St. Jude Children's Research Hospital

St. Jude BioHackathon

Title
Determine the effects of imputation on ssGSEA/GSVA methods in scRNA-seq

Category
Processing Pipelines And Methods

Challenge
Single-sample GSEA (ssGSEA), an extension of Gene Set Enrichment Analysis (GSEA) originally developed for bulk gene expression analysis, has been commonly used to score single cell or cluster of cells. The algorithms such as ssGSEA and GSVA rely on statistics that are inappropriate for single cell transcriptomes mostly composed of zero values, and where the missing genes of each cell vary extensively across the entire scRNA-seq dataset. Thus the derived scores may not be reliable. The excess zero values include those of genes not expressed in the cell, and those produced due to dropout events. Whether imputing zeros due to dropout events can improve the performance of single-cell GSEA remains to be an open question and one well worth answering.

Benefit
The single cell GSEA scores can be used to cluster cells or assign cell types such as crispr-edited cells vs non-edited cells. More robust scores that are less affected by the data scarcity and cell to cell variability inherent to some single cell RNA-seq platforms would increase confidence in higher-level interpretation of transcriptonal changes between popluations.

Helpful Tools, Packages, or Software
impupte: https://pubmed.ncbi.nlm.nih.gov/35017482/, ssGSEA: GSVA or escape R package

Test Data
Data for 4k and 8k PBMC obtained by 10× Genomics 3' chemistry V2: 10× Genomics website (https://support.10xgenomics. com/single-cell-gene-expression/datasets/2.1.0/pb-mc8k, https://support.10xgenomics.com/single-cell-geneexpression/ datasets/2.1.0/pbmc4k).