Novel YSPH framework helps identify genes associated with disease

15 March 2019

Hongyu Zhao, Ph.D. Credit: Yale University

A powerful analytical tool, known as UTMOST, developed by Hongyu Zhao, Ph.D., the Ira V. Hiscock Professor of Biostatistics at the Yale School of Public Health, and colleagues could allow researchers to design therapeutic drugs that more effectively combat disease.

Zhao explained the framework's purpose and potential in identifying genetic associations to disease in a recent interview.

The study is published in *Nature Genetics* .

What challenges currently exist in identifying genetic associations to disease and complex biological traits?

HZ: Understanding the genetic architecture underlying complex human traits and diseases (e.g. Alzheimer's disease, obesity, cancer and others) is important. Although much progress has been made to date through genome-wide association studies, a lot more can be learned by integrating different data sets together to extract more information. Our framework, UTMOST (Unified Test for Molecular SignaTures), is designed to leverage the rich information from major National Institutes of Health initiatives, such as the GTEx Project, to better identify genes associated with many complex diseases.

So, what does the UTMOST platform do?

HZ: UTMOST is a comprehensive and powerful framework capable of performing gene-level association analysis that leverages gene expression information across many human tissues simultaneously. Instead of directly studying associations between phenotypes and genetic variations across a population sample, disease-associated genes are inferred through evaluating correlation between disease status and imputed (predicted) gene expression levels from genetic variations either in single tissues or across a set of tissues.

The most critical step in this approach is the imputation accuracy. Compared with the single-tissue methods commonly used today, our approach achieved an average of 39% improvement in imputation accuracy and generated effective imputation models for an average of 120% more genes. The improved imputation accuracy for a larger number of genes will enable researchers to identify more disease-associated genes in the future.

What are some of the ways that this new tool can be applied in the field of public health?

HZ: With our UTMOST framework, we identified 68 significant genes that may be associated with late-onset Alzheimer's disease. We were also able to gain greater insight and understanding into the biological mechanisms associated with different gene expressions. This information not only helps us better understand the underlying causes of diseases like Alzheimer's and certain biological traits, it may also lead to the development of novel therapeutics, especially as more and more data are generated in regard to individual human genomes, cell types and molecular phenotypes.

Are there any limitations to UTMOST?

HZ: The statistical and computational methods for genomic studies are constantly evolving driven by ever increasing data types and volumes. Despite the many improvements UTMOST provides over existing methods, researchers need to be cautious

in interpreting findings from UTMOST analysis. Most importantly, gene-level associations identified in UTMOST do not imply causality.

More information: undefined undefined et al. A statistical framework for cross-tissue transcriptome-wide association analysis, *Nature Genetics* (2019). DOI: 10.1038/s41588-019-0345-7

Provided by Yale University