Advancing Drug Discovery and Development through Single-Cell RNA Sequencing and Deep Learning

Name: Nhi Luu
University of Maryland, Baltimore County
BTEC453: Biochemical Engineering
Dr. Fahimeh Mirakhori
Submission Date: 12/08/2023

Table of Contents

Abstract

In recent years, single-cell RNA sequencing (scRNA-seq) technology has enabled researchers to investigate further into insights of cellular processes and disease mechanisms, enhancing the current research on drug discovery and development process at the transcriptomic levels. Continuous advancements and optimizations of scRNA-seq with more sophisticated deep learning (DL) models play a significant role in the discovery of biomarkers, validation tools for targeted therapies, as well as prediction models, overcoming limitations of traditional and current screening methods (Van de Sande et al., 2023). Despite its great opportunities in biomedical research, scRNA-seq is still considered a novel approach, and several challenges need to be addressed before making its way into the industry and clinical settings. Nevertheless, scRNA-seq and its integration with DL are undeniably promising in advancing single-cell omics and clinical applications in the near future.

Introduction

Every day, scientists and researchers around the world make important discoveries in biological systems, looking at questions related to the complexities of cellular processes, how nature works, understanding the fundamental mechanisms underlying many illnesses. These curiosity-driven research projects are the foundations that aim to support further studies in drug development and clinical research, providing advanced treatments, applications, and solutions to human health. The development of a new medicine that starts from basic science as a molecular target to the medicine cabinet as an approved drug can take up to 15 years, through dozens of steps and many potential failure points. According to Nature Reviews in Drug Discovery, this journey of a new medical intervention often takes “too long, costs too much, and too often ends in failure” (Gilliland et al., 2016). Consequently, scientists have introduced various methods and technologies to accelerate the drug discovery and development process in the past decades, from automation to more advanced sequencing strategies. Furthermore, recent advancements in single-cell RNA sequencing (scRNA-seq) technologies paired with Deep Learning (DL) have proven success by improving the speed and quality of the drug candidates in the discovery and clinical development process (Van de Sande et al., 2023). Although there are many challenges regarding the industrial and clinical implementation of scRNA-seq and its technological advancements, it is undeniably one of the technologies that drive the evolution of targeted and personalized therapies, paving the way for advanced omics study in the near future.

Significance of scRNA-seq in Drug Discovery and Development

Drug discovery and development can be viewed as the total steps to identify a new substance, transforming it into an approved product to be used by patients, from drug discovery, pre-clinical development, to clinical development broken down into 5 different phases and

eventually regulatory approval by the FDA for commercialization (Mohs & Greig, 2017). Consequently, many drug candidates do not move through the entire process due to the strict and rigorous requirements and evaluation which may take 10-15 years on average and over several billion dollars to pull off (Mohs & Greig, 2017). One of the limitations of the drug discovery and development process arises from a reductionist view of a disease. For more than two decades, drug discovery has operated by using concepts of targets and pipelines, under the assumption that a single gene product can trigger a therapeutic response and that a compound in vitro can mimics or reflects its effectiveness in vivo (Duval, 2018). However, as human biology is complex and characterized by a high degree of heterogeneity, while focusing on isolated molecular targets or pathways, the approach may not capture the full spectrum of factors that contribute to disease and treatment outcomes, which prompts scientists to explore the finer details of cell transcriptomics.

Ribonucleic acid (RNA) has multiple forms and plays a critical role in cell growth and differentiation in the central dogma of life. RNA expression is tightly regulated in protein synthesis, playing a central role in the translation of genetic information from DNA into functional proteins. Understanding RNA provides insights into how genes are expressed and regulated in biological processes, cellular functions, and disease mechanisms. Often, abnormal expression of RNA may lead to the development and progression of diseases, and different gene expression profiles may reflect different disease subtypes, as well as tumor microenvironment (Li & Wang, 2021). Thus, RNA sequencing offers a comprehensive view of gene expression, including the identification of differentially expressed genes, alternative splicing events, genetic variations like single nucleotide polymorphisms (SNPs), and RNA editing. Since its first appearance in 2007, bulk RNA-seq has been the most extensively used tool in disease

classification, biomarker discovery, gene fusion, as well as disease prediction (Li & Wang, 2021). However, this approach, along with other conventional techniques such as amplification-based techniques (microarrays) or qRT-PCR could only assess cell and tissue characteristics in bulk by measuring RNA transcripts in a bulk of cells and cannot distinguish subpopulations of cell types (Van de Sande et al., 2023). Techniques like fluorescence-activated cell sorting (FACS), immunohistochemistry, and cytometry, though valuable, have limitations related to their small-scale targets and the requirement for prior biological insights (Van de Sande et al., 2023). Compared to other established methods, scRNA-seq technology is still in its early stages of development. However, scRNA-seq is extensively used in research and development (R&D) and its analytical workflow is continuously developing, being refined, and improved with newer cutting-edge tools and automation. Furthermore, scRNA-seq has also been used as a validation tool to repurpose other existing technologies, such as FACS to optimize the protocols for biomarker discovery in cancer research (Yang et al., 2020). Given the diverse gene activities in transcription at the individual cell level, the need for the characterization of individual cell types and personalized treatments, drives the advancement and utilization of scRNA-seq technologies (Van de Sande et al., 2023; Li & Wang, 2021). When combined with emerging concepts in Deep Learning (DL) and Artificial intelligence (AI), as well as collaborative efforts of global public health data, scRNA-seq is contributing to improve disease understanding, target identification, and biomarker discovery in both preclinical and clinical drug development phases.

Drug Discovery and Development with scRNA-seq and Deep Learning

Overview of scRNA-seq Processing Pipeline

Since the first scRNA-seq study was published in 2009, numerous scRNA-seq technologies have emerged, enabling in-depth analysis of transcriptional differences and cellular

heterogeneity at the single-cell level (Andrews et al., 2021; Haque et al., 2017). Along with the development of the 10X genomic chromium system, more than 20 scRNA-seq platforms and protocols are generated, each with specific experimental approaches and key research questions that have been developed in less than a decade (Figure 1) (Svensson et al., 2018).

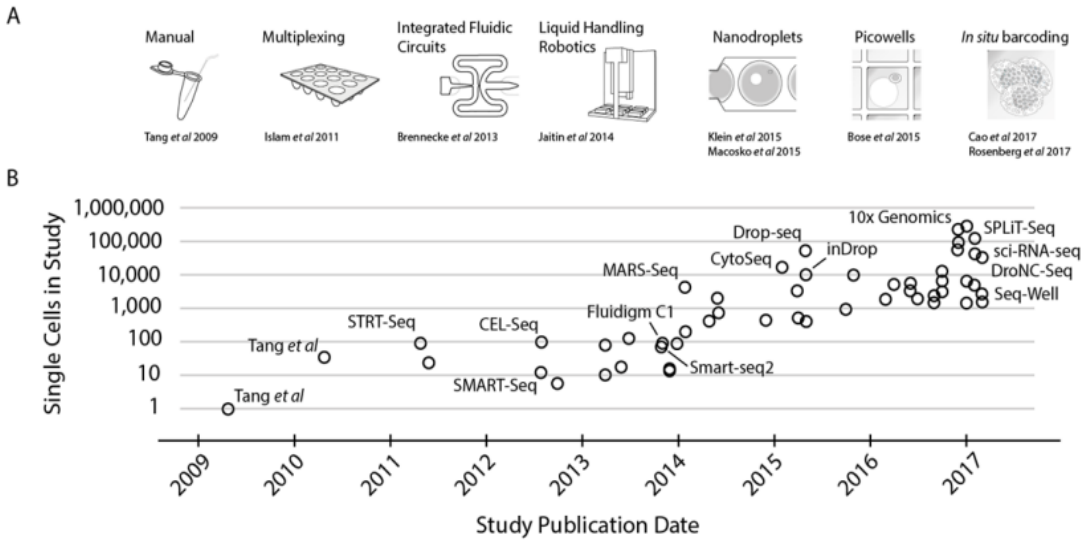

Figure 1. Advancements of scRNA-seq Sequencing Platforms. The number of analyzed single cells grow exponentially with the advancement of scRNA-seq studies and sequencing platforms (Svensson et al., 2018).

Most scRNA-seq tool aims to measure a distribution of expression levels across a population of cell, mostly to study heterogeneity of individual cells while keeping low required RNA material and high cellular resolution (Haque et al., 2017). While different in the way single cells are captured as well as quantification of RNA abundance, they adhere to a general pipeline which shares three key steps: Library generation, data pre-processing and downstream analysis, where DL plays a crucial role (Svensson et al., 2018; Andrews et al., 2021; Van de Sande et al., 2023; Haque et al., 2017).

Library Generation

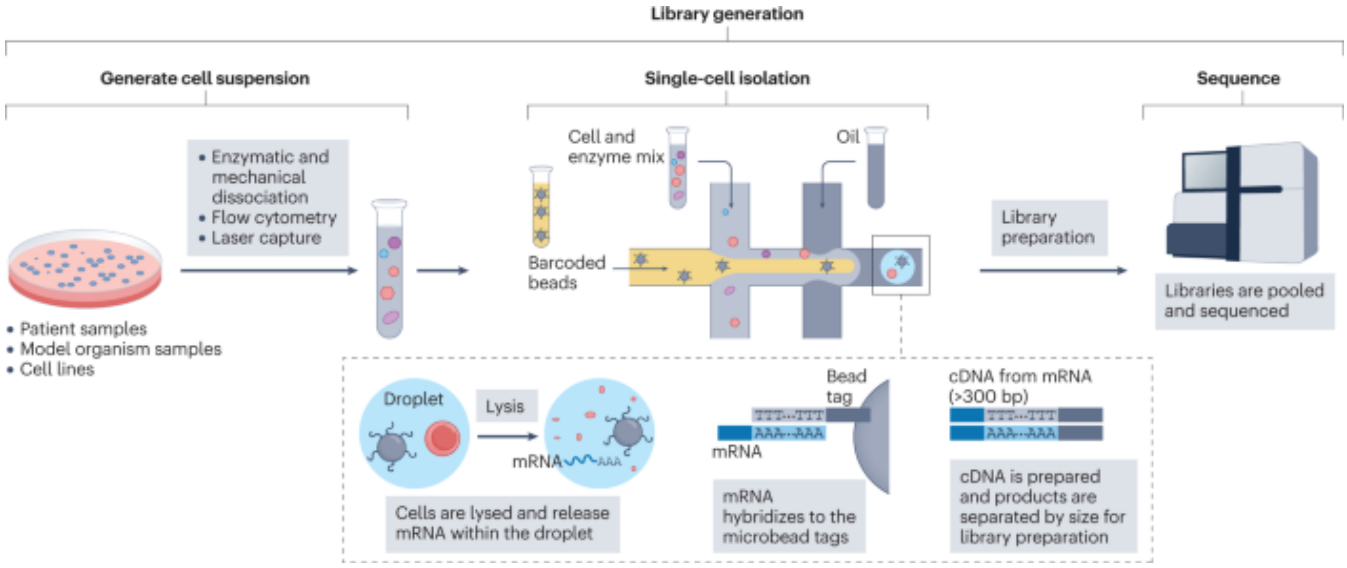

Figure 2. Library Generation Process in scRNA-seq. The primary steps in scRNA-seq processing pipelines involve the isolation of single cells and selection of sequencing technology for further downstream processing (Van de Sande et al., 2023).

Sample preparation is the first step to ensure individual cells or nuclei are isolated, while preserving quality for cell lysis and RNA capture in the next steps. Figure 2 shows the generation of sequencing library through multiple steps from single-cell suspension, isolation, to library construction and sequencing steps. Input material for an scRNA-seq experiment can be varied from established cell lines such as CHO, HEK293 or biological tissue samples from patients and model organisms like tumor biopsies, organoids, and 3D cultures (Luecken & Theis, 2019; Van de Sande et al., 2023). Depends on the data and research questions (information depth, cell numbers, time, and costs), the most widely used sequencing options are microwell- based, such as CEL-seq2, Drop-seq, or droplet-based techniques, such as 10X Chromium technology (Wolfien et al., 2021). Subsequently, RNA transcripts from each cell are tagged with unique molecular identifiers (UMIs) and a cDNA library is formed by utilizing full-length or tag-based protocols (Van de Sande et al., 2023). Full-length-based protocols seek to provide even

sequencing coverage across the entire length of RNA transcripts, while tag-based protocols selectively capture either the 5'-end or the 3'-end of RNA molecules (Wolfien et al., 2021). It is crucial to determine the appropriate sequencing methods to answer specific scientific questions as the technique used determines the accuracy, depth, and applicability of subsequent analysis for later processing steps computationally.

Deep Learning Integration: Data Pre-processing and Downstream Analysis

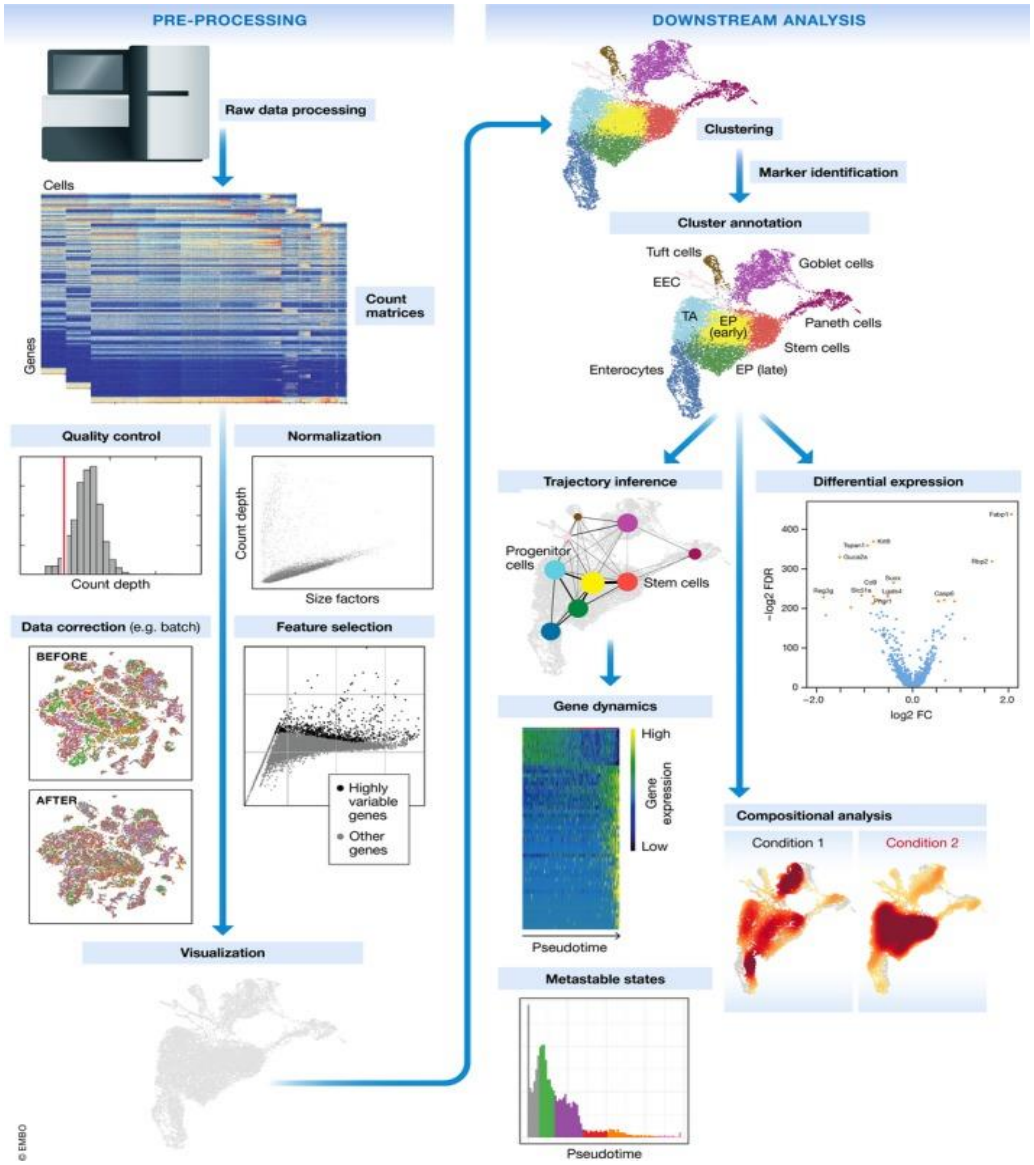

Figure 3. Overview of scRNA-seq Analysis Workflow. From previous library generation steps, raw sequencing reads are transformed into count matrices, a standard format for scRNA-seq analysis, then processed through subsequential steps for further downstream analysis and visualization with advance DL methods (Luecken & Theis, 2019).

In recent years, machine learning models have shown tremendous breakthrough in research, analyze complex biomedical data and accelerate discoveries. DL, a subcategories of machine learning, is a powerful tool to discover hidden and informative patterns within complex data through an artificial neural network, inspired from the structure and function of the human brain. Thus, DL models are extensively applied to scRNA-seq analysis steps due to their ability to analyze thousands to millions of trainable parameters, enabling them to extract non-linear patterns directly of the data from a "deep", multilayer network structure (Brendel et al., 2022). DL frameworks such as recurrent neural networks (RNN), convolutional neural networks (CNN), and autoencoders (AE), Variational AE (VAE), as well as generative adversarial networks (GAN) are all popular DL approaches extensively applied towards advanced research, not only scRNA-seq, but also spatial transcriptomics and 3D image analysis (Erfanian et al., 2023).

Several scRNA-seq studies have demonstrated the tremendous potential of deep learning in uncovering underlying patterns from scRNA-seq data, effectively addressing issues related to sequencing reads out, noise, and sparsity inherent in scRNA-seq workflows as seen in Figure 3 (Brendel et al., 2022; Ma & Xu, 2022; Luecken & Theis, 2019). In the data pre-processing steps, first, raw reads must be converted with alignment tools, into appropriate format called count matrices or read matrices which shows information about the number of times each gene is expressed in each cell (Wolfien et al., 2021; Luecken & Theis, 2019). Depends on the strategies used to sequence and align the reads, further downstream analysis will require different approach and additional steps during cleaning and filtration (Wolfien et al., 2021). For instance, traditional methods for bulk genome or transcriptome alignment and quantification can be utilized for reads from plate-based technologies while droplet-based platforms required specialized tools to handle

data with multiple cells and accurately assign UMI counts to cell barcodes (Wolfien et al., 2021). One of the technical challenges of scRNA-seq during the preprocessing step is that most of the retrieved reads are “noisy”. Noisy data can arise from a variety of sources, such as technical error, cDNA amplification bias, inefficiency in mRNA capture, dropouts leading to zero counts for genes, and potential cell state changes or death during isolation (Haque et al., 2017; Brendel et al., 2022). In scRNA-seq, quality control (QC) is performed to ensure the counts are correctly captured and of high quality (minimal noise), and that all counts are correspond to viable cells (count depth) (Wolfien et al., 2021). For data correction, such as batch effect removal, imputation and denoising as well as feature selections, DL-based approaches such as DeepImpute, scIGAN, and AE-based methods like scGMAI, SAVER-X, DCA, and ZINBAE, address non-linearity and offer improved imputation, potentially remove bias in selection of genes or cells (Andrews et al., 2021).

scRNA-seq sets apart from other sequencing methods due to their computational downstream analysis supported by multiple DL frameworks. As figure 3 depicts, the downstream analysis of scRNA-seq workflows reveals several high-resolution biological insights at the single cell level. For instance, dimensionality reduction techniques, such as t-distributed stochastic neighbor embedding (t-SNE) or uniform manifold approximation and projection (UMAP) provide the visualization of cell groupings in two- or three-dimensional space, useful for annotation of cluster and marker identifications within tumors (Van de Sande et al., 2023; Luecken & Theis, 2019). The annotated cluster can go through further studies, such as identification of distinctive genes through differential expression analysis, annotation of cell types, as well as cell-to-cell communications to study disease mechanism, or trajectory analysis

to track cell differentiation that are extensively used towards research on cancer disease progression (Van de Sande et al., 2023; Liu et al., 2021).

Predictive Model in Drug Discovery and Potential Applications

Drug Screening and Target Discovery

The initial phase of the drug development process starts with a target discovery program, focusing on a specific disease area and develop potential drug candidates. Modern drug discovery program utilizes systematic approaches such as high-throughput screening (HTS) to test thousands of potential targets and chemical compounds to identify a new drug-target combination, or rational drug design to design specific compounds based on known structure (Duval, 2018). For instance, the current HTS tests roughly 100,000 to a million compounds at a single dose, single condition to find "hits" with desired biological activity and further refine them into the lead compound preclinical production and testing (Van de Sande et al., 2023). Even though these methods remained the basis for drug discovery, it can take a lot of time, efforts, and a lot of starting materials to end with just one or two potential targets, perhaps none. While novel HTS approaches with scRNA-seq reads out test for a smaller number of compounds (roughly 100-1000), they are capable of testing multiple dosages and various conditions simultaneously, providing more comprehensive analysis on the drug mode of action, and specificity at the single cell levels in addition to UMAP, resulting in higher success rates towards lead compound optimization (Van de Sande et al., 2023; Luecken & Theis, 2019). In addition to compounds screening in drug discovery, scRNA-seq has also made its way through enhancing the process of biomarker discovery. For instance, early-phase clinical trials have incorporated scRNA-seq screening to predict biomarkers, associated with effective drug response or resistant

and provide an unbiased characterization of patients based on their cellular profiles and potential disease progression (Van de Sande et al., 2023).

Recent years, the biopharma industry has directed towards the exploration of targeted therapy, such as Chimeric antigen receptor (CAR)-T cell therapy, a type of immunotherapy that turn the body's T-cells into efficient cancer-fighting machines. Despite some clinical successes of CAR-T cell therapy, the major setback of the treatment is the potential toxicity "on-target, off-tumor" effects, possibly resulting in adverse events and limited response rates (Lareau et al., 2021). Thus, the success of CAR-T cell therapies relies on the specificity of receptor design and the appropriate modifications to T cells in newer CAR-T cell generations, which result in precise activation of T cells. In one study, scRNA-seq is utilized to identify tumor antigens, creating a tumor antigen map with predicted toxicity and safety profiles, including known targets that have been used in clinical settings (Lareau et al., 2021). Furthermore, researchers can access CAR-T cells expression profiles at different stages of treatment through trajectory analysis and record the transcriptional heterogeneity and trace the differences to further improve next generation designs (Castellanos-Rueda et al., 2021). Similar to CAR-T cell therapy, scRNA-seq has also been integrated with CRISPR/Cas9, another rising targeted therapy, to understand and explore how individual cells react to genetic modifications or disruptions from the technology, providing the more information on the precision of gene editing (Nomura, 2020).

Understanding Diseases through Preclinical Models

Through multilayers of learned DL networks and algorithms, as well as large-scale projects and databases of the species atlas, scRNA-seq provides a sophisticated, unbiased, and noninvasive approach to the insights of many disease mechanisms in cell development, cancer biology, immunology, microbiology, and more (Van de Sande et al., 2023; Jovic et al., 2022). In

complex diseases, such as cancer, tumor heterogeneity and microenvironment have been a tough challenge as current biological models still rely on simplifications and reductions of biological systems, which may not fully capture the depth of the disease. To date, scRNA-seq has been extensively applied to study disease heterogeneity, genetic and phenotypic variations, as well as cell development trajectory (Liu et al., 2021).

Most successful applications of scRNA-seq over other analytical methods is the exploration of cell and gene developmental trajectory at the single cell level with high resolution, specifically in cancer research relating to tumor and cancer stem cells (Yang et al., 2020). In one study on pancreatic ductal adenocarcinoma (PDAC) utilizing scRNA-seq, researchers revealed malignant tumors with complex heterogeneity and expression profiles that enable various growth and metastatic abilities (Liu et al., 2021). Popular trajectory analysis algorithms such as CellRouter or STREAM allow researchers to discover the developmental stages of the cells, observing their transitions and differentiations, especially in oncogenes, malignant cells, and fate of stem cells of high clinical values (Liu et al., 2021; Li et al., 2021). scRNA-seq also highlights the heterogeneity of individual epithelial cell, stromal cell, immune cell and most importantly, the profiling of distinct cells within tumors of highly malignant and heterogeneous cancer such as lung cancer, melanoma, and acute myeloid leukemia (Van de Sande et al., 2023; Li et al., 2021). Not only cancers, scRNA-seq has also been applied in mouse models to study the gene expression in neurodegenerative diseases such as Parkinson or Alzheimer's, predicting the potential genomic alterations that may lead to impairment of neural functions (Van de Sande et al., 2023). In infectious diseases, scRNA-seq has aided in analyzing the immune profiles of the disease to determine the severity, potential targets, and corresponding treatments. In a study about the metabolic changes due to COVID-19, scRNA-seq revealed the complex interaction

between different types of immune cells, such as T-cells, with more alterations and expansion that correspond differently in mild and severe cases of COVID-19 (Liu et al., 2022). The study indicates that it is promising to combine the high-resolution profiling and transcriptomic analysis of scRNA-seq with T- and B-cell receptor lineage analysis to assist in vaccine design and antibody discovery (Liu et al., 2022). scRNA-seq is also used in the characterization of bacterial colonies in biofilms, and heterogenous evolution during infections, such as in tuberculosis (Van de Sande et al., 2023).

Challenges and Opportunities of scRNA-seq

With the continuous development and success of scRNA-seq and its DL algorithms, the technology has become one of the most powerful tools utilized in drug discovery and development to unlock new findings. While scRNA-seq applications have demonstrated significant progress and opportunities over the last few years, they have been accompanied by several challenges and questions related to their economic perspective, regulatory requirements, and data accessibility as well as ethical considerations in clinical settings.

Economic and Regulatory Perspective

According to the statistics of Clinical and Translational Medicine (Figure 4), the rise of scRNA-seq technology in 2009 have contributed to the increase in the number of analyzed cells along with a significant decrease in cost of sequencing per cell (Jovic et al., 2022). As researchers continuously enhance algorithms, more sophisticated tools have been developed to address not only the broader aspects of biological systems but also the finer details embedded within their heterogeneity, such as in situ barcoding, spatial transcriptomics, and pseudo time analysis on disease progression and more. However, integration of scRNA-seq on the current economic and regulatory requirements is still a challenging aspect.

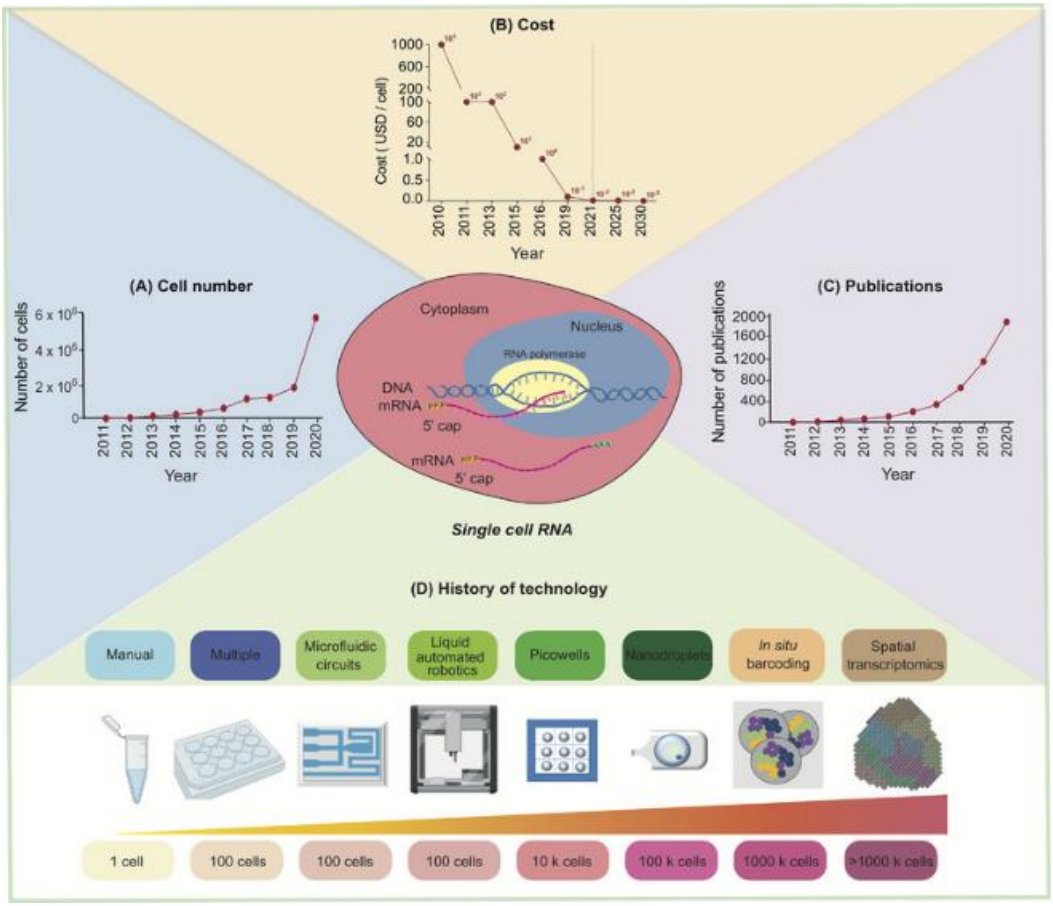

Figure 4. Advancements of scRNA-seq Technologies. A graphical overview of advancements in scRNA-seq technologies relating to sequenced number per cell, cost, and rising publications and newer scRNA-seq methods**.** (Jovic et al., 2022)

First, as there are currently no established and standardized protocols for scRNA-seq processing, most scRNA-seq clinical applications are still in exploratory phases or stay within R&D. In economic perspective, even though scRNA-seq reduce the cost per cell, the labor and sequencing cost per sample is still relatively high and not worth the efforts to change the current production infrastructure (Jovic et al., 2022). Additionally, computational capacity for sample processing and automation, as well as data storage and scale-up capability may not be as economically friendly as it seems. In regulatory perspective, scRNA-seq processing are not beginner-friendly, requiring advanced knowledge in the sequencing steps and adaptability of bioinformatics tools usage to the specific research questions. One lab personnel cannot simply be

trained to handle the specialized equipment and protocols but need the in-depth knowledge and computational skills to use the tools, ensuring consistency and quality in scRNA-seq products. Second, scRNA-seq processing produces a lot of data and they are still relatively complex and have not yet been fully studied (Liu et al., 2021). Most scRNA-seq readouts have limited applicability as the analysis is very specific and based on researchers' standpoint and available packages, which may be difficult to interpret and validate to federal guidelines. With the rapid growth of data and new ways to interpret scRNA-seq data, standardization of scRNA-seq protocols can be very complex (Luecken & Theis, 2019). Thus, there is no reason consider a change in production as current established methods are still capable of achieving similar outputs. For instance, most companies are aiming at large-scale drug screening for precision therapies and the potential candidates can be achieved through the low-cost, simple, and fast detection speed of gene expression microarray (Yang et al., 2020). Furthermore, HTS and bulk screening are still the basis for drug screening, both of which are optimized for quite sometimes and are able screen more compounds that scRNA-seq can (Van de Sande et al., 2023).

Data Availability and Ethical Considerations

Another challenge of scRNA-seq in the current pipeline is the available data and whether its integrity, quality and accessibility are up to standards for ethical considerations and clinical implementation. First, it has been reported that current scRNA-seq technologies are still not optimized to detect low-abundance transcripts, in which the lost RNA content can be reached up to 60% (Wang et al., 2021). While scRNA-seq provides better resolution on cell heterogeneity, low copy transcripts in individual cell may exclude importance characteristics of the cells. Second, while most scRNA-seq data and packages are made publicly available and follow common scRNA-seq processing steps, it is challenging to utilize the results from different

scRNA-seq projects as the training and testing sets for DL predictions. Additionally, most scRNA-seq workflows utilize available sequencing tools, such as the 10X Chromium system or BD Rhapsody (Haque et al., 2017). However, the sequencing reads out are usually processed and analyzed through automated and personalized scripts with no criteria to validate on the quality, consistency and efficiency of these data (Van de Sande et al., 2023; Jovic et al., 2022). Third, with no standardize protocols and validation steps, the format of scRNA output and terminology used are also differ by the studies. As mentioned, depends on the library construction protocol, the raw data generated by the sequencing machines, such as 10X Chromium or BD Rhapsody, and whether the protocols incorporate UMIs, the reads out will be in different formats (Luecken & Theis, 2019). The normalization methods and quality control metrics to filter out cells used in data pre-processing will also impact downstream analysis significantly. Moreover, studies may use different identifiers for the same or similar individual cells that potentially lead to confusion or duplicated data, leading to challenges in generalizability of the data (Van de Sande et al., 2023). Lastly and most importantly, ethical considerations and consent of patient samples, such as tumor biopsy, used towards scRNA-seq is still an issue that needs to be addressed. In clinical settings, tissues or cells from participants are usually harvested for research purposes upon informed consent in which patients are notified and agreed to the purpose, potential risks, and future implications regarding privacy of the data (Liu et al., 2022). However, as scRNA-seq research is still developing with advanced DL networks with most scRNA-seq data are available publicly, there is a very fine line between the primary and other usage of the data, either storage for research or for any other economical means. Such as in the case of Henrietta Lacks, collectively known as the HeLa cell line with tissue samples taken for research and industrialized for profits without her knowledge or consent (Beskow, 2016). Similarly, scRNA-seq technology

can advance so rapidly with large public databases that the current informed consent guidelines now may not be applicable in the future, especially with the growth of DL and more advanced artificial intelligence.

Conclusion

Since the initial introduction of scRNA-seq, the technology has been utilized extensively in R&D sector and continuously been optimized on its downstream analysis to overcome the limitations of traditional bulk sequencing methods and other screening technologies (Van de Sande et al., 2023). The development of scRNA-seq and its integration with DL frameworks stands out as one of the most promising technologies over the past decades across multiple fields in the life sciences, from target discovery to vaccine design and creation of various prediction models. Indeed, scRNA-seq has significantly impacted the drug discovery and development pipeline, advancing understanding of biological insights at the single cell level, enabling researchers to explore the cellular processes, disease mechanisms and progression, as well as therapeutic responses and more to come in the near future.

Along with the rapid applications and successes, as well as advancements of scRNA-seq technology, it is crucial to address the challenges that hinder its implementation in industrial and clinical settings. While scRNA-seq is a powerful technique, the sequencing and regulatory costs, non-uniform protocols, and requirements for skilled personnel prevent the implementation of scRNA-seq in industrial pipeline, especially when alternative methods are still adequate for the current production (Jovic et al., 2022). Furthermore, the growth of scRNA-seq technology indicates the growth of big data availability publicly, requiring the needs for standardize protocols, data integrity, and requirements on consent, especially when implementing scRNA-seq into clinical interventions (Luecken & Theis, 2019).

With the rise of single cell RNA sequencing, researchers are diving into further insights, such as spatial transcriptomics, and single cell analysis with other omics study. One promising application of scRNA-seq is its integration into personalized medicine and clinical diagnoses (Jovic et al., 2022). For instance, scRNA-seq trajectory analysis can be applied to study transitional states of cells in early stages of diseases, such as cardiac or cancer to predict and develop appropriate course of treatments (Wang et al., 2021). Another application of scRNA-seq is its combination technology with other targeted therapy techniques, such as CAR-T or CRISPR-based genetic screening, to improve upon precision and specificity, minimizing the "on-target, off-tumor" effects (Jovic et al., 2022). Taken together, scRNA-seq technologies will greatly expand biological insights in various types of diseases and ultimately set a foundation for future precision medicine, paving the way for newer technologies in single cell omics with enhanced DL and artificial intelligence.

References

Andrews, T. S., Kiselev, V. Y., McCarthy, D., & Hemberg, M. (2021). Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data. *Nature Protocols*, *16*(1), 1–9. https://doi.org/10.1038/s41596-020-00409-w

Beskow, L. M. (2016). Lessons from HeLa Cells: The Ethics and Policy of Biospecimens. *Annual Review of Genomics and Human Genetics*, *17*(1), 395–417. https://doi.org/10.1146/annurev-genom-083115-022536

Brendel, M., Su, C., Bai, Z., Zhang, H., Elemento, O., & Wang, F. (2022). Application of Deep Learning on Single-cell RNA Sequencing Data Analysis: A Review. *Genomics, Proteomics & Bioinformatics*, *20*(5), 814–835. https://doi.org/10.1016/j.gpb.2022.11.011

Castellanos-Rueda, R., Roberto, R. B. D., Schlatter, F. S., & Reddy, S. T. (2021). Leveraging Single-Cell Sequencing for Chimeric Antigen Receptor T Cell Therapies. *Trends in Biotechnology*, *0*(0). https://doi.org/10.1016/j.tibtech.2021.03.005

Duval, M. X. (2018). The inadequacy of the reductionist approach in discovering new therapeutic agents against complex diseases. *Experimental Biology and Medicine*, *243*(12), 1004–1013. https://doi.org/10.1177/1535370218794365

Erfanian, N., Heydari, A. A., Feriz, A. M., Iañez, P., Derakhshani, A., Ghasemigol, M., Farahpour, M., Razavi, S. M., Nasseri, S., Safarpour, H., & Sahebkar, A. (2023). Deep learning applications in single-cell genomics and transcriptomics data analysis. *Biomedicine & Pharmacotherapy*, *165*, 115077. https://doi.org/10.1016/j.biopha.2023.115077

Gilliland, C. T., Zuk, D., Kocis, P., Johnson, M., Hay, S., Hajduch, M., Bietrix, F., Aversa, G., Austin, C. P., & Ussi, A. E. (2016). Putting Translational Science on to a Global Stage. *Nature Reviews Drug Discovery*, *15*(4), 217–218. https://doi.org/10.1038/nrd.2016.33

Haque, A., Engel, J., Teichmann, S. A., & Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine*, *9*(1). https://doi.org/10.1186/s13073-017-0467-4

Hsieh, C.-Y., Wen, J.-H., Lin, S.-M., Tseng, T.-Y., Huang, J.-H., Hsuan Cheng Huang, & Hsueh Fen Juan. (2023). scDrug: From single-cell RNA-seq to drug response prediction. *Computational and Structural Biotechnology Journal*, *21*, 150–157. https://doi.org/10.1016/j.csbj.2022.11.055

Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., & Luo, Y. (2022). Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and Translational Medicine*, *12*(3). https://doi.org/10.1002/ctm2.694

Lareau, C. A., Parker, K. R., & Satpathy, A. T. (2021). Charting the tumor antigen maps drawn by single-cell genomics. *Cancer Cell*, *39*(12), 1553–1557. https://doi.org/10.1016/j.ccell.2021.11.005

Li, L., Xiong, F., Wang, Y., Zhang, S., Gong, Z., Li, X., He, Y., Shi, L., Wang, F., Liao, Q., Xiang, B., Zhou, M., Li, X., Li, Y., Li, G., Zeng, Z., Xiong, W., & Guo, C. (2021). What are the applications of single-cell RNA sequencing in cancer research: a systematic review. *Journal of Experimental & Clinical Cancer Research*, *40*(1). https://doi.org/10.1186/s13046-021-01955-1

Li, X., & Wang, C.-Y. (2021). From bulk, single-cell to spatial RNA sequencing. *International Journal of Oral Science*, *13*(1). https://doi.org/10.1038/s41368-021-00146-0

Liu, J., Xu, T., Jin, Y., Huang, B., & Zhang, Y. (2021). Progress and Clinical Application of Single-Cell Transcriptional Sequencing Technology in Cancer Research. *Front. Oncol.*, *10*(Molecular and Cellular Oncology). https://doi.org/10.3389/fonc.2020.593085

Liu, W., Ye, X., An, Z., & Zhao, Z. (2022). The challenges and opportunities of scRNA-seq in COVID-19 research and clinical translation. *Virology Journal*, *19*(1). https://doi.org/10.1186/s12985-022-01923-x

Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology*, *15*(6). https://doi.org/10.15252/msb.20188746

Ma, Q., & Xu, D. (2022). Deep learning shapes single-cell data analysis. *Nature Reviews Molecular Cell Biology*, *23*(5), 303–304. https://doi.org/10.1038/s41580-022-00466-x

Mohs, R. C., & Greig, N. H. (2017). Drug discovery and development: Role of basic biological research. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *3*(4), 651–657. NCBI. https://doi.org/10.1016/j.trci.2017.10.005

Nomura, S. (2020). Single-cell genomics to understand disease pathogenesis. *Journal of Human Genetics*, *66*(1), 75–84. https://doi.org/10.1038/s10038-020-00844-3

Svensson, V., Vento-Tormo, R., & Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols*, *13*(4), 599–604. https://doi.org/10.1038/nprot.2017.149

Van de Sande, B., Lee, J. S., Mutasa-Gottgens, E., Naughton, B., Bacon, W., Manning, J., Wang, Y., Pollard, J., Mendez, M., Hill, J., Kumar, N., Cao, X., Chen, X., Khaladkar, M., Wen, J., Leach, A., & Ferran, E. (2023). Applications of single-cell RNA sequencing in drug discovery and development. *Nature Reviews Drug Discovery*, *22*(6), 496–520. https://doi.org/10.1038/s41573-023-00688-4

Wang, M., Gu, M., Liu, L., Liu, Y., & Tian, L. (2021). Single-Cell RNA Sequencing (scRNA-seq) in Cardiac Tissue: Applications and Limitations. *Vascular Health and Risk Management*, *Volume 17*, 641–657. https://doi.org/10.2147/vhrm.s288090

Wolfien, M., Róbert Dávid, & Galow, A.-M. (2021). Single-Cell RNA Sequencing Procedures and Data Analysis. *Bioinformatics*, *Chapter 2*, 19–35. https://doi.org/10.36255/exonpublications.bioinformatics.2021.ch2

Yang, X., Kui, L., Tang, M., Li, D., Wei, K., Chen, W., Miao, J., & Dong, Y. (2020). High-Throughput Transcriptome Profiling in Drug and Biomarker Discovery. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00019