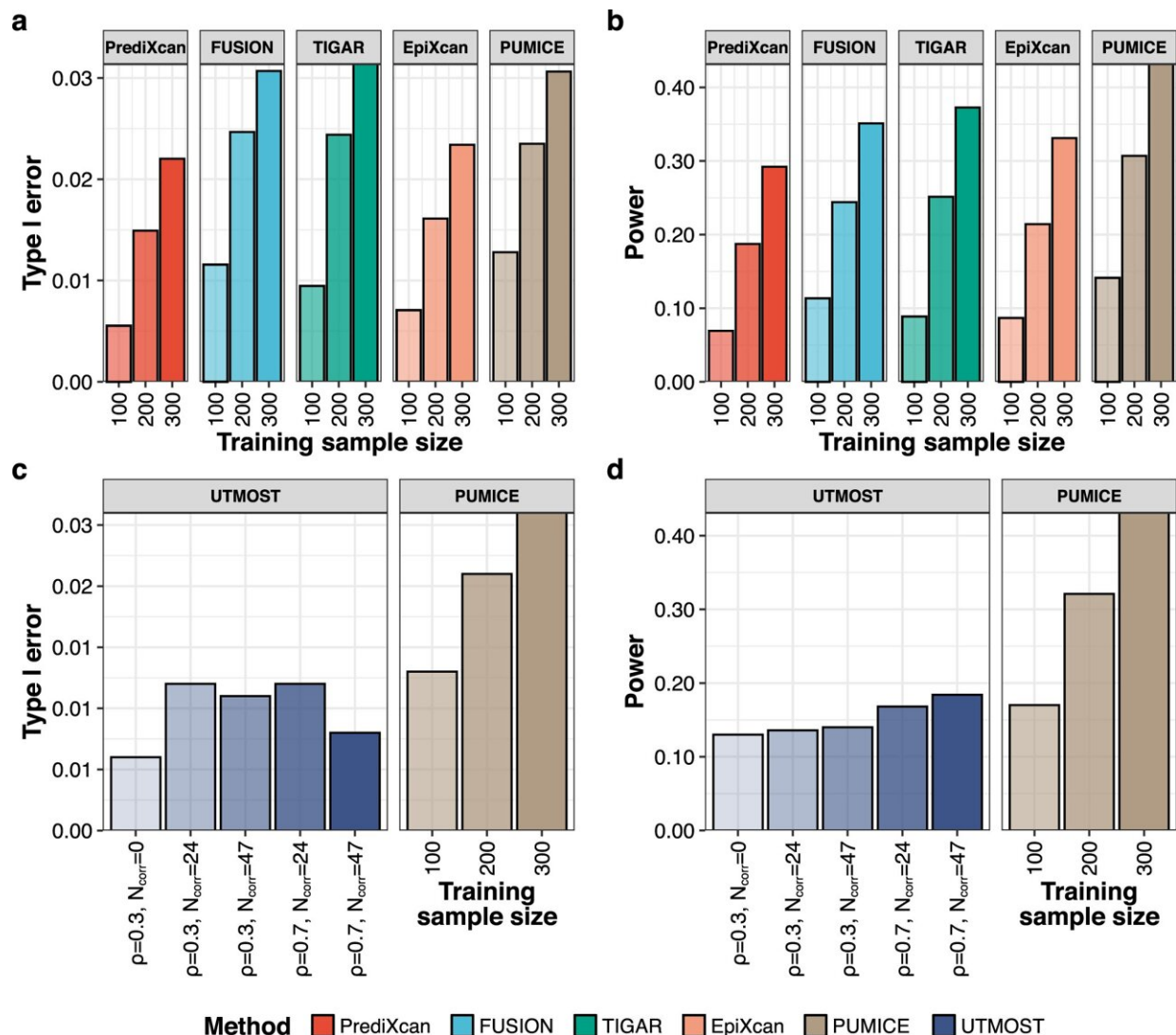# New machine learning technique shows how drugs can be repurposed

June 29 2022

Simulation studies comparing the performance of PUMICE to other TWAS methods. Panels (a, b) illustrates the comparison of PUMICE to other single-tissue TWAS methods for type I error (a) and power (b). Panels (c, d) illustrates

the comparison of PUMICE to multi-tissue TWAS method (UTMOST) for type I error (c) and power (d). For UTMOST, we evaluate its performance across different combinations of genetic correlation between causal and correlated tissues (ϱ) and number of correlated tissues (N_{corr}). Shadings represent different training sample sizes used to train gene expression prediction models for single-tissue TWAS methods and ϱ/N_{corr} combinations for multi-tissue TWAS method. Credit: *Nature Communications* (2022). DOI: 10.1038/s41467-022-30956-7

A new machine learning method to model gene expression levels might improve the identification of genes that cause human diseases, according to a new study by Penn State College of Medicine researchers. Through information from the three-dimensional (3D) structure of genomes and epigenetics—how genes and environment jointly influence diseases—the investigators were able to identify genes associated with complex traits and diseases. These identified disease genes also help to nominate drugs that may be repurposed to treat new disorders.

Developing and approving new prescription medications can be a costly and time-consuming process. However, findings from this study could partially change that moving forward. According to investigators, instead of developing new medicines, pharmaceutical companies could save time and money by repurposing drugs that have already been approved by the Food and Drug Administration to treat other disorders.

The human genome is composed of genetic instructions, or DNA that is fundamental to health and disease. In order to carry out these instructions, DNA must be read and expressed, and gene expression will be influenced by genetic variation. The same gene may be expressed higher (or lower) in people with certain mutations, which may cause diseases. Scientists analyze collections of gene readouts—or transcriptome—present in cells on hundreds of thousands of individuals. Transcriptome analyses can identify genes differentially expressed

between people with and without diseases, and thus lead to a new understanding of the genes associated with certain conditions.

For the new data method, PUMICE (Prediction Using Models Informed by Chromatin conformations and Epigenomics), Penn State researchers integrated transcriptomic, epigenomic and 3D genomic data using a novel machine learning approach. According to the study, PUMICE was successful at identifying drugs that could reverse the expression level of disease genes and may be repurposed to treat several human diseases.

"Traditional approaches that analyze one drug and one disease at a time can be very inefficient," said Dajiang Liu, co-senior author and associate professor of public health sciences and biochemistry and molecular biology at Penn State. "In contrast, a machine learning approach based on big data, such as PUMICE, can revolutionize biological and clinical research. It will greatly accelerate the process of identifying promising therapeutic targets, and fast forward drug development."

Using PUMICE, the researchers identified potential treatments for medical conditions, including COVID-19, Alzheimer's disease and autoimmune diseases, such as Crohn's disease, rheumatoid arthritis, ulcerative colitis and vitiligo, a skin pigmentation condition. They noted that some of the identified medications are already being evaluated in clinical trials, including Baracitinib, a drug for treating COVID-19.

"Being able to rediscover drugs that are already in clinical trials showcase the power of our approach," said Bibo Jiang, co-senior author and assistant professor of public health sciences at Penn State. "We will design follow-up experiments to validate new drugs and identify the most promising ones to further test in cell lines and animal models and eventually in clinical trials."

The study was published in *Nature Communications*.

Provided by Penn State College of Medicine