Single cell biology and single cell RNA sequencing

Anastasia Ramanchanka
Modeling of Complex Biological System

April 6, 2025

1 Motivation

Humans are highly organized systems composed of approximately 3.7210^{13} cells of various types forming harmonious microenvironments to maintain proper organ functions and normal cellular homeostasis. Almost all cells in the human body share the same genetic material, but their transcriptomes reflect unique cellular activities. While traditional approaches focused on gene expression differences between tissue types or conditions, research over the last three decades has revealed significant heterogeneity among cells within a single tissue. This cellular heterogeneity has driven the need for genome-wide RNA profiling at single-cell resolution, revolutionizing our understanding of cellular states, transcription processes, and gene regulation mechanisms [1].

The resolution of target data remains a crucial consideration. For instance, studies from 70 years ago demonstrated that inductive cues often produce all-or-none responses in individual cells, yet these appear as gradual changes when measured across cell populations [2]. Equally important are intercellular interactions, as cells constantly engage in signal transduction.

2 Basics of single cell RNA-seq

Since its introduction in 2009 [3], single-cell RNA sequencing (scRNA-seq) has undergone rapid technological advancements, resulting in reduced costs and increased throughput. The standard scRNA-seq workflow comprises: single-cell isolation, cell lysis, reverse transcription (converting RNA to cDNA), cDNA amplification, and library preparation.

While different techniques employ various cell capture methods, they all incorporate unique barcoding of transcripts following cDNA conversion. Common isolation approaches include limiting dilution, fluorescence-activated cell sorting (FACS), magnetic-activated cell sorting (MACS), microfluidic systems, and laser microdissection. However, scRNA-seq faces methodological challenges such as artificial transcriptional stress responses, where cell dissociation procedures may induce stress-related gene expression, alter transcriptomes, and potentially compromise accurate cell type identification [4].

Following first-strand cDNA synthesis, amplification occurs via either polymerase chain reaction (PCR) or in vitro transcription (IVT). To address amplification biases, unique molecular identifiers (UMIs) are employed to label individual mRNA molecules during the reverse transcription step [5].

Library preparation in scRNA-seq must contend with both technical variability (including RNA capture efficiency, stochastic transcription, and batch effects) and biological noise (stemming from cellular heterogeneity and dynamic states). While cost considerations often restrict sequencing to 3′ or 5′ ends, this approach necessarily sacrifices full-length transcript information [6]. Consequently, protocol optimization to minimize RNA loss and enhance data precision remains a critical focus.

As important as the correct laboratory preparation of data is, so is the correct preprocessing and analysis of the scRNA-seq data. The classic roadmap for data analysis mainly consists of data preprocessing, general analyses and exploratory analyses:

- **Data preprocessing** includes quality control, alignment and quantification.
- **General analyses** include low-quality cell filtering, normalization, Highly Variable Gene selection, dimension reduction, clustering and annotation of cell types.
- **Exploratory analyses** include Differential Gene Expression analysis (identify marker genes between clusters), Function Enrichment (pathway analysis), Gene Set Variation analysis (score cells for pathway activity).

3 Application

scRNA-seq enables precise identification and classification of cell types within complex tissues. Unlike bulk RNA-seq, which averages gene expression across thousands of cells, scRNA-seq captures the transcriptional signatures of individual cells, primarily through clustering and dimensionality reduction techniques.

A particularly valuable application involves mapping both stable and transient cellular states during differentiation or reprogramming processes. These transcripts may appear either as highly expressed in rare cell populations (for example, present at ten copies in one of 100,000 cells) or as low-level ("leaky") expression across larger cell subsets.

Tumors are ecosystems of malignant cells and stromal components (immune cells, fibroblasts, vasculature). The technology reveals alterations in gene expression programs occurring in both transformed cells and their surrounding microenvironment.

Current developmental directions for scRNA-seq technology include spatial transcriptomics applications, multi-omics integration approaches, and expanded clinical implementations.

References

[1] Rickard Sandberg. "Entering the era of single-cell transcriptomics in biology and medicine". In: *Nature America* (2014). DOI: 10.1038/nmeth.2764.

[2] A. Weiner Novick. In: *Proc. Natl. Acad. Sci. USA* (1957).

[3] Fuchou Tang et al. "mRNA-Seq whole-transcriptome analysis of a single cell". In: *Nature America* (2009). DOI: 10.1038/nmeth.1315.

[4] Van Den Brink et al. "Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations". In: *Nat Methods.* (2017). DOI: 10.1038/nmeth.4437.

[5] Glenn K. Fu et al. "Counting individual DNA molecules by the stochastic attachment of diverse labels". In: *Proc. Natl. Acad. Sci. U.S.A.* (2011). DOI: 10.1073/pnas.1017621108.

[6] Legnini I. et al. "FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control". In: *Nat Methods* (2019). DOI: 10.1038/s41592-019-0503-y.