

New tool more accurately uses genomic data to predict disease risk across diverse populations

May 5 2022



Credit: CC0 Public Domain

Polygenic risk scores (PRS) are promising tools for predicting disease risk, but current versions have built-in bias that can affect their accuracy in some populations and result in health disparities. However, a team of researchers from Massachusetts General Hospital (MGH), the Broad

Institute of MIT and Harvard, and Shanghai Jiao Tong University in Shanghai, China, have designed a new method for generating PRS that more accurately predict disease risk across populations, which they report in *Nature Genetics*.

Alterations in a gene's DNA sequence can produce a genetic variant that increases the risk for disease. Some genetic variants are closely linked to certain diseases, such as the BRCA1 mutation and breast cancer.

"However, most common human diseases—such as type 2 diabetes, [high blood pressure](#), and depression, for example—are influenced not by single genes, but by hundreds or thousands of genetic variants across the genome. Each variant contributes a small effect," says Tian Ge, Ph.D., an applied mathematician and biostatistician in the Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine at MGH, and co-senior author of the paper. PRS aggregate the effects of genetic variants across the genome and have shown promise for one day being used to predict individual patients' chances of developing diseases. That would allow clinicians to recommend [preventive measures](#) and monitor patients closely for early diagnosis and intervention.

However, a PRS must be "trained" to predict disease risk using data from studies in which genomic information is collected from large groups of individuals. While many disease-causing variants are shared, explains Ge, there are important differences in the genetic basis of a disease between individuals of different ancestries. For example, a common genetic variant that is associated with a specific disease in one population may have a lower frequency or even be missing in other populations. When a genetic variant linked to a disease is shared across different populations, its effect size, or how much it increases risk, may also vary from one ancestral group to another, explains Ge. PRS trained using data from one [population](#) therefore often have attenuated, or reduced, performance when applied to other populations.

"A major problem with existing methods for PRS calculation is that, to date, most of the genomic studies used data collected from individuals of European ancestry," says Ge. That creates a Eurocentric bias in existing PRS, he says, producing substantially less-accurate predictions and raising the possibility that they could over- or underestimate disease risk in non-European populations.

Fortunately, investigators have increased efforts to collect genomic data from underrepresented populations. Leveraging these resources, Ge and his colleagues created a new tool called PRS-CSx that can integrate data from multiple populations and account for genetic similarities and differences between them. While there's still significantly more genomic data on individuals of European ancestry, the investigators used computational methods that allowed them to maximize the value of non-European data and improve prediction accuracy in ancestrally diverse individuals.

In the study, the investigators used genomic data from individuals in several different populations to predict a wide range of physical measures (such as height, body mass index, and [blood pressure](#)), blood biomarkers (such as glucose and cholesterol), and the risk for schizophrenia. Then they compared the predicted trait or [disease risk](#) with actual measures or reported disease status to measure PRS-CSx's prediction accuracy. The study's results demonstrated that PRS-CSx is significantly more accurate than existing PRS tools in non-European populations.

"The goal of our work was to narrow the gap between the prediction accuracy in underrepresented populations relative to European individuals, and narrow the gap in health disparities when implementing PRS in clinical settings," says Ge, who notes that the new tool will continue to be refined with the hope that clinicians may one day use it to inform treatment choices and make recommendations about patient care.

PRS-CSx could also have a role in basic research, says the study's lead author, Yunfeng Ruan, Ph.D., a postdoctoral research fellow at the Broad Institute of MIT and Harvard. It could be used, for example, to explore gene-environment interactions, such as how the effect of genetic risk would depend on the level of environmental risk factors in global populations.

Even with PRS-CSx, the gap in prediction accuracy between European and non-European populations remains considerable. Broadening the sample diversity across global populations is crucial to further improve the prediction accuracy of PRS in diverse populations. "The expansion of non-European genomic resources, coupled with advanced analytic methods like PRS-CSx, will accelerate the equitable deployment of PRS in [clinical settings](#)," says Hailiang Huang, Ph.D., a statistical geneticist in the Analytic and Translational Genetics Unit at MGH and the Stanley Center for Psychiatric Research at the Broad Institute, and co-senior author of the paper.

More information: Yunfeng Ruan et al, Improving polygenic prediction in ancestrally diverse populations, *Nature Genetics* (2022). [DOI: 10.1038/s41588-022-01054-7](https://doi.org/10.1038/s41588-022-01054-7)

Provided by Massachusetts General Hospital

Citation: New tool more accurately uses genomic data to predict disease risk across diverse populations (2022, May 5) retrieved 18 April 2024 from <https://medicalxpress.com/news/2022-05-tool-accurately-genomic-disease-diverse.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
