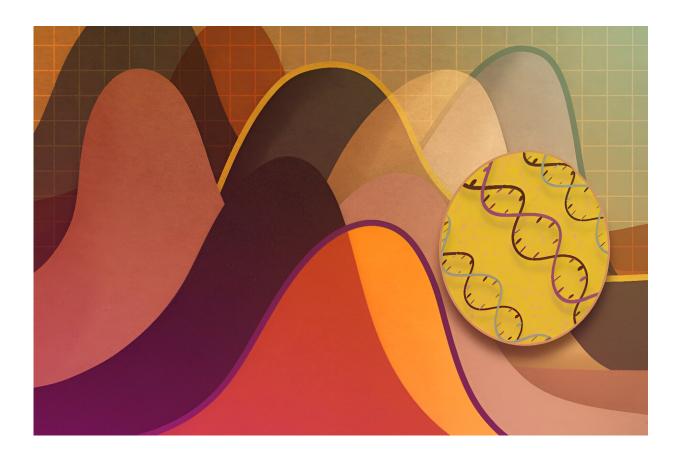# Medical Xpress

Making genetic prediction models more inclusive

October 26 2023, by Anne Trafton

In an effort to make polygenic risk scores more inclusive, MIT researchers have created a new model that takes into account genetic information from people of a wider variety of genetic backgrounds. Credit: Jose-Luis Olivares, MIT

While any two human genomes are about 99.9% identical, genetic

variation in the remaining 0.1% plays an important role in shaping human diversity, including a person's risk for developing certain diseases.

Measuring the cumulative effect of these small genetic differences can provide an estimate of an individual's genetic risk for a particular disease or their likelihood of having a particular trait. However, the majority of models used to generate these "polygenic scores" are based on studies done in people of European descent, and do not accurately gauge the risk for people of non-European ancestry or people whose genomes contain a mixture of chromosome regions inherited from previously isolated populations, also known as admixed ancestry.

In an effort to make these genetic scores more inclusive, MIT researchers have created a new model that takes into account genetic information from people from a wider diversity of genetic ancestries across the world. Using this model, they showed that they could increase the accuracy of genetics-based predictions for a variety of traits, especially for people from populations that have been traditionally underrepresented in genetic studies.

"For people of African ancestry, our model proved to be about 60% more accurate on average," says Manolis Kellis, a professor of computer science in MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) and a member of the Broad Institute of MIT and Harvard. "For people of admixed genetic backgrounds more broadly, who have been excluded from most previous models, the accuracy of our model increased by an average of about 18%."

The researchers hope their more inclusive modeling approach could help improve health outcomes for a wider range of people and promote health equity by spreading the benefits of genomic sequencing more widely across the globe.

"What we have done is created a method that allows you to be much more accurate for admixed and ancestry-diverse individuals, and ensure the results and the benefits of human genetics research are equally shared by everyone," says MIT postdoc Yosuke Tanigawa, the lead and co-corresponding author of the paper, which appears today in open-access form in the *American Journal of Human Genetics*. The researchers have made all of their data publicly available for the broader scientific community to use.

More inclusive models

The work builds on the Human Genome Project, which mapped all of the genes found in the human genome, and on subsequent large-scale, cohort-based studies of how genetic variants in the human genome are linked to disease risk and other differences between individuals.

These studies showed that the effect of any individual genetic variant on its own is typically very small. Together, these small effects add up and influence the risk of developing heart disease or diabetes, having a stroke, or being diagnosed with psychiatric disorders such as schizophrenia.

"We have hundreds of thousands of genetic variants that are associated with complex traits, each of which is individually playing a weak effect, but together they are beginning to be predictive for disease predispositions," Kellis says.

However, most of these genome-wide association studies included few people of non-European descent, so polygenic risk models based on them translate poorly to non-European populations. People from different geographic areas can have different patterns of genetic variation, shaped by stochastic drift, population history, and environmental factors—for example, in people of African descent,

genetic variants that protect against malaria are more common than in other populations.

Those variants also affect other traits involving the immune system, such as counts of neutrophils, a type of immune cell. That variation would not be well-captured in a model based on genetic analysis of people of European ancestry alone.

"If you are an individual of African descent, of Latin American descent, of Asian descent, then you are currently being left out by the system," Kellis says. "This inequity in the utilization of genetic information for predicting risk of patients can cause unnecessary burden, unnecessary deaths, and unnecessary lack of prevention, and that's where our work comes in."

Some researchers have begun trying to address these disparities by creating distinct models for people of European descent, of African descent, or of Asian descent. These emerging approaches assign individuals to distinct genetic ancestry groups, aggregate the data to create an association summary, and make genetic prediction models. However, these approaches still don't represent people of admixed genetic backgrounds well.

"Our approach builds on the previous work without requiring researchers to assign individuals or local genomic segments of individuals to predefined distinct genetic ancestry groups," Tanigawa says. "Instead, we develop a single model for everybody by directly working on individuals across the continuum of their genetic ancestries."

In creating their new model, the MIT team used computational and statistical techniques that enabled them to study each individual's unique genetic profile instead of grouping individuals by population. This methodological advancement allowed the researchers to include people

of admixed ancestry, who made up nearly 10% of the UK Biobank dataset used for this study and currently account for about one in seven newborns in the United States.

"Because we work at the individual level, there is no need for computing summary-level data for different populations," Kellis says. "Thus, we did not need to exclude individuals of admixed ancestry, increasing our power by including more individuals and representing contributions from all populations in our combined model."

Better predictions

To create their new model, the researchers used genetic data from more than 280,000 people, which was collected by UK Biobank, a large-scale biomedical database and research resource containing de-identified genetic, lifestyle, and health information from half a million U.K. participants.

Using another set of about 81,000 held-out individuals from the UK Biobank, the researchers evaluated their model across 60 traits, which included traits related to body size and shape, such as height and body mass index, as well as blood traits such as white blood cell count and red blood cell count, which also have a genetic basis.

The researchers found that compared to models trained only on European-ancestry individuals, their model's predictions are more accurate for all genetic ancestry groups. The most notable gain was for people of African ancestry, who showed 61% average improvements, even though they only made up about 1.5% of samples in UK Biobank. The researchers also saw improvements of 11% for people of South Asian descent and 5% for white British people. Predictions for people of admixed ancestry improved by about 18%.

"When you bring all the individuals together in the training set, everybody contributes to the training of the polygenic score modeling on equal footing," Tanigawa says. "Combined with increasingly more inclusive data collection efforts, our method can help leverage these efforts to improve predictive accuracy for all."

The MIT team hopes its approach can eventually be incorporated into tests of an individual's risk of a variety of diseases. Such tests could be combined with conventional risk factors and used to help doctors diagnose disease or to help people manage their risk for certain diseases before they develop.

"Our work highlights the power of diversity, equity, and inclusion efforts in the context of genomics research," Tanigawa says.

The researchers now hope to add even more data to their model, including data from the United States, and to apply it to additional traits that they didn't analyze in this study.

"This is just the start," Kellis says. "We can't wait to see more people join our effort to propel inclusive human genetics research."

More information: Yosuke Tanigawa et al, *American Journal of Human Genetics* (2023).

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Making genetic prediction models more inclusive (2023, October 26) retrieved 21 May

2024 from https://medicalxpress.com/news/2023-10-genetic-inclusive.html