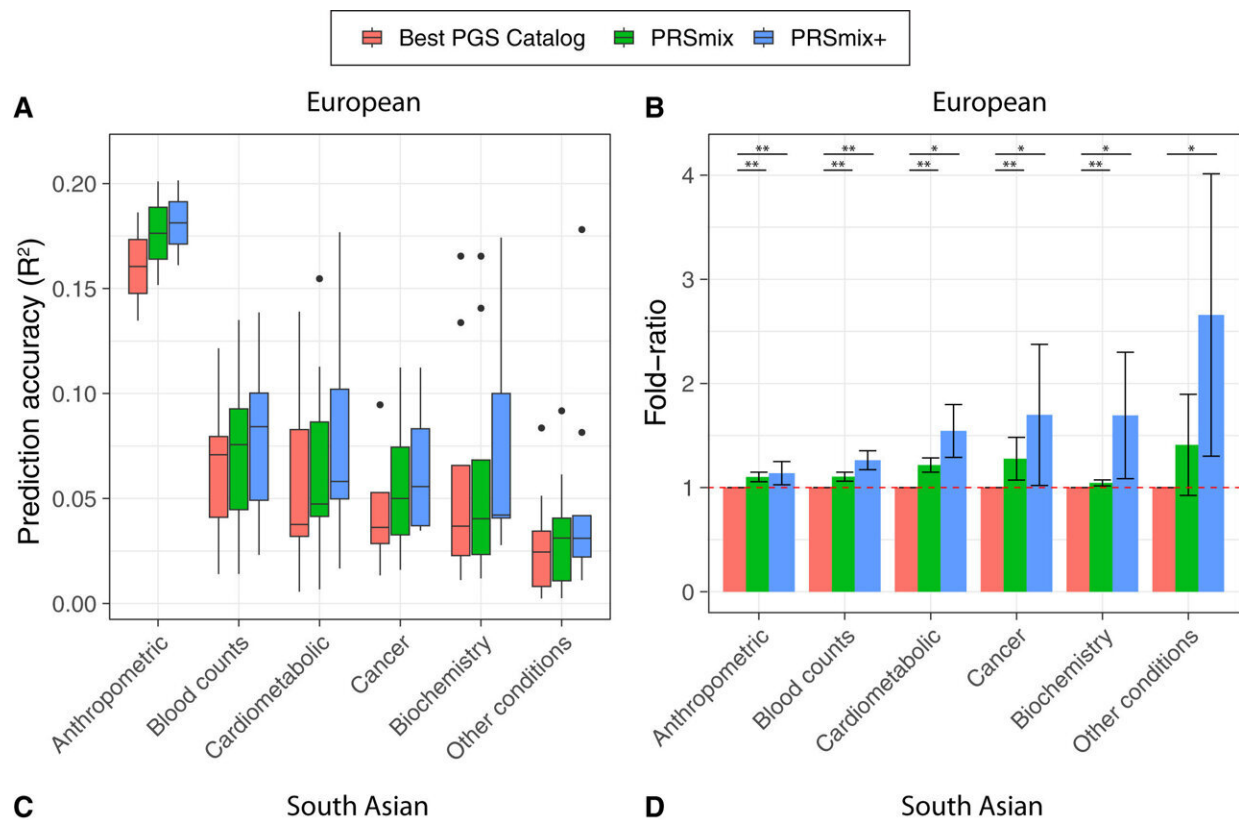# Researchers roll out a more accurate way to estimate genetic risks of disease

March 20 2024, by Claire Hendershot

Prediction accuracy and improvement across various types of traits in the European and South Asian ancestry. Credit: *Cell Genomics* (2024). DOI: 10.1016/j.xgen.2024.100523. https://www.cell.com/cell-genomics/fulltext/S2666-979X(24)00065-X

Researchers have developed statistical tools called polygenic risk scores

(PRSs) that can estimate individuals' risk for certain diseases with strong genetic components, such as heart disease or diabetes. However, the data on which PRSs are built is often limited in diversity and scope. As a result, PRSs are less accurate when applied to populations that differ demographically from the PRS training data.

A [new scoring approach](link) featured in *Cell Genomics* and developed by researchers at the Broad Institute of MIT and Harvard and Massachusetts General Hospital (MGH) uses a comprehensive approach to generate more accurate and informative PRSs.

Aptly named PRSmix due to its ability to "mix" all previously developed PRSs for a given trait, the approach generates scores that estimate a patient's genetic disease risk more accurately than PRSs generated from individual studies.

"A major challenge with PRSs is that they're derived in one population and then unleashed broadly with the assumption that the scores can be generalized," explained Pradeep Natarajan, the study's corresponding author. Natarajan is an associate member in Broad's Cardiovascular Disease Initiative and director of preventive cardiology at MGH. "The overall motivation for this work is to better identify individuals who are prematurely at high risk for heritable conditions."

When applied to an array of diseases, PRSmix was 20% more accurate on average at predicting risk for a given trait compared to individual PRSs. The improvements held true across two ancestral groups, suggesting that PRSmix scores may be more applicable across diverse populations.

While PRSmix offers marked improvements in genetic risk assessment, the tool's scope is limited only to genetic variants that are directly associated with a given trait.

"Most PRSs are trait-specific. For example, if you want to predict risk of coronary artery disease, most models use data trained only on coronary artery disease. But we know that the clinical risk is affected by more factors than just the genetic variants related to cardiovascular disease," said Buu Truong, the study's first author. Truong is a computational biologist at the Broad in Natarajan's lab and a Ph.D. student in the Price lab at the Harvard T. H. Chan School of Public Health. "If we aggregate more genetic information, how much more accurate can our PRSs be?"

To answer this question, the team developed an additional approach called PRSmix+ that aggregates all existing PRSs for a given trait plus all PRSs for related traits, such as heart disease and lipid levels.

By considering cross-trait influences, PRSmix+ showed even greater accuracy improvements than PRSmix. For example, PRSmix+'s estimates of risk of coronary artery disease represent a 3.27-fold increase in accuracy over previously developed combined methods.

PRSmix+ also accounts for pleiotropic variants, or single variants that influence multiple traits, and can help reveal otherwise unrecognized connections. For instance, age of menopause onset and ischemic risk stroke seem unrelated on the surface, but PRSmix+ identified a link between the two.

"We wouldn't have known this if we didn't do an unbiased scan," said Truong.

Despite PRSmix+'s predictive abilities, the team said there are still scenarios in which researchers should opt for PRSmix instead. PRSmix+ has a longer run time because it considers more genetic scores—and therefore must incorporate far more data—than PRSmix. Additionally, PRSmix+ only slightly outperformed PRSmix on predicting some highly

heritable traits like height. For researchers studying highly heritable traits, PRSmix could provide similar predictive accuracy in less time than PRSmix+. For researchers interested in clinical utility, PRSmix+ may be the better option.

An accessible algorithm

To ensure that PRSmix continually aggregates the most up-to-date genetic information, the PRSmix framework is publicly available on AnVIL, a data repository on biomedical data sharing platform Terra, and as an R package. "This workflow enables anyone to use the platform to generate PRSs for their own research," said Truong. "The user can just drop the file on Terra. They don't need to do any sophisticated computational work." AnVIL features checkpoints throughout the workflow to help users of all skill levels avoid computational errors.

So far, this accessible approach is proving successful. "Several people have already used the R package and reported that it generated a much better score than the scores from individual studies," said Truong.

PRSmix's accessible workflow combined with PRSmix+'s potential clinical utility makes the team's PRS framework more usable for researchers and clinicians alike.

"As we think about moving PRSs into the clinic, we need to anticipate that scores will perform differently across health care systems based on the different groups of individuals and contexts where the health care practices are located," said Natarajan. "We now have a method that provides a framework for training and recalibration, and that uses all currently available information to generate the best possible score within that dataset."

Provided by Broad Institute of MIT and Harvard