What is TWAS and how do we use it in integrating gene expression data

Quan Long
University of Calgary

Abstract

The transcriptome-wide association studies (TWAS) is a pioneering approach utilizing gene expression data to identify genetic basis of complex diseases. Its core component is called "genetically regulated expression (GReX)". GReX links gene expression information with phenotype by serving as both the outcome of genotype-based expression models and the predictor for downstream association testing. Although it is popular and has been used in many high-profile projects, its mathematical nature and interpretation haven't been rigorously verified. As such, we have first conducted power analysis using NCP-based closed forms (Cao et al, PLoS Genet 2021), based on which we realized that the common interpretation of TWAS that looks biologically sensible is actually mathematically questionable. Following this insight, by real data analysis and simulations, we demonstrated that current linear models of GReX inadvertently combine two separable steps of machine learning - feature selection and aggregation - which can be independently replaced to improve overall power (Cao et al, Genetics 2021). Based on this new interpretation, we have developed novel protocols disentangling feature selections and aggregations, leading to improved power and novel biological discoveries (Cao et al, BiB 2021; Genetics 2021). To promote this new understanding, we moved forward to develop two statistical tools utilizing gene expressions in identifying genetic basis of gene-gene interactions (Kossinna et al, in preparation) and low-effect genetic variants (Bian et al, in preparation), respectively. Looking forward, our mathematical characterization of TWAS opens a door for a new way to integrate gene expressions in genetic studies towards the realization of precision medicine.