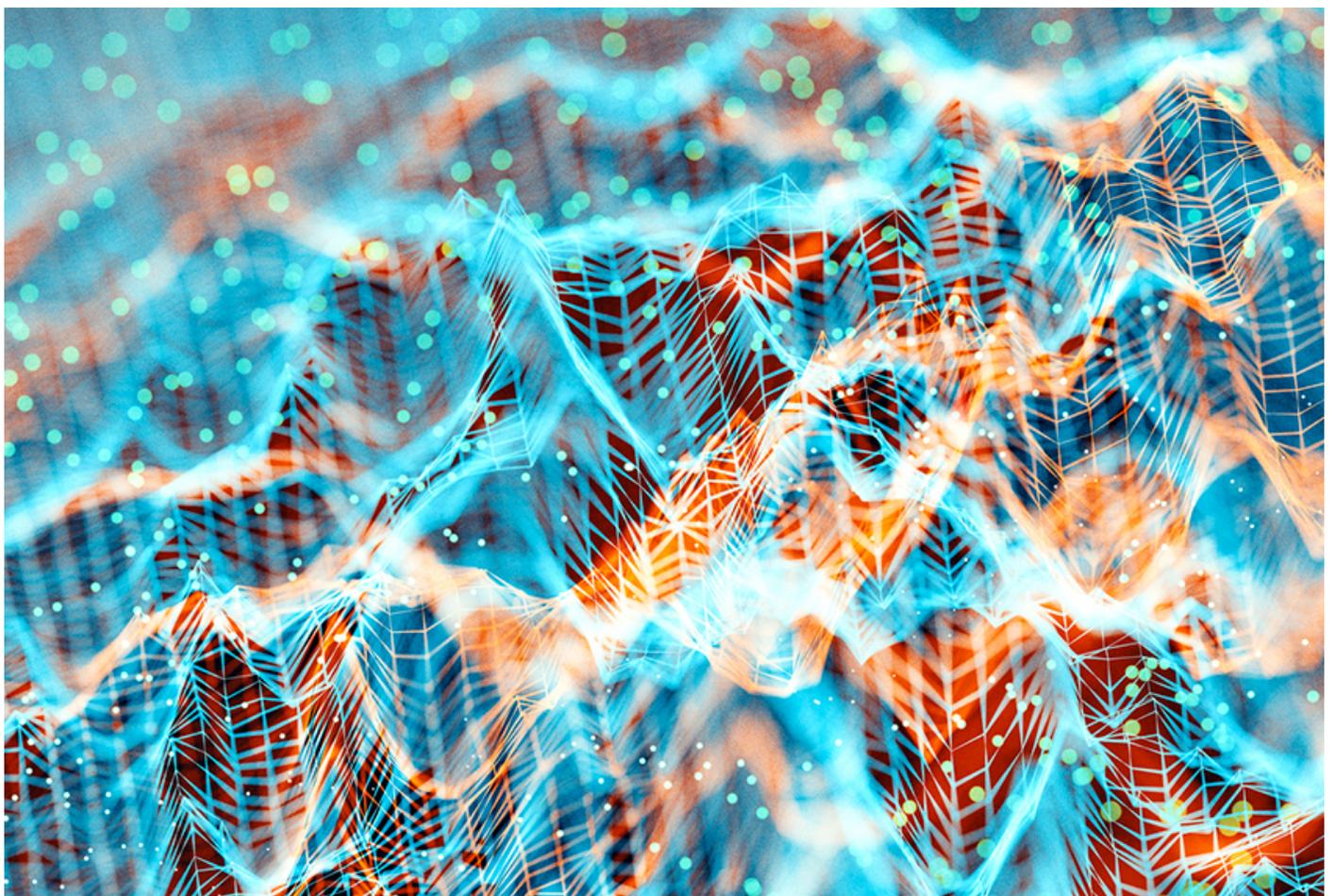


TOOLBOX

New approach aids search for genetic roots of complex conditions

BY CHLOE WILLIAMS

5 JULY 2019



A new method enables researchers to test algorithms for spotting genes that contribute to a complex trait or condition, such as autism¹.

Researchers often study the genetics of complex traits using **genome-wide association studies** (GWAS). In these studies, scientists compare the genomes of people who have a condition with those of people without the condition, looking for genetic variants likely to contribute to the condition. These studies often require tens of thousands of people to yield statistically significant results.

GWAS have identified more than 100 genomic regions associated with **schizophrenia**, for example, and **12 linked to autism**. Results are often difficult to interpret, however. Causal variants for a condition may be inherited with nearby sections of DNA that do not play a role.

To pick out the causal variants, scientists have developed algorithms that score variants based on their biological relevance to a condition. These algorithms may use information about a variant's shared biological pathways, common tissue expression levels or similar protein interaction networks.

Evaluating an algorithm's accuracy, however, is problematic. Most approaches draw comparisons with variants that have known causal links to a condition, which steers researchers back toward known genes. The approach also falters when there is little existing knowledge about a condition.

Divide and conquer:

The new method is unbiased because its results do not rely on making comparisons with known variants. Called Benchmarker, it is based on the machine-learning concept of 'cross validation' — training an algorithm on part of a dataset and then applying it to the rest.

In this case, researchers remove one chromosome at a time from a GWAS dataset and then apply an algorithm to the remaining chromosomes. The algorithm looks for patterns in genes or variants associated with a condition or trait. Once trained to spot those patterns, Benchmarker applies the algorithm to the removed chromosome and scores genes that fit the pattern by their biological relevance. It deems the 10 percent of genes or variants with the top scores from each chromosome as potentially causal.

The researchers compare the results for each algorithm by calculating how much its top variants contribute to a condition's heritability. The more a variant contributes to heritability, the more likely it is to be causal.

The researchers used Benchmarker on three algorithms and 20 GWAS datasets, including those related to height, schizophrenia, blood pressure and menopause.

Each algorithm highlighted slightly different genes, with some overlap. Two of the algorithms flagged 931 of the same genes, on average, and each algorithm also selected more than 750 other genes. When two or more algorithms flag the same genes, those genes are more likely to be

causal than genes identified by only one algorithm, the researchers reported in June in the *American Journal of Human Genetics*.

Benchmarker could be used to identify the strongest combination of algorithms for studying the genetics of a particular condition, the researchers say. It may soon be useful for autism, as **GWAS** are approaching a size that can produce robust results.

REFERENCES:

1. Fine R.S. *et al. Am. J. Hum. Genet.* **104**, 1025-1039 (2019) **PubMed**