SPL – III Project Proposal

Single-Cell RNA-seq Analysis

Course: Software Project Lab – III

Course No: SE – 801

Submitted by

Ishrat Jahan Emu

BSSE 0927

Session: 2016–17

Supervised by

Sumon Ahmed

Assistant Professor

1. Broad Domain

Single-cell RNA-seq has enabled gene expression to be studied at an unprecedented resolution. The promise of this technology is attracting a growing user base for single-cell analysis methods. As more analysis tools are becoming available, it is becoming increasingly difficult to navigate this landscape and produce an up-to-date workflow to analyse one's data. Here, we detail the steps of a typical single-cell RNA-seq analysis, including pre-processing (quality control, normalization, data correction, feature selection, and dimensionality reduction) and cell- and gene-level downstream analysis.

2. Problem Statement:

Due to technical limitations and biological factors, scRNA-seq data are noisier and more complex than bulk RNA-seq data. The high variability of scRNA-seq data raises computational challenges in data analysis.
Generating single-cell data from a biological sample requires multiple steps. Typical workflows incorporate single-cell dissociation, single-cell isolation, library construction, and sequencing.
Make a single cell RNA-seq analysis tool which is faster and efficient for big dataset.

3. Project Overview:

Recent developments in single-cell RNA sequencing (scRNA-seq) platforms have vastly increased the number of cells typically assayed in an experiment. Analysis of scRNA-seq data is multidisciplinary in nature, requiring careful consideration of the application of statistical methods with respect to the underlying biology.
Few analysis packages exist that are at once robust, are computationally fast, and allow flexible integration with other bioinformatics tools and methods.
Single-cell RNA sequencing (scRNA-seq) enables transcriptome-wide gene expression measurement at single-cell resolution, allowing for cell type clusters to be distinguished (for an early example, see [2]), the arrangement of populations of cells according to novel hierarchies, and the identification of cells transitioning between states. This can lead to a much clearer view of the dynamics of tissue and organism development, and on structures within cell populations that had so far been perceived as homogeneous. In a similar vein, analyses based on single-cell DNA sequencing (scDNA-seq) can highlight somatic clonal structures (e.g., in cancer, see [3, 4]), thus helping to track the formation of cell lineages and provide insight into evolutionary processes acting on somatic mutations.

The opportunities arising from single-cell sequencing (sc-seq) are enormous: only now is it possible to re-evaluate hypotheses about differences between pre-defined sample groups at the single-cell level—no matter if such sample groups are disease subtypes, treatment groups, or simply morphologically distinct cell types. It is therefore no surprise that enthusiasm about the possibility to screen the genetic material of

the basic units of life has continued to grow. A prominent example is the Human Cell Atlas [5], an initiative aiming to map the numerous cell types and states comprising a human being.

4. Motivation: Single-cell RNA sequencing (scRNA-seq) is increasingly used to study gene expression at the level of individual cells. However, preparing raw sequence data for further analysis is not a straightforward process. Biases, artifacts and other sources of unwanted variation are present in the data, requiring substantial time and effort to be spent on pre-processing, quality control (QC) and normalization.
Rapid development of technologies and methods that permit a detailed analysis of the genome and transcriptome of a single cell is offering a window into the extent and nature of genomic and transcriptomic heterogeneity which occurs in both normal development and disease. Single-cell approaches stand poised to revolutionize our capacity to understand the scale of genomic, epigenomic, and transcriptomic diversity that occurs during the lifetime of an individual organism. This session will review the major technological and biological breakthroughs achieved, describe the remaining challenges to overcome, and provide a glimpse into the promise of recent and future developments.

5.Challenges

- The way single-cell transcriptomics are entering the realm of big data.
- Technical and scientific challenges which are unique to the processing and analysis of large numbers of cells.
- Opportunities opened through big data driven statistical confidence and machine learning performance.
- optimized to improve computation and memory efficiency .

6. Timeline

Reference:

1. Single-Cell Analysis in Biotechnology, Systems Biology, and Biocatalysis:https://www.researchgate.net/publication/223981252_Single-Cell_Analysis_in_Biotechnology_Systems_Biology_and_Biocatalysis
2. Advances And Future Perspectives In Single Cell Analysis:https://www.pmwcintl.com/session/advances-and-future-perspectives-in-single-cell-analysis_2020sv/
3. Eleven grand challenges in single-cell data science: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-1926-6