Polygenic score construction with mean field variational inference to model SNP-SNP interactions

2023-10-26

Polygenic score

- Uses genome-wide SNP risk allele counts to produce a number - for each person - that summarizes genetic risk for a disease of interest, like coronary artery disease
- Typically rely on SNP effect estimates from genome-wide association studies
- Current uses include risk stratification for preventive interventions

Risk Stratification with Polygenic Scores (Lu et al. 2022)

Key question

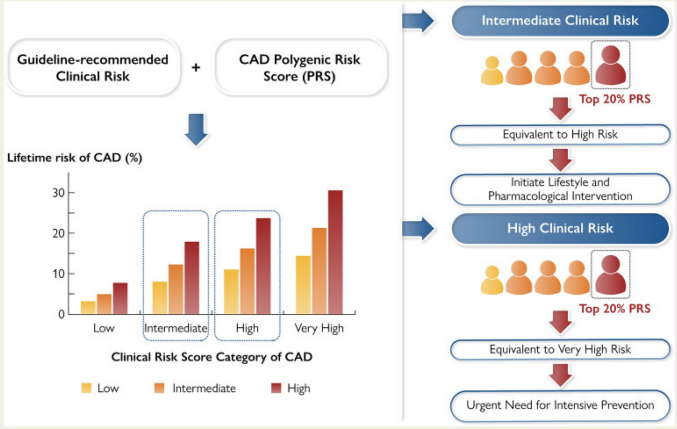
- The potential clinical utility of the polygenic risk score (PRS) for coronary artery disease (CAD), especially integrating with the current established clinical risk, was unknown among East Asian populations with significant disparities in both genetics and lifestyles.

Key finding

- The PRS comprising 540 genetic variants could stratify individuals into different trajectories of CAD risk, and further refine risk stratification for CAD within each clinical risk category.

Take-home message

The incorporation of polygenic risk into clinical care setting may provide a valuable risk stratification guidance to identify high-risk individuals for targeted intervention in primary prevention of CAD.



Genome-wide association studies

- Probe millions of markers, one at a time, across the genome to identify associations with disease
- Often use case-control study design
- Sample sizes in the hundreds of thousands
- Provides estimates of SNP effects

SNP-SNP interactions

- Existing polygenic score construction methods consider only SNP main effects and neglect SNP-SNP interactions

Current methods:

$$PGS_i = \sum_j g_{ij}\hat{\beta}_j$$

Proposed methods:

$$PGS_i = \sum_j g_{ij}\hat{\beta}_j + \sum_{j<k} g_{ij}g_{ik}\hat{\gamma}_{jk}$$

SNP-SNP Interactions

- We hypothesize that variational inference methods, and the associated reduction in computing requirements, will enable construction of polygenic risk scores that model SNP-SNP interactions
- We expect polygenic risk scores that model SNP-SNP interactions (and include SNP main effects) to outperform current methods in predictive accuracy for some traits, including those with high SNP-SNP heritability

Specific Aim 1

We will use mean field variational methods to provide analytic approximations to the posterior distribution for a Bayesian model with sparsity-inducing priors for polygenic risk scores

Rationale for Aim 1

Our variational inference-based strategy will diminish computing time and memory requirements while maintaining predictive ability of the sampling-based strategy of Privé, Arbel, and Vilhjálmsson (2020)

Outcomes for Aim 1

- A computationally scalable and efficient method for constructing polygenic risk scores
- Reproducible, open source implementation ensures transparency in our research and provides a valuable analytic tool to human genetics researchers

Specific Aim 2

We will develop a Bayesian statistical model for polygenic risk scores based on SNP effect estimates and estimates for SNP-SNP interaction effects

Rationale for Aim 2

- SNP-SNP interactions, in some diseases, explain a non-negligible proportion of the trait variance
- Therefore, polygenic risk scores that explicitly model SNP-SNP interactions for these diseases is expected to be more predictive of disease than scores from current models that ignore interactions

Outcomes for Aim 2

- A new class of Bayesian statistical models for polygenic risk scores using SNP main effects and SNP-SNP interactions
- Well designed, thoroughly tested, reproducible, open source software package that implements our methods

Thank You!

References

Lu, Xiangfeng, Zhongying Liu, Qingmei Cui, Fangchao Liu, Jianxin Li, Xiaoge Niu, Chong Shen, et al. 2022. "A Polygenic Risk Score Improves Risk Stratification of Coronary Artery Disease: A Large-Scale Prospective Chinese Cohort Study." *European Heart Journal* 43 (18): 1702–11.

Privé, Florian, Julyan Arbel, and Bjarni J Vilhjálmsson. 2020. "LDpred2: Better, Faster, Stronger." *Bioinformatics* 36 (22-23): 5424–31.