GMR
The Original

SINGLE-CELL TRANSCRIPTOME ANALYSIS FOR IDENTIFYING CELLULAR HETEROGENEITY IN BIOLOGICAL SYSTEMS

Tharani Munusamy[1], Ramnath V[2], Dr. S. A. Keerthi[3], Kakarla Hari Kishore[4]

[1]Department of Research, Meenakshi Academy of Higher Education and Research
[2]Professor, Meenakshi College of Allied Health Sciences, Meenakshi Academy of Higher Education and Research
[3]Assistant Professor, Biochemistry, Sree Balaji Medical College and Hospital, Bharath Institute of Higher Education and Research, ORCID: https://orcid.org/0009-0007-7450-5027
[4]Professor, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India, Email: kakarla.harikishore@kluniversity.in

ABSTRACT
The cellular heterogeneity is involved in the basic interpretation of complex biological systems, which shapes such processes as the development, development of disease, and a response to therapy. Conventional bulk RNA sequencing methods tend to obscure cell-to-cell differences and thus restrict the characterization of rare or functionally separate classes of cell. With a revolution in molecular biology, nomenclature, and analysis, single-cell transcriptomics, specifically single-cell RNA sequencing (scRNA-seq) now allows the examination of the state of individual cells in detail, including gene expression and state comparisons across cells in vivo. The review gives a general overview of the latest development in single-cell transcriptome studies and their applications in identifying cellular heterogeneity in a wide range of biological systems. The most important technological platforms, which include Smart-seq, Drop-seq, and 10x Genomics, are mentioned and computational methodologies, including data preprocessing, normalization, dimensionality reduction, clustering and differential gene expression analysis. The recent uses of scRNA-seq in cancer biology, immunology and plant sciences are analyzed critically with emphasis to the discoveries of new cell subpopulations and mechanisms of regulation. Besides, the latest issues, such as technical noise, drop out events, and batch effects are discussed, and possible solutions are examined. The possible avenues of emerging multi-omics integration and spatial transcriptomics are also investigated. This survey seeks to present perspectives on the potential and constraints of single-cell transcriptomics to inform future studies about more specific and scalable analytical models.

KEYWORDS: Single-cell RNA sequencing, cellular heterogeneity, transcriptomics, gene expression analysis, clustering algorithms, biological systems

1. INTRODUCTION
The complexity of biological systems has been a major issue of concern in molecular and cellular biology (Wang and Bodovitz, 2010; Tanay and Regev, 2017). Conventional gene expression analyses have used bulk RNA sequencing methods meaningfully, to measure the mean transcriptional activity in a large population of cells. Although these methods have been useful in getting an idea about gene regulation and biological pathways, by nature, they obscure the variability that does exist between individual cells (Islam et al., 2011; Ramsköld et al., 2012). The effects of such averaging cause the resulting populations of cells to be blind to the existence of rare cell populations, transitional cellular states, and subtle regulatory variations that are usually essential to comprehend biological processes in greater depth. This means that numerous relevant biological phenomena will be elusive to study at a bulk level. The discussion of gene expression at individual cell level has turned out to be possible with the advent of single-cell transcriptomics, especially single-cell RNA sequencing (scRNA-seq) technology enabling the study of cells at single-cell resolution (Islam et al., 2011; Picelli et al., 2013; Hwang et al., 2018). This change of paradigm has enabled scientists to go beyond the averages of the population and the complex cellular diversity underlying complex biological systems (Tanay and Regev, 2017; Svensson et al., 2018). The phenomenon of cellular heterogeneity is a typical feature of multicellular organisms, which is very important in the development, physiological functioning, and the course of the disease (Shalek et al., 2013; Tirosh et al., 2016). Within tissues that seem to be uniform to the eye, individual cells in the same tissue can differ greatly in the patterns of their gene expression, their functioning and sensitivity to environmental stimulus. Although these differences are not accidental, they are also usually fundamental to biological functioning and flexibility. As an example, cellular heterogeneity in embryonic development participates in the differentiation of the stem cells into specialized types of cells that eventually results in development of complex tissues and organs. Heterogeneity helps maintain tissue homeostasis and tissue regeneration by allowing the dynamic response of cells to internal and external conditions in adult organisms. Cellular heterogeneity in the context of disease,

especially cancer, is one of the principal causes of tumor occurrence, growth, metastasis, and drug resistance (Tirosh et al., 2016). Tumours typically consist of a wide range of distinct subpopulations of cells with different genetic and transcriptional profiles, and are thus challenging to effectively treat with non-targeted therapeutic approaches. Likewise, heterogeneous populations of immune cells in immunology play a determinant role in specificity and strength of immune responses, affecting the outcome in the case of infections, autoimmune diseases and immunotherapies (Shalek et al., 2013). Cellular heterogeneity controls developmental events and responses to environmental stresses in vegetation systems in response to drought, salinity, and pathogen onslaughts (Denyer et al., 2019). Thus, the problem of cellular heterogeneity is paramount and requires determination and definition to further our knowledge of normal biological functioning as well as of disease pathogenesis. The invention of single-cell transcriptome analysis has now offered an unprecedented chance to uncover diversity at an extremely high resolution and scale, in the context of cells (Macosko et al., 2015; Svensson et al., 2018). ScRNA-seq can be used to identify individual cell types, infrequent subpopulations, and dynamic cell states, which traditional methods could have never accessed before, as it is able to capture the transcriptomic profile of isolated cells (Picelli et al., 2013; Macosko et al., 2015). Moreover, the single-cell methods can also be used to reconstruct their developmental dynamics and lineage relations, gaining knowledge of the cells that change immune response to different stimuli in connection with the lapse of time (Satija et al., 2015; Denyer et al., 2019). In tandem with technological developments, much has been done in the field of computational analysis of high-dimensional single-cell data. Quality control filtering, normalization, dimensionality reduction, clustering, and differential gene expression techniques are techniques that have become the norm in obtaining pertinent biological data out of the voluminous datasets (Chen et al., 2019; Luecken and Theis, 2019). The use of advanced algorithms and machine learning methods is beginning to increase with the goal of improving the quality and scalability of the analysis of single-cell data, making it possible to consider millions of cells in large-scale research (Kiselev et al., 2019; Lotfollahi et al., 2022). Although it is evident by the impressive advances made in single-cell transcriptomics that this technology has reached maturity, there are some obstacles that still restrain its potential. The variability is created by technical noise, dropout events, and batch effects, which may make data interpretation more challenging and ensure low reproducibility (Kiselev et al., 2019; Luecken and Theis, 2019). Single-cell data, which are also high-dimensional and sparse, represent an additional major burden to computationally demanding analytical frameworks (Zappia et al., 2018; Chen et al., 2019). Also, multiple-source and multi-platform data integration is a complicated task, especially when integrating the transcriptomic data into other omics layers, including proteomics, epigenomics, and metabolomics (Butler et al., 2018; Stuart et al., 2019). The furthest concern in improving the trustworthiness and utility of single cell analyses across research and clinical practices is to tackle these obstacles. To this end, the current review will critically summarize the concept of single-cell transcriptome analysis as an effective methodology in the study of heterogeneity of cells in biological systems. It is a review of the recent developments in sequencing technologies, computational approaches, and analytical plans that have influenced the emergence of the field (Hwang et al., 2018; Chen et al., 2019; Luecken and Theis, 2019). Significant applications are also mentioned in terms of various fields in the review, such as cancer biology, immunology, and plant sciences, which indicate the wide-ranging relevance of single-cell methods to the study of complex biological processes (Shalek et al., 2013; Tirosh et al., 2016; Denyer et al., 2019). Moreover, it mentions future limitations and challenges, as well as the rising trends of new research in this field, including multi-omics integration, spatial transcriptomics, and artificial intelligence-powered analysis, which are likely to make a difference on the new level of research in this area (Stahl et al., 2016; Stuart et al., 2019; Lotfollahi et al., 2022). The compilation of existing knowledge and the presence of gaps in the literature will ensure that this work will inform future researchers in their work and enable the further development of single-cell technologies to gain a better understanding of the human biological aspect and help to achieve better treatment results.

2. LITERATURE REVIEW AND BACKGROUND.

Transcriptomics has experienced a radical revolution during the last twenty years, changing the level of population-wide gene expression analysis to the single-cell level of resolution (Wang and Bodovitz, 2010; Tanay and Regev, 2017). First transcriptomic work was mainly guided by microarray technologies that allowed the simultaneous measurement of the thousands of genes but did not offer more dynamic range, predetermined probes, and sensitivity. A significant development was represented by the introduction of bulk RNA sequencing that enabled unbiased transcript detection with a better level of accuracy and increased coverage. Bulk RNA-seq made a great contribution to the analysis of gene expression patterns during development, disease, and functional genomics. Nevertheless, regardless of the advantages, bulk RNA sequencing is inherently a procedure that quantifies an average transcriptional signal of a homogenised group of cells. Such averaging effect obscures the heterogeneity of actual cell states in tissues and does not allow identifying rare subpopulations, intermediates in lineage, and minor expression variations which can have very significant biological implications (Islam et al., 2011; Ramskoeld et al., 2012). Due to this, the increasing complexity of the cellular aggregate that needed to be resolved at a more detailed level emerged as one of the primary contributors to the emergence of the single-cell transcriptomic technologies. Single-cell RNA sequencing became a groundbreaking answer to this problem since it allowed defining transcriptome profiles at the cell level (Islam et al., 2011; Hwang et al., 2018). Such a change in the analysis of biological systems at the bulk to the cell scale has redefined the knowledge of the biological

systems, specifically where tissue heterogeneity is playing a significant role (Tanay & Regev, 2017). The initial scRNA-seq represents sequencing low numbers of cells with high sensitivity that enables the finding of previously unidentified transcriptional diversity (Islam et al., 2011; Ramskoeld et al., 2012). Smart-seq and its enhanced versions were among the first technologies that allowed complete coverage of transcripts with high sensitivity and therefore were applicable to characterising transcripts in detail, splice variant identification, and low-input samples (Picelli et al., 2013). All these methods however cost a lot and had a low throughput. In order to meet the need of high-scale profiling, droplet-based technologies, including Drop-seq and subsequent 10x Genomics Chromium, demonstrated a significant improvement in scalability. Those platforms enabled mass profiling thousands to tens of thousands of cells simultaneously by encasing the single cells into microdroplets and labelling their transcripts with distinct molecular barcodes (Macosko et al., 2015; Svensson et al., 2018). This technology has immensely widened the use of scRNA-seq in complicated tissues, developmental domains, and disease microenvironment. Additional technology combinations, such as Seq-Well, CEL-seq, MARS-seq and combinatorial indexing designs also led to a further fragmentation of the technological base with each providing various trade-offs in terms of sensitivity, throughput, cost and technical complexity. All these gains have made scRNA-seq one of the most significant domains in current genomics (Hwang et al., 2018; Chen et al., 2019). Along with the rapid growth of scRNA-seq technologies, there has been a corresponding growth in computational analysis technology. Single-cell data are high-dimensional, sparse, and technically noisy; therefore, raw sequencing results require a huge amount of processing, until biological inferences can be made using them. It has also resulted in powerful quality control, normalisation, batch correction, dimensionality reduction, clustering, and differential expression analysis as well as cell-type annotation computational pipelines (Chen et al., 2019; Luecken and Theis, 2019). Located at the heart of analysing single cells more informatically and normalised, the widely used software ecosystems are Seurat, Scanpy, Monocle and SC3 (Zappia et al., 2018; Luecken and Theis, 2019). Principal component analysis, t-distributed stochastic neighbour embedding, uniform manifold approximation and projection are two examples of dimensionality reduction methods used to visualise dysfunctional transcriptional landscapes (Satija et al., 2015; Kiselev et al., 2019). Additional clustering and trajectory inference techniques based on graphs have allowed the detection of cell populations, developmental lineages and transitional states (Kiselev et al., 2019; Luecken and Theis, 2019). Such advances in the use of analysis have not only enhanced the interpretation of data; but have also increased the biological questions to be answered by using the study of single cell. Specifically, cross-linking of machine learning and network-based technologies has created fresh prospects in detecting regulatory schemes, marker genes and latent cellular designs in massive datasets (Lotfollahi et al., 2022). The transformational power of single-cells transcriptome analysis is now known to have significant effects in diverse areas of biological science. ScRNA-seq has also found extensive applications in cancer biology to dissect the intratumoral heterogeneity, define therapy-resistant subclones, define tumour microenvironment, and reveal transcriptional programmes associated with metastasis and disease progression (Tirosh et al., 2016). Such studies have demonstrated that the previously believed rather homogenous tumours are actually a collection of various malignant and non-malignant cell types performing different functional activities. The single-cell transcriptomics has been used in immunology with the ability to generate high-resolution profiles of immune cell subsets, to understand dynamic responses to infection, inflammation, vaccination, and immunotherapy (Shalek et al., 2013). This has greatly enhanced the knowledge on immune regulation and communication between cells. ScRNA-seq in developmental biology has facilitated comprehensive maps of lineage specification and differentiation enabling scientists to trace developmental events as well as define transient progenitor states (Satija et al., 2015). The approaches have also been useful to plant biology, where single-cell study has provided cell-type-specific gene expression patterns in root development, leaf morphology, vascular organisation and stress adaptation (Denyer et al., 2019). On the same note, research studies in neuroscience, regenerative medicine, and stem cell biology have further established that the extent and dynamics of cellular richness is much broader and wider than thought before. Collectively, all available literature demonstrates that scRNA-seq is one of the most formidable platforms to explain the heterogeneity of cells and the organisation of molecules in a complex system (Tanay and Regev, 2017; Hwang et al., 2018). Through these developments, however, significant challenges and unaddressed issues, which still form the sphere, can be found in the literature as well. A large number of weaknesses include technical noise and dropout instances where transcripts reflected in a cell are not detected by a low capture efficiency or sequencing depth. This leads to low coverage of datasets and makes downstream interpretation more difficult (Kiselev et al., 2019; Luecken and Theis, 2019). Another longstanding issue is batch effects caused by variations in sample processing, library preparation, sequencing conditions, and so on, especially in large multi-sample/ multi-centre studies (Butler et al., 2018; Stuart et al., 2019). Moreover, many computational tools are proposed, yet there is not a universally accepted pipeline of analysis and the output can be different depending on which workflow is used. This standardisation will impact on reproducibility and render cross-study comparisons challenging (Zappia et al., 2018; Chen et al., 2019). The other crucial gap is the biological interpretation. Most studies within this area give much attention to computational clustering and visualisation but offer less mechanistic validation of the discovered populations of cells or regulatory patterns. In others, computationally obtained clusters are considered to be definitive biological objects that cannot be supported with adequate experimental factors. Moreover, single-cell transcriptomics is a measure of transcriptional diversity, but it is no longer representative of other dimensions of cellular regulation,

namely chromatin accessibility, protein abundance, metabolic state, or spatial organisation within tissues. This disadvantage has encouraged swelling enthusiasm toward multi-omics and spatially resolved technologies, even though these areas are in the long run and has their own technical and computational constraints (Stahl et al., 2016; Stuart et al., 2019). It is also noted in the literature that most of the currently existing research is highly focused in specific fields; especially oncology and immunology and there are other biological systems that have relatively not been explored. Single-cell transcriptomics is becoming more applicable in plant sciences, ecological biology, non-model organisms, and translational clinical modes of application (Denyer et al., 2019). In addition, interpretation scalability has been a bottleneck although high-throughput technologies have made it possible to analyse large datasets (Macosko et al., 2015; Svensson et al., 2018). It continues to be a challenge to apply large amounts of single-cell data to the biological actionable knowledge the presence of a robust computational framework, reliable annotations, and significant validation schemes all remains needed to do so. The other gap that is very important is when it comes to integration of data across platforms, species, and modalities. Differences in sequencing chemistry, capturing abilities, and definition of features may make it challenging to harmonise data and reduce the usefulness of publicly available data (Butler et al., 2018; Stuart et al., 2019; Lotfollahi et al., 2022). Therefore, despite the rapid development of the field, there are still serious methodological and conceptual problems. Considering these findings, it is possible to state that the existing collection of studies indicates that the single-cell transcriptomics is at the phase of extensive applicability, but not yet the full maturity (Tanay & Regev, 2017; Luecken and Theis, 2019). The discipline has already proven the importance of studying gene expression on a cellular scale, but further innovation is required to make it more sensitive, minimise technical artefact, standardise computations, and better biologically interpret the results. Future research should go further than the documentation of cell clusters to a more integrative perspective of the functionality, mechanism, and cellular communication. It is against this need that this review is placed. It suggests that by synthesising the development of transcriptomics and critically evaluating scRNA-seq technologies, compiling representative studies, and pinpointing those gaps in research that could be persistent, it is possible to offer a consistent historical background of understanding through which such single-cell transcriptome analysis can further develop the study of the heterogeneity of cells in biological systems.

3. Single Cell Transcriptomic Data Processing.

The processing of single-cell transcriptomic data is the initial step of the scRNA-seq analysis, which is devoted to converting raw sequencing data into a format that could be analyzed. The phase is paramount due to the fact that the quality of preprocessing is directly proportional to subsequent analysis accuracy and biological interpretation. Data acquisition, quality control (QC), and normalization are the usual steps of the pipeline, that is, to ensure that true biological signals are left despite the presence of technical noise and biases. The stage to start data acquisition is the generation of raw sequencing reads that come out of single cells using 10x Genomics, Smart-seq, or Drop-seq. These reads are then mapped to a reference genome to generate a gene expression matrix, with rows (genes) and columns (individual cells). And on the foundation of this matrix all analyses henceforth are founded. Figure 1 shows an overview of the entire workflow of processing single-cell transcriptome, starting with the preparation and concluding with the analysis of samples, thus offering a general impression of the scRNA-seq pipeline, comprising the isolation of cells, the creation of libraries, sequencing, and the analysis of the obtained data.

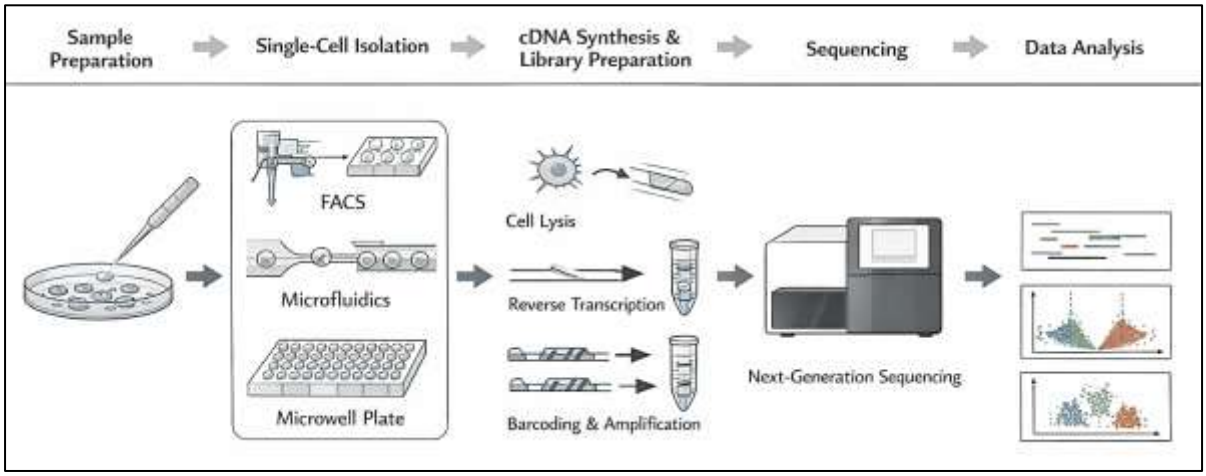

Figure 1. Schematic Representation of the Single-Cell RNA Sequencing (scRNA-seq) Workflow Pipeline

The quality control is a necessary process designed to eliminate the cells of low quality and erratic measurements of the genes that could emerge as a result of technical shortcomings. The standard QC measures are defined as the total detected genes per cell, percentage of these who are expressed in the mitochondrion, and sequencing depth. Extremely few counts of genes in cells can indicate dead or damaged cells and excessively high numbers may indicate doubled or tripled cells. Likewise, genes that are presented in a small number of cells are often

filtered to minimise noise. Such data filtering techniques increase the reliability of the data and also guarantee that subsequent analysis is conducted on signals that have biological significance. Normalisation is then done following quality control in order to consider differences in sequencing depth and library size between cells. In the absence of normalisation, cell comparisons of gene expression levels would not be based on inherent biological variation but instead be biased by technical variation. One of the popular ways to normalise the data is to scale the gene expressions values using the total expression of each cell, and multiply the resulting one by a constant factor in order to simplify the data. Equation (1) gives the mathematical representation of this process:

$$X_{norm} = \frac{X_{ij}}{\sum_j X_{ij}} \times 10^6, \quad (1)$$

where X_{ij} denotes the raw expression of gene i in cell$j$, and the denominator represents the total expression across all genes in that cell. This type of normalisation, also known as counts per million (CPM) allows an easy comparison of the levels of gene expression between different cells due to the elimination of the library size effects. In general, the preprocessing phase brings about the idea of converting the raw scRNA-seq data to a high-quality normalised dataset that can undergo further analysis. This step has the groundwork by systematically considering tissue variability and noise to be able to accurately identify cellular heterogeneity and interpretation of biological needs.

4. ANALYSIS METHODS OF HETEROGENEITY OF CELLS.

All the processes of identifying and interpreting cellular heterogeneity in single-cell transcriptomic data have become dependent on sophisticated analysis tools that reduce high-dimensional mass of gene expression data to biologically significant knowledge. After preprocessing, analytical solutions are used to reveal concealed structures of the data that allow recognizing separate cell populations, functional conditions, and governing patterns. They are mostly going to involve techniques of dimensionality reduction, clustering, differential gene expression analysis and cell-type identification, which can be combined to achieve a more complete view of cellular diversity. Dimensionality reduction is a highly sought after phase of the analytical step, which simplifies but do not erase important biological variation of data concerning high-dimensional gene expression. Single-cell dataset often contains thousands of genes in a single cell, and it is not easy to visualise and analyse these data directly. Dimensional reduction methods like Principal Component Analysis (PCA) are usually applied to first of all identify significantly large sources of variation within the data. Nonlinear algorithms like Uniform Manifold Approximation and Projection (UMAP) and t-distributed Stochastic Neighbour Embedding (t-SNE) have, however, become the most popular because they are able to preserve both local and global data layouts in a better way. The methods allow intuitively visualisation of cell populations in two-dimensional space and allow the detection of clusters representing different cellular subtypes. Such dimensionality reduction methods show good effectiveness in distinguishing organised cellular groupings, as demonstrated in Figure 2 where well-distinct clusters are the heterogeneous cell groups as obtained using single-cell transcriptomic data.

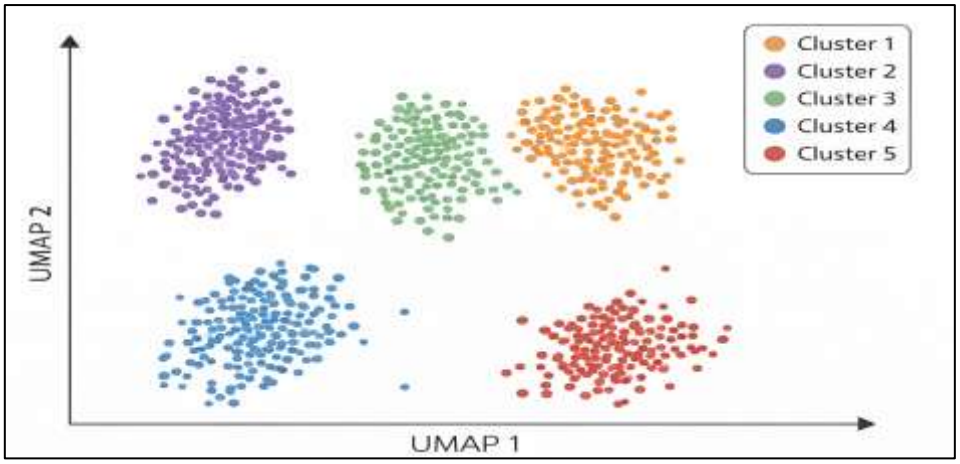

Figure 2: UMAP-Based Visualization of Cellular Clusters Revealing Heterogeneity in Single-Cell Transcriptomic Data

After analysis of dimensionality, clustering is used to cluster cells that share similar expression of genes within specific cluster. It is assumed that these clusters are biologically relevant cell types or states. Several clustering algorithms have been designed to cluster single-cell data, such as k-means clustering, hierarchical clustering as well as graph-based (the Louvain and Leiden algorithms). Among them, the graph-based type of clustering has emerged as the standard of clustering because it is a scalable data classification method with the capability to establish more complex community patterns in large datasets. The techniques can be used to identify fine differences between cell population by building a nearest-neighbour graph and subdividing it into clusters, thereby identifying rare or otherwise character-understudied cell types. To further describe the identified clusters, the

analysis of differential gene expression (DGE) is conducted to reveal a set of genes that are upregulated or downregulated significantly under various conditions of assessing different cell populations. This is an essential step in realising the molecular biology of cellular heterogeneity and in knowing markers genes of a particular cell type. The most common metric of measuring changes in gene expression between clusters is the fold change which is mathematically modelled by Equation (2):

$$FC = \log_2\left(\frac{Mean_A}{Mean_B}\right), \quad (2)$$

where $Mean_A$ and $Mean_B$ represent the average expression levels of a gene in two distinct cell clusters, respectively. Variance is stabilised with the logarithmic transformation and can be interpreted symmetrically between upregulation and down regulation. If the genes with high fold change values exist, these genes are usually discussed as possible biomarkers, and are utilised to differentiate between various cellular populations. The last part of the analytical pipeline is cells-type identification, which annotates clusters according to the existing biological knowledge. It includes analysis of the expression of identified clusters and comparing them to known marker genes of reference databases or published works. It has been a focus of automated annotation tools and reference-based mapping methodologies, which simplify this process but manual validation is still necessary to ensure that there is a biological accuracy. New cell types or intermediate forms can often also be identified, too, which underscores the potential of single-cell transcriptomics in revealing what was still unknown cellular diversity. All in all, these methods of analysis allow obtaining meaningful biological insights of complex single-cell data through a combination of methods. Combining dimensionality reduction, clustering, differential expression, and cell-type annotation, researchers are able to systematically describe the cellular heterogeneity of biological systems and reveal the molecular mechanism underlying the particular biological systems. Their combination, as Figure 2 shows, yields an interpretable and clear picture of separate cell groups, which motivates the role of advanced computational analysis in current single-cell studies.

5. BIOLOGICAL SYSTEMS USES.

The study of transcriptome at the single-cell level has greatly contributed to the knowledge on the complex biological systems by facilitation of the determination of heterogeneity of cells in various domains. It has been used in various fields and areas like cancer biology, immunology and even in the plant sciences among others where it has offered quality information on the composition of cells, functional diversity, and even dynamic processes in biology never known before. Single-cell RNA sequencing (scRNA-seq) has emerged as an effective instrument in cancer biology used to understand the intratumoral heterogeneity. Tumours are very complicated and composed of various populations of malignant, non-malignant cells and stroma, as well as, beneficiary and immune cells and cancer stem-like cells. This diversity is not represented in the traditional sequencing of bulk where results tend to be incomplete or confusing. Scalable scRNA-seq, on the contrary, allows identifying separate tumor subclones, cell groups that resist therapy, and metastasis and disease-related transcriptional programs. These high-resolution profiling techniques have enabled identification of new biomarkers and drugs of interest, hence playing a role in the generation of precision oncology approaches. In immunology, single-cell transcriptomics has transformed the research on performer and variety of immune cells. The immune system is a diverse assembly of cell types that have a specific role, and their dynamic interplay is the key to successful immune responses. ScRNA-seq makes it possible to characterize the immune cell subset, such as T cells, B cells, macrophages, and dendritic cells, in detail, in different physiological and pathological modes. it has been used in inferring immune responses in cases of infections, autoimmune diseases and cancer immunotherapy. Single-cell analysis methods can reveal the cellular activation status, lineage differentiation and network of intercellular communication by detecting transcriptional heterogeneity across cell populations associated with the immune response. ScRNA-seq has opened new prospects in plant biology in which plant development, tissue organization, and stress responses are not only studied, but also understood. Plant tissues are comprised of varying types of cells that help in growth and in transportation of nutrients as well as adapting to the environment. Transcriptomic studies of single cells have allowed one to determine cell type-specific patterns of gene expression in roots, leaves, and vascular tissues. Such insights have helped improve the study of developmental pathways, cellular differentiation and abiotic and biotic stress responses to drought, salinity and pathogen infection. Moreover, scRNA-seq has enabled identifying regulatory networks that control the growth and adaptation of plants, which can be used in crop enhancement and sustainable farming. The outcome of such applications is closely related to the sequencing technology chosen, with various platforms providing sensitivity, throughput and cost with differing degrees of sensitivity. Table 1 shows the main features of widely used scRNA-seq technologies, and their applicability to various biological problems.

Table 1. Comparison of Single-Cell RNA Sequencing Technologies

Technology	Throughput	Sensitivity	Cost	Key Features	Typical Applications
Smart-seq / Smart-seq2	Low	High	High	Full-length transcript coverage	Detailed gene analysis, isoform detection

Drop-seq	High	Moderate	Low	Droplet-based, scalable	Large-scale cell profiling
10x Genomics	Very High	Moderate	Medium	Commercial platform, high reproducibility	Clinical studies, large datasets
CEL-seq / CEL-seq2	Moderate	High	Medium	Linear amplification	Quantitative gene expression studies
Seq-Well	High	Moderate	Low	Portable and cost-effective	Field studies, resource-limited settings

On the whole, single-cell transcriptome studies have emerged as an essential tool to study the heterogeneity of cells within various biological systems. It allows the analysis of individual cells at a high resolution level which allows gaining a deeper understanding of intricate biological processes and helps further development in the fields of medicine, immunology and agricultural sciences.

6. CHALLENGES AND LIMITATIONS

Although the single-cell RNA sequencing (scRNA-seq) has made a major breakthrough, various obstacles and weaknesses still persist in the accuracy, scalability, and biological interpretation of the information. The causes of these challenges are both technical and computational factors based on high-dimensional single-cell data sets. The most noticeable problem in scRNA-seq data analysis is called dropout which occurs when transcripts actually expressed in a cell are not present in the data because of low capture efficiency or sequencing depth. This has the effect of giving sparse gene expression matrices that contain too many zero values and it is hard to tell when there is a real biological absence of the gene and when there is technical failure. There is a huge effect of dropout events on downstream analysis like clustering, and differential gene expression which might consequently result in a biased interpretation. Technical noise is another problem that has become very critical and is caused by the variability that is brought about in the course of sample preparation, amplification and sequencing. This sound may cause legitimacy of real biological signals and lower the dependability of identified patterns of gene expression. Besides that, batch effects are the systematic differences resulting due to differences in experimental conditions, reagent batches or sequencing platform. These effects may cause artificial aggregating of cells due to technical artefact and not because of actual biological variation and thus invalidate the analysis outcome. Scalability is also an expanding issue because the current scRNA-seq technology has the potential to produce datasets in the tens of thousands to millions of cells. The process of such large-scale data demands large computational resources and processing, storage and analysis efficient algorithms. The data is even more challenging, as the dimensionality and complexity of the data increase which means the need to create strong and scalable analytical systems. Table 2 provides an approximate overview of the most common problems and possible solutions in the analysis of scRNA-seqs, with the key challenges highlighted and a list of possible mitigation strategies.

Table 2. Challenges and Solutions in scRNA-seq Analysis

Challenge	Description	Common Solutions
Dropout Events	Missing gene expression due to low capture efficiency	Imputation methods (MAGIC, SAVER), data smoothing
Technical Noise	Variability introduced during sequencing and amplification	Normalization, noise modeling techniques
Batch Effects	Non-biological variation across experiments	Batch correction tools (ComBat, Harmony, MNN)
High Dimensionality	Large number of genes and sparse data	PCA, UMAP, dimensionality reduction techniques
Scalability	Large datasets requiring high computational resources	Parallel computing, cloud-based analysis platforms

Figure 3 conceptually depicts these challenges and their interactions in the research of single cell transcriptomics, in which a central position of single-cell RNA-seq data is linked to major problems, including data complexity, technical noise, dropout events, batch effects, and interpretation problems.

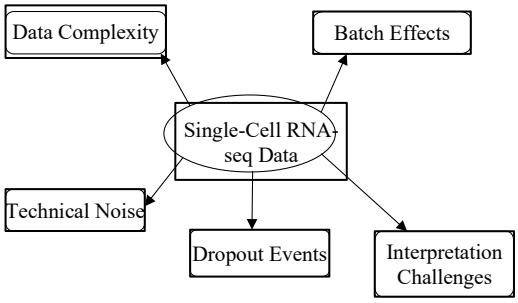

Figure 3. Key Challenges and Limitations in Single-Cell RNA Sequencing (scRNA-seq) Analysis

All in all, to enhance the strength, reproduction, and biological applicability of single-cell transcriptomic analyses, there is a necessity to address the following challenges. The naval innovations in experimental protocols and computational methods continue to improve the accuracy and scale of the scRNA-seq analysis of future studies as well.

7. EMERGING TRENDS AND FUTURE DIRECTIONS.

Single-cell transcriptomics is a rapidly developing field that faces ongoing developments in experimental techniques and computing algorithms. Although the traditional scRNA-seq has been used to offer meaningful insights regarding cellular heterogeneity, innovative methods are currently offering the ability to understand biological systems in a comprehensive and context-sensitive manner. Other important trends in the future of this area are spatial transcriptomics, integration of multi-omics, and the use of the methods of analytical artificial intelligence (AI). Spatial transcriptomics is one of the most notable developments that remediate one of the most considerable flaws of the traditional scRNA-seq whereby the cells in tissues are left in their natural environment. Common single-cell system methods involve dissociation of tissues, which causes the loss of positional cues relevant to cell interactions and tissue structure. The spatial transcriptomics technologies provide the capability to map the information on the gene expression to directly to the tissue sections and the researchers may examine the impact of cellular organization on the biological functions. The technique has been especially useful in cancer biology, neuroscience and developmental biology where cellular positions regarding one another are vital towards functional consequences. Through the combination of spatial and transcriptomic data, scientists will have a better understanding of the microenvironment of tissues and the cellular communication network. Multi-omics integration is another new direction, as it consists in integrating transcriptomic data with other biological information layers, including genomics, epigenomics, proteomics and metabolomics. Although scRNA-seq gives a view of gene expression, it fails to give the complete picture about cellular regulation. Multi-omics methods allow the holistic approach to the cellular function by connecting the pattern of gene expression/regulatory pattern with protein activity and metabolic status. Multi-modal sequencing platforms, single-cell ATAC-seq, CITE-seq, and similar technologies are enabling simultaneous measures of a wide range of molecular features in single cells. Nevertheless, uniting all these various datasets is still quite a computation task and specific algorithms and standardized frameworks are needed in order to make the interpretation correct and meaningful. Single-cell data analysis is also adopting artificial intelligence and machine learning. The Big Data of single-cell datasets are increasing in their dimensionality, sparsity, and complexity, and are being processed by AI-based algorithms. Deep learning, graph neural networks, and unsupervised clustering algorithms are all techniques allowing more precise cell type identification and prediction, computations of gene regulatory networks, and cellular trajectories. The AI-based methods also improve data integrations, noise, and identification of patterns, and allow to derive greater biological understanding out of large datasets. Along with increased and increasingly complex datasets, AI will become a crucial factor in enhancing scalability, automation, and accuracy of analysis. Moving forward, an intersection of these new developments will transform the future of single-cell biology. The combination of spatial information, multi-omics data and AI-based analytics will make the cellular heterogeneity a more comprehensive and systems-level comprehension. This will be completed in future studies that aim at the standardization of data, greater efficiency of computational and the gap to understanding results that results in the generation of large amounts of data. Such developments have major application potential in the areas of precision medicine, drug discovery, and sustainable agriculture, which eventually lead to a higher level of comprehending complex biological systems and better health conditions.

CONCLUSION

Single-cell transcriptome has become a revolutionary intervention of learning about cellular heterogeneity in complicated biological frameworks. Single-cell RNA sequencing (scRNA-seq) goes beyond the constraints of existing bulk sequencing technologies by offering finer details about the nature of cellular diversity, cellular states, and their regulation by enacting the expression of multiple genes in parallel on at least one cell. This review has underscored the chronicle of the transcriptomic technologies, improvements in the elaborate scRNA-seq systems and the significance of computational methods to the processing and evaluation of the high-dimensional single-cell data. Transcriptomic analysis of single cells in various fields of study such as cancer biology, immunology,

and plant sciences has exhibited its extensive application in the discovery of new populations of cells, biomarkers, and in better understanding of complex biological processes. Concurrently, various issues similar to the events of dropouts, technical noise, and batch effects, and scalability still restrain the maximum potential of the given technology. Improving experimental methods and sound computational schemes has been a major agenda of the discipline to such an extent that these weaknesses are addressed. With its publication, new trends have arisen such as spatial transcriptomics, multi-omics integration and analysis driven by artificial intelligence that are likely to further improve the resolution, accuracy and interpretability of single-cell studies. Such developments will allow a more systems and general view of the biological organisation, spanning the divide between molecular and functional data. In short, the analysis of single cell transcriptomes is a strong and fast-growing instrument to study heterogeneity of cells. Further development of technology, as well as data analysis, will be vital in order to unlock the full potential of it, and eventually, it will lead to the progress of precision medicine, biotechnology, and sustainable biological research.

REFERENCES

1. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology, 36(5), 411–420.
2. Chen, G., Ning, B., & Shi, T. (2019). Single-cell RNA-seq technologies and related computational data analysis. Frontiers in Genetics, 10, 317.
3. Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., & Timmermans, M. C. (2019). Spatiotemporal developmental trajectories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing. Developmental Cell, 48(6), 840–852.
4. Hwang, B., Lee, J. H., & Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. Experimental & Molecular Medicine, 50(8), 1–14.
5. Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J. B., Lönnerberg, P., & Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Research, 21(7), 1160–1167.
6. Kiselev, V. Y., Andrews, T. S., & Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. Nature Reviews Genetics, 20(5), 273–282.
7. Lotfollahi, M., Naghipourfar, M., Luecken, M. D., Khajavi, M., Büttner, M., Wagenstetter, M., ... & Theis, F. J. (2022). Mapping single-cell data to reference atlases by transfer learning. Nature Biotechnology, 40(1), 121–130.
8. Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular Systems Biology, 15(6), MSB188746.
9. Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., ... & McCarroll, S. A. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell, 161(5), 1202–1214.
10. Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., & Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature Methods, 10(11), 1096–1098.
11. Ramsköld, D., Luo, S., Wang, Y. C., Li, R., Deng, Q., Faridani, O. R., ... & Sandberg, R. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nature Biotechnology, 30(8), 777–782.
12. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology, 33(5), 495–502.
13. Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., ... & Regev, A. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature, 498(7453), 236–240.
14. Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., & Frisén, J. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science, 353(6294), 78–82.
15. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., & Satija, R. (2019). Comprehensive integration of single-cell data. Cell, 177(7), 1888–1902.
16. 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.
17. Tanay, A., & Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. Nature, 541(7637), 331–338.
18. Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., ... & Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science, 352(6282), 189–196.
19. Wang, D., & Bodovitz, S. (2010). Single cell analysis: the new frontier in 'omics'. Trends in Biotechnology, 28(6), 281–290.
20. Zappia, L., Phipson, B., & Oshlack, A. (2018). Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. PLoS Computational Biology, 14(6), e1006245.