Identifying effector genes of human GWAS variants by INFIMA

Genome-wide association studies (GWAS) have revealed many non-coding single nucleotide variants that are statistically associated with complex traits and diseases. However, the effector genes through which these disease risk variants mediate their effects are largely unknown. Surprisingly, model organism studies have also largely remained as an untapped resource for unveiling such effector genes. A recent well-powered expression quantitative locus (eQTL) study in islets from Diversity Outbred (DO) mice identified thousands of eQTLs; however, it lacked the resolution to pinpoint causal single nucleotide variants and the regulatory mechanisms responsible for the wide range in susceptibility to diabetes due to high linkage disequilibrium. To address this bottleneck and leverage eQTLs derived from DO mice for elucidating effector genes of human GWAS variants, we developed a statistical data integration model, INFIMA for Integrative Fine-Mapping with Model Organism Multi-Omics Data. INFIMA capitalizes on multi-omics data modalities such as chromatin accessibility and transcriptome from the eight DO mice founder strains to fine-map DO islet eQTLs. In addition, INFIMA employs footprinting and in silico mutation analysis to reveal regulatory genetic variants that mediate strain-specific expression differences. We applied INFIMA to identify novel effector genes in pancreatic islets for human GWAS variants associated with diabetes. We computationally validated INFIMA predictions with high-resolution chromatin capture data sets from mouse and human islets. Our results demonstrate that INFIMA is a powerful method for leveraging model organism multi-omics data to identify candidate effector genes of non-coding human GWAS variants and performs better than baseline alternatives.