-ancestry PRS training with seven methods, allowing users to upload their own data or query from more than 27,000 datasets in the GWAS Catalog, submit jobs, and download trained PRS models. Additionally, we applied our pseudo-training pipeline to train PRS models for over 8,000 phenotypes and made their PRS weights publicly accessible. In summary, PennPRS provides a novel cloud computing solution to improve the accessibility of PRS applications and reduce disparities in computational resources for the global PRS research community.

Bio: Dr. Jin is an Assistant Professor of Biostatistics at the University of Pennsylvania, Perelman School of Medicine. Her research focuses on statistical genetics/genomics and multi-modal data integration.

