HORNET: Software and methods to perform whole-genome searches for causal gene networks

Noah Lorincz-Comi , Yihe Yang, Xiaofeng Zhu
Department of Population & Quantitative Health Sciences, Case Western Reserve University

The extent to which genes in disease-associated loci actually cause disease risk is mostly unknown. Many non-experimental approaches to investigate causality combine summary statistics from disease and gene expression (eQTL) GWAS. However, these approaches may produce incorrect inferences because of confounding by other genes and misspecified LD structure. We introduce the HORNET software to perform robust multivariable Mendelian Randomization (MR) with variable selection genome-wide using eQTL and phenotype GWAS summary statistics.

HORNET estimates direct causal effects of gene expression, constructs gene regulatory networks, performs Bayesian LD estimation, imputes missing data, and estimates local disease heritability. We applied HORNET to the 500Kb region surrounding the *APOE* gene and Alzheimer's disease (AD) using eQTL GWAS from lung tissue (n=515) and AD GWAS data (n=455k).

Simulations demonstrate that HORNET can provide valid causal inference across a range of real-world conditions in which other methods cannot. The software itself is computationally fast, spending approximately 1 second for every gene tested. When applied to AD, 5 genes in a 1Mb window around APOE explained 87.6% of the local AD heritability. APOE, APOC2/4, and DMPK all had direct causal effects on AD (P-values<5E-5), whereas PPM1N only caused AD by regulating APOE (P-value=9.4E-8) and APOC4 (P-value<1E-10).

The HORNET software provides researchers with an accessible and robust tool for identifying genes with causal evidence. The software is publicly available and may identify promising candidate genes for follow-up experimental testing.