Genetic mapping of cell type specificity for complex traits

Key points

○ Single-cell RNA sequencing data was used to create specific transcriptome profiles and aligned with genome-wide association studies in order to implicate cell type specificity in various traits.

○ Current methods struggle with limited datasets and integration difficulties due to complex batch effects due to variety of sources and protocols.

○ The authors made a workflow that systematically evaluates and analyzes within and between datasets, circumventing batch effects and uncovering associations of traits with specific cell types.

Background

○ **Genome-wide association studies (GWAS)**: an observational study of a genome-wide set of genetic variants in various individuals to see if any variant is associated with a trait.

  • GWAS studies on human data take a phenotype-first approach by comparing varying phenotypes for a trait or disease.

  • GWAS have yielded genetic risk variants for a wide degenerative, cardiovascular, metabolic disease, and even quantitative traits such as IQ, educational attainment, and height.

  • High polygenicity of traits continue to pose challenges for GWAS and biological insight in general.

  • The phenotype-first contrasts with genetic epidemiological studies, which use genotype-first, wherein individuals are first grouped by a statically common genotypes gathered from prior clinical phenotypic classifications.

  • GWAS typically focuses on associations between single-nucleotide polymorphisms (SNPs), but this study uses in combination with single-cell RNA sequencing.

○ **Single cell RNA sequencing (scRNA-seq)**: a subset of single-cell sequencing that uses microarrrays and RNA-sequences to analyze expression of RNA from large populations or cells.

  • Provides expression profiles of individual cells and is currently the "gold standard" for defining cell states and phenotypes.

  • Uses clustering analysis (data mining technique of grouping like objects) to uncover the existence of rare cell types within a cell population that may have never been seen before.

- Traditional single cell sequencing also use bulk RNA-sequences, but scRNA-seq can characterize transcriptome profiles (TMPs) resulting in much higher resolution.

- Recent studies linking cell-specific expression profiles with GWAS have already implicated cell types involved in schizophrenia, IQ, and neuroticism.

  - However, these use limited datasets, confined cell types from single tissue, and replication difficulties.

Methods

- This study addresses the currently limited use of publicly available scRNA datasets and purpose a workflow for other multiple scRNA-seq resources to use.

- 43 publicly available datasets were curated from 32 studies across a variety of tissues and organs from human and mouse samples.

- A large barrier to overcome is the difficulties (batch-effects) involved in generated unified scRNA-seq datasets from the variety of sources and differences in protocols.

  - Often there is a requirement of highly similar cell types that are not available.

  - Unique information can easily be lost that may be needed to reach higher resolutions.

- A publicly available workflow was made available via a web application (http://fuma.ctglab.nl) made by the authors that systematically evaluates association as a means to help combat batch-effect problems.

  - The authors illustrated the use of the workflow by applying it to 26 traits covering 6 domains in which GWAS were available.

Results/Discussion

- The results of this study honestly is not in the scope of my current understanding. This paper sides more on data science than genetics. Overall, this paper was more about describing the workflow they created for others to use.

- Overall, their comparison method of cell type specific expression across datasets showed that same cell types cluster together across such datasets despite difference in taxonomy or tissue types in samples.

- The application of their method did show a success across 26 traits using 43 scRNA-seq datasets, demonstrating that multi-step conditional analyses can disentangle relationships between associated cell types, revealing independent signals of specific cell types.

References

Adhikari, A. N., Gallagher, R. C., Wang, Y., Currier, R. J., Amatuni, G., Bassaganyas, L., … others (2020). The role of exome sequencing in newborn screening for inborn errors of metabolism. *Nature Medicine*, *26*(9), 1392–1397.