

Analytical methods to identify genes for complex traits in Genome-Wide Association Studies (GWAS)

Fabio Macciardi

Dept of Science and Biomedical Technologies, University of Milano.
fabio.macciardi@unimi.it

Abstract

Current methods for GWA studies look for the association of simple DNA variants (eg, SNPs) with a complex trait of interest reducing the complexity of the approach to “n” simple univariate tests, with “n” equal to the total number of DNA variants under scrutiny. In this case, analyzing the genetic bases of, say, schizophrenia in a case-control study with a 1M SNPs array, resolve into calculating 1M chi-square tests, with or without a proper correction for multiple testing.

While this approach makes a GWA analysis easy to perform, we lose the “big picture” that can be captured with analyses that are closer to the underlying biological reality. In fact, a thorough analysis must include a detailed evaluation of the genomic context for the SNPs found significantly associated (the LD structure and the SNP-to-gene relationship), of the gene/block-based haplotypes and of the multivariate SNP-SNP (as proxies for gene-gene) interactions, in addition to extended annotation for the associated SNPs. Jointly with SNP association, analysis of Copy Number Variants (CNVs) and Polymorphisms (CNP) is warranted. A second step of the analysis should try to (re)build the pathway(s) and/or the network of genes found associated, thus moving toward the generation of functional hypotheses. The careful consideration of the phenotypic complexity (e.g., the clinical heterogeneity) of the trait and the need to replicate the findings in a different and independent sample, possibly via a joint analysis, would reasonably give more grounds to any finding.

Such an integrated method not only needs advanced multivariate statistical techniques, but also requires sound bioinformatics approaches from computational to graph analyses.