



May 6, 2022

Untangling our genes

Updated: Jan 12, 2024



Did you know **about Genome Wide Association Studies, or GWAS?** It aims to identify genetic causes of specific diseases or other outcomes. The good news is that understanding how a disease develops at a molecular level allows us to better design drugs or other treatments.

GWAS is done by sequencing

DNA of individuals, some with disease and some without, and analyzing this sequencing data to see which parts are found more often in persons with disease – known as a **positive association**. The genes occurring in these DNA sequences are identified and undergo functional analysis to determine how they may be causing the disease we see.

It is vital to look across the entire genome of the studies because most diseases are caused by several interacting genes. In addition, most diseases are not solely caused by genetics but may be triggered by environmental or lifestyle factors during their development.

To deal with this complex system, GWAS has been getting larger, with studies including over 1 million individuals now published, which means more data that increases the difficulties and computational burden of analysis. Researchers are now turning to Machine Learning (ML) to provide better strategies for dealing with very large data sets generated.

One **key problem** with GWAS is figuring out which associations are causative and which ones are not. This may be solved with supervised ML methods, where statistical models are applied to the list of associations to accurately classify them based on if they are causal or not.

There are several other ways ML can be used to improve GWAS data analysis, like multi-loci feature selection to find novel groups of genes working together, neural networks or deep

learning for incomplete data, or permutation testing to simplify analysis by removing non-significant data.

In addition, genes found by existing ML methods are being used in clinics. For example, a drug targeting Uromodulin is tested in patients to treat hypertension after this gene was identified as relevant by the OPEN unbiased ML GWAS approach.

So, ML is fast becoming a standard part of the GWAS analysis pipeline, a trend which is expected to continue.

For more info on using Machine Learning in Biology, [take a look at our video](#).

Written: Shelley Edmunds

Updated: Marie Sorivelle