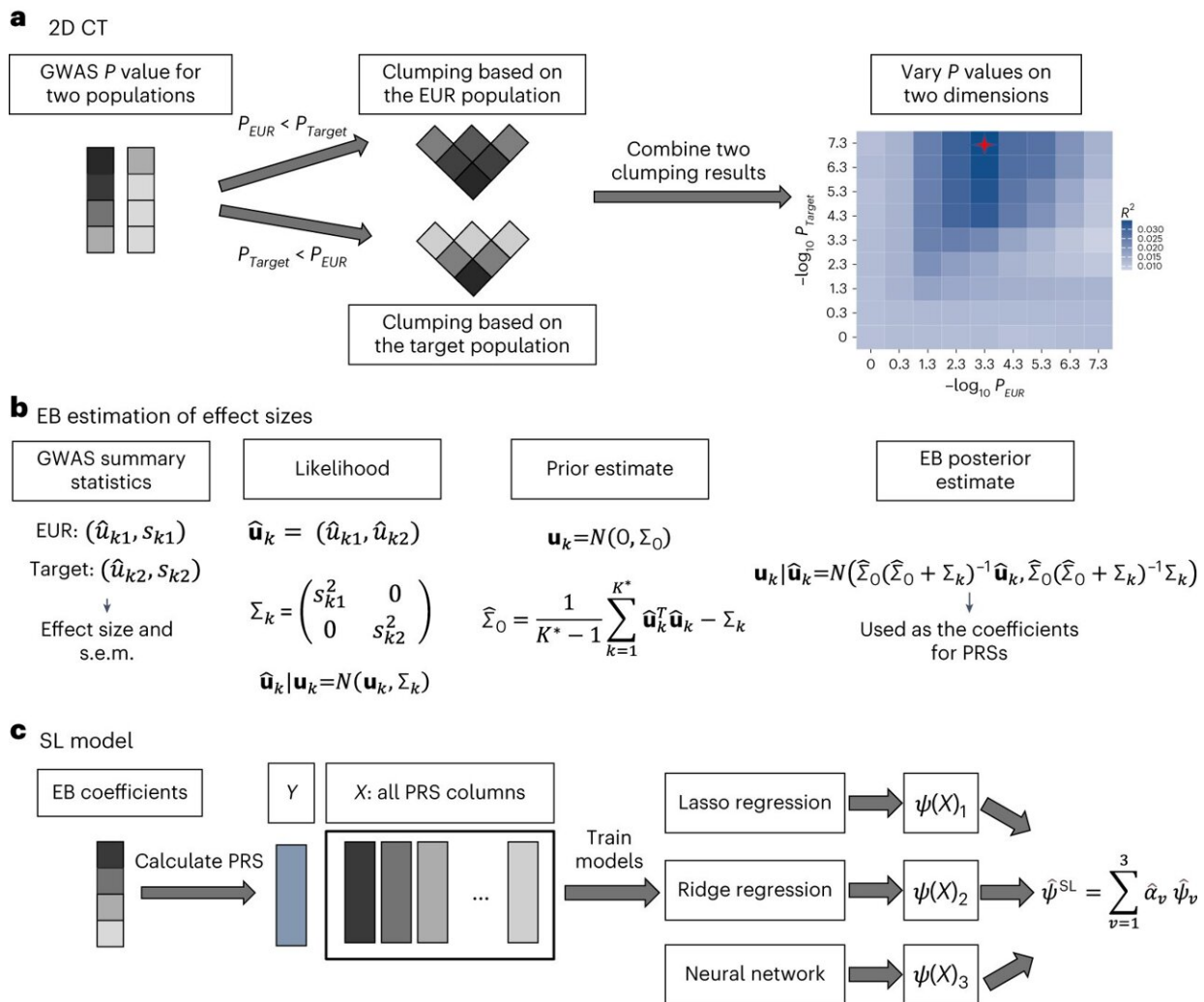


New method can improve assessing genetic risks for non-white populations

September 25 2023



CT-SLEB workflow. a–c, The method has three key steps: CT method for selecting SNPs (a); EB procedure for incorporating correlation in effect sizes of genetic variants across populations (b); and SL model for combining the PRSs derived from the first two steps under different tuning parameters (c). GWAS

summary statistics data were obtained from the training data. The tuning dataset was used to train the SL model. The final prediction performance was evaluated using an independent validation dataset. s.e.m., standard error of the mean.

Credit: *Nature Genetics* (2023). DOI: 10.1038/s41588-023-01501-z

A team led by researchers at Johns Hopkins Bloomberg School of Public Health and the National Cancer Institute has developed a new algorithm for genetic risk-scoring for major diseases across diverse ancestry populations that holds promise for reducing health care disparities.

Genetic risk-scoring algorithms are considered a promising method to identify high-risk groups of individuals who could benefit from preventive interventions for various diseases and conditions, such as cancers and heart diseases. These risk-scoring algorithms are based on large-scale genetic studies that link certain DNA variants to higher or lower disease risks.

The vast majority of subjects in these genetic studies have been people of European ancestry. The resulting risk-scoring algorithms have not always performed well in other populations, due to genetic differences across populations.

The new method, described in [a paper that appears online today](#) in *Nature Genetics*, has been applied to data from genetic studies from 23andMe Inc. and other sources involving more than 5 million individuals across diverse populations to generate genetic scores for 13 traits, including health conditions like coronary artery diseases and depression, in five different ancestry categories: European, African, Latino, East Asian, and South Asian. The researchers also tested the new method in large-scale simulation studies.

"We showed that our method can help close the risk-scoring performance gap for non-European-ancestry populations," says study senior author Nilanjan Chatterjee, Ph.D., Bloomberg Distinguished Professor in the Bloomberg School's Department of Biostatistics. "At the same time, we also concluded that we can't fully close the gap with new methods alone—we also need larger datasets on these populations."

Many risk-scoring models derived from [genetic studies](#) in non-European-ancestry populations often fall short because those studies typically are relatively small in scale. This results in a performance gap in risk-scoring between European-ancestry and other-ancestry populations, which may contribute to health care disparities.

The new method—which the researchers call CT-SLEB—used a combination of AI techniques including [machine learning](#) and Bayesian statistical modeling. In addition to the 23andMe database, the researchers "trained" CT-SLEB on data from the Global Lipids Genetics Consortium, the National Institutes of Health's All of Us research program, and UK Biobank.

The research team's benchmarking analyses showed that these new ancestry-specific risk-scoring models for the non-European populations generally outperformed standard polygenic risk score models that are based on mostly European-ancestry datasets, or are based on smaller non-European-ancestry datasets.

The researchers also compared CT-SLEB to a number of alternative methods. They found the proposed method is particularly helpful to improve genetic risk scores in African ancestry populations where scoring accuracy is generally the lowest. The team also found that CT-SLEB is computationally much faster compared to its closest competitors, and thus could be amenable to analyzing much larger numbers of DNA variants and more populations.

The team is now working with more advanced methods that are even better performing but are still computationally fast, Chatterjee says.

He also emphasizes that as the team's calculations in the study showed, having polygenic risk score models that work equally well in non-European-ancestry and European-ancestry populations will require more [genome-wide association studies](#) in non-European-ancestry populations.

"A lot of people think machine-learning and AI can do magic but without large, well-designed studies, algorithms will not be as useful," Chatterjee says.

The paper's lead author is Haoyu Zhang, Ph.D., who was a doctoral student at the Bloomberg School at the time the study began and is currently an investigator at the National Cancer Institute. Researchers from 23andMe contributed to development of the new method and the analysis of the data. The CT-SLEB code is publicly available via [GitHub](#). The code availability section in the paper includes a link to GitHub which includes the CT-SLEB code.

The paper, titled "A new method for multi-ancestry polygenic prediction improves performance across diverse populations," was co-authored by Haoyu Zhang, Jianan Zhan, Jin Jin, Jingning Zhang, Wenxuan Lu, Ruzhang Zhao, Thomas Ahearn, Zhi Yu, Jared O'Connell, Yunxuan Jiang, Tony Chen, Dayne Okuhara, 23andMe Research Team, Montserrat Garcia-Closas, Xihong Lin, Bertram Koelsch, and Nilanjan Chatterjee.

More information: Haoyu Zhang et al, A new method for multi-ancestry polygenic prediction improves performance across diverse populations, *Nature Genetics* (2023). [DOI: 10.1038/s41588-023-01501-z](https://doi.org/10.1038/s41588-023-01501-z)

Provided by Johns Hopkins University Bloomberg School of Public Health

Citation: New method can improve assessing genetic risks for non-white populations (2023, September 25) retrieved 2 May 2024 from <https://medicalxpress.com/news/2023-09-method-genetic-non-white-populations.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.