Das, S., Rai, A., Merchant, M. L., Cave, M. C., & Rai, S. N. (2021). A comprehensive survey of statistical approaches for differential expression analysis in single-cell RNA sequencing studies. *Genes*, *12*(12), 1947. https://doi.org/10.3390/genes12121947

Definitions

- **Single-cell RNA Sequencing (scRNA-seq):** A technology that allows scientists to examine the gene expression of individual cells.
- **Differential Expression (DE):** Identifying genes that show different levels of expression between different conditions or cell types.
- **Bulk RNA Sequencing:** A method that measures the average gene expression across many cells, without distinguishing individual cells.
- **Negative Binomial (NB) Model:** A statistical model used to handle overdispersed count data, common in RNA sequencing.
- **Zero-Inflated Models (ZIM):** Models that account for excess zeros in data, useful in scRNA-seq due to many genes not being expressed in certain cells.

Key Findings

- Several DE analysis methods originally designed for bulk RNA-seq can be adapted for scRNA-seq, but specialized methods for scRNA-seq perform better.
- The DECENT and EBSeq methods are identified as top performers for DE analysis in scRNA-seq.
- Performance varies significantly depending on the dataset and specific criteria used for evaluation.

Introduction

The paper reviews statistical methods used for analyzing single-cell RNA sequencing (scRNA-seq) data, focusing on differential expression (DE) analysis. This analysis is crucial for identifying gene markers specific to different cell types, which can lead to better understanding of cellular functions and disease mechanisms.

Main Content

Background

Single-cell RNA sequencing (scRNA-seq) is a powerful technique that allows scientists to study gene expression at the level of individual cells. Unlike bulk RNA sequencing, which averages gene expression across many cells, scRNA-seq captures the heterogeneity between cells. DE analysis helps identify genes

that are expressed differently between conditions or cell types. However, the presence of noise and technical variability in scRNA-seq data presents challenges for DE analysis.

Objectives

The study aims to provide a comprehensive review of DE analysis methods for scRNA-seq, evaluating their performance using multiple criteria and real datasets. The goal is to guide researchers in selecting appropriate methods for their specific experimental settings.

Methods

- **Data Collection:** Eleven scRNA-seq datasets from different studies were used to evaluate 19 DE analysis methods.
- **Performance Metrics:** Methods were assessed using metrics like true positive rate (TPR), false positive rate (FPR), accuracy, and area under the receiver operating characteristic curve (AUROC).
- **Comparison:** Both bulk RNA-seq methods adapted for single-cell data and methods specifically designed for scRNA-seq were included in the comparison.

Results

- **Method Performance:** The study found that some bulk RNA-seq methods, such as edgeR and DESeq2, perform competitively with single-cell specific methods under certain conditions.
- **Top Performers:** DECENT and EBSeq emerged as the best methods for DE analysis in scRNA-seq data based on multiple performance criteria.
- **Data Specificity:** The performance of DE methods varied significantly depending on the dataset characteristics and evaluation criteria.

Conclusion

The study provides a detailed evaluation of DE analysis methods for single-cell RNA sequencing. It highlights the strengths and limitations of various methods and offers practical guidelines for selecting the most appropriate tool based on specific experimental settings. The findings underscore the importance of choosing the right statistical approach to accurately interpret scRNA-seq data and advance our understanding of cellular heterogeneity and disease mechanisms.

Word Count: 491