SEQUIN: interactive web app for rapid, reproducible bulk and single cell RNA-Seq analysis

C. Weber[1,2], M. B. Hirst[3], B. Ernest[3], N. J. Schaub[2], K. M. Wilson[2], K. Wang[2], H. M. Baskir[2], P-H. Chu[2], C. Tristan[2], I. Singec[2]
[1]National Eye Institute, Rockville, MD
[2]National Center for Advancing Translational Sciences, Rockville, MD
[3]Rancho BioSciences, San Diego, CA

Highlights

- **SEQUIN: an R-Shiny app to analyze bulk *and* single-cell RNA-Seq data interactively in a browser**
- **No** prior bioinformatics expertise required
- **Data visualization** and generation of publication-ready figures and analysis result tables
- **iPSC Profiler tool:** measure and compare pluripotent and differentiated cell types against reference profiles
- Advanced sample and cell **clustering** options for differential expression tests and manual clusters (subset, merge, create)
- **Gene set enrichment** analysis via EnrichR (GO, KEGG, etc.)

Introduction

SEQUIN is a web-based R/Shiny application that allows fast and intuitive analysis of RNA-sequencing data derived for model organisms, tissues, and single cells. Integrated app functions enable uploading datasets, quality control, differential gene expression analysis (DGE), gene set enrichment (GSE), and data visualization. We developed the **iPSC Profiler**, a practical gene module scoring tool that helps to measure and compare pluripotent and differentiated cell types. Freely available to the public, SEQUIN empowers scientists using interdisciplinary methods to investigate and present transcriptome data firsthand with state-of-the-art statistical methods. SEQUIN helps democratize and increase the throughput of analysis using next-generation sequencing data with single-cell resolution.

Methods

Here, we analyzed a new pair of unpublished datasets from human iPSCs of cell lines cultured with differing media, in order to determine optimal conditions for growth and self-renewal: E8, E8+Albumin, and mTeSR. Cell lines were: LiPSC GR1.1, GM25256, NCRM5, and WA09. Bulk RNA-Seq had 27 samples in triple technical replicates, while the scRNA had a total of 24,186 cells. We performed pairwise DGE and PCA-based clustering (and UMAP for scRNA), followed by GSE.

Results

In both datasets we found samples were most tightly clustered by growing media first, then by cell line. In PCA plots, E8 and mTeSR media were more similar by overall expression than E8 and Albumin. After performing DGE with media as contrast (E8 vs. E8-Albumin) and evaluating iPSC Profiler gene module scores, pluripotency was equally high across media types and cell lines. By the intersection of significant DE genes in bulk and scRNA-Seq, for the same media comparison, there were 90 genes (adj. p-value < 0.001, abs. value log2 fold change ≥ 0.5). We performed GSE with this gene set in SEQUIN and found enriched Gene Ontology terms relating to prostaglandin biosynthesis and cell membrane raft organization. These terms included ANXA (annexin) gene family which relates to phospholipid binding. Since the media differ by lipid concentration, this may explain upregulation of ANXA. These experiments exemplify the use of SEQUIN for rapid and reproducible analysis leading to actionable insights.