



## OPEN ACCESS

EDITED BY  
Paola A. Marignani,  
Dalhousie University, Canada

REVIEWED BY  
David Molik,  
United States Department of Agriculture  
(USDA), United States  
Ron Stewart,  
Morgridge Institute for Research,  
United States

\*CORRESPONDENCE  
Jia Qian Wu,  
jiaqianwu@uth.tmc.edu  
Iván Velasco,  
ivelasco@ifc.unam.mx

SPECIALTY SECTION  
This article was submitted to Molecular  
and Cellular Pathology,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

RECEIVED 27 February 2022  
ACCEPTED 06 October 2022  
PUBLISHED 24 October 2022

CITATION  
Cuevas-Díaz Duran R,  
González-Orozco JC, Velasco I and  
Wu JQ (2022), Single-cell and single-  
nuclei RNA sequencing as powerful  
tools to decipher cellular heterogeneity  
and dysregulation in  
neurodegenerative diseases.  
*Front. Cell Dev. Biol.* 10:884748.  
doi: 10.3389/fcell.2022.884748

COPYRIGHT  
© 2022 Cuevas-Díaz Duran, González-  
Orozco, Velasco and Wu. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Single-cell and single-nuclei RNA sequencing as powerful tools to decipher cellular heterogeneity and dysregulation in neurodegenerative diseases

Raquel Cuevas-Díaz Duran<sup>1</sup>, Juan Carlos González-Orozco<sup>2</sup>,  
Iván Velasco<sup>2,3\*</sup> and Jia Qian Wu<sup>4,5,6\*</sup>

<sup>1</sup>Tecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Monterrey, Mexico, <sup>2</sup>Instituto de Fisiología Celular—Neurociencias, Universidad Nacional Autónoma de México, Mexico City, Mexico, <sup>3</sup>Laboratorio de Reprogramación Celular, Instituto Nacional de Neurología y Neurocirugía “Manuel Velasco Suárez”, Mexico City, Mexico, <sup>4</sup>The Vivian L. Smith Department of Neurosurgery, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, United States, <sup>5</sup>Center for Stem Cell and Regenerative Medicine, UT Brown Foundation Institute of Molecular Medicine, Houston, TX, United States, <sup>6</sup>MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX, United States

Neurodegenerative diseases affect millions of people worldwide and there are currently no cures. Two types of common neurodegenerative diseases are Alzheimer’s (AD) and Parkinson’s disease (PD). Single-cell and single-nuclei RNA sequencing (scRNA-seq and snRNA-seq) have become powerful tools to elucidate the inherent complexity and dynamics of the central nervous system at cellular resolution. This technology has allowed the identification of cell types and states, providing new insights into cellular susceptibilities and molecular mechanisms underlying neurodegenerative conditions. Exciting research using high throughput scRNA-seq and snRNA-seq technologies to study AD and PD is emerging. Herein we review the recent progress in understanding these neurodegenerative diseases using these state-of-the-art technologies. We discuss the fundamental principles and implications of single-cell sequencing of the human brain. Moreover, we review some examples of the computational and analytical tools required to interpret the extensive amount of data generated from these assays. We conclude by highlighting challenges and limitations in the application of these technologies in the study of AD and PD.

## KEYWORDS

single-cell sequencing, single-nuclei sequencing, Alzheimer’s disease, Parkinson’s disease, cellular heterogeneity, cellular vulnerability

## Introduction

Neurodegenerative diseases are characterized by a chronic and progressive structural and functional degeneration of cells in the central or peripheral nervous system, leading to massive neuronal loss. Unfortunately, there are currently no effective treatments for neurodegenerative diseases nor biomarkers for their early detection (Heemels, 2016). Data from the Global Burden of Diseases (GBD, 2020) demonstrates that Alzheimer's and Parkinson's diseases are the top neurodegenerative disorders contributing to the highest disability-adjusted life years (DALY) among the groups aged 75 or older (GBD 2019 Diseases and Injuries Collaborators, 2020). These common neurodegenerative diseases are predominantly idiopathic and have overlapping clinical and pathological features. Genome-wide association studies (GWAS) have identified genetic variants related to the risk of developing Alzheimer's disease (AD) and Parkinson's disease (PD). However, our understanding of the specific cell types dysregulated in these diseases remains elusive.

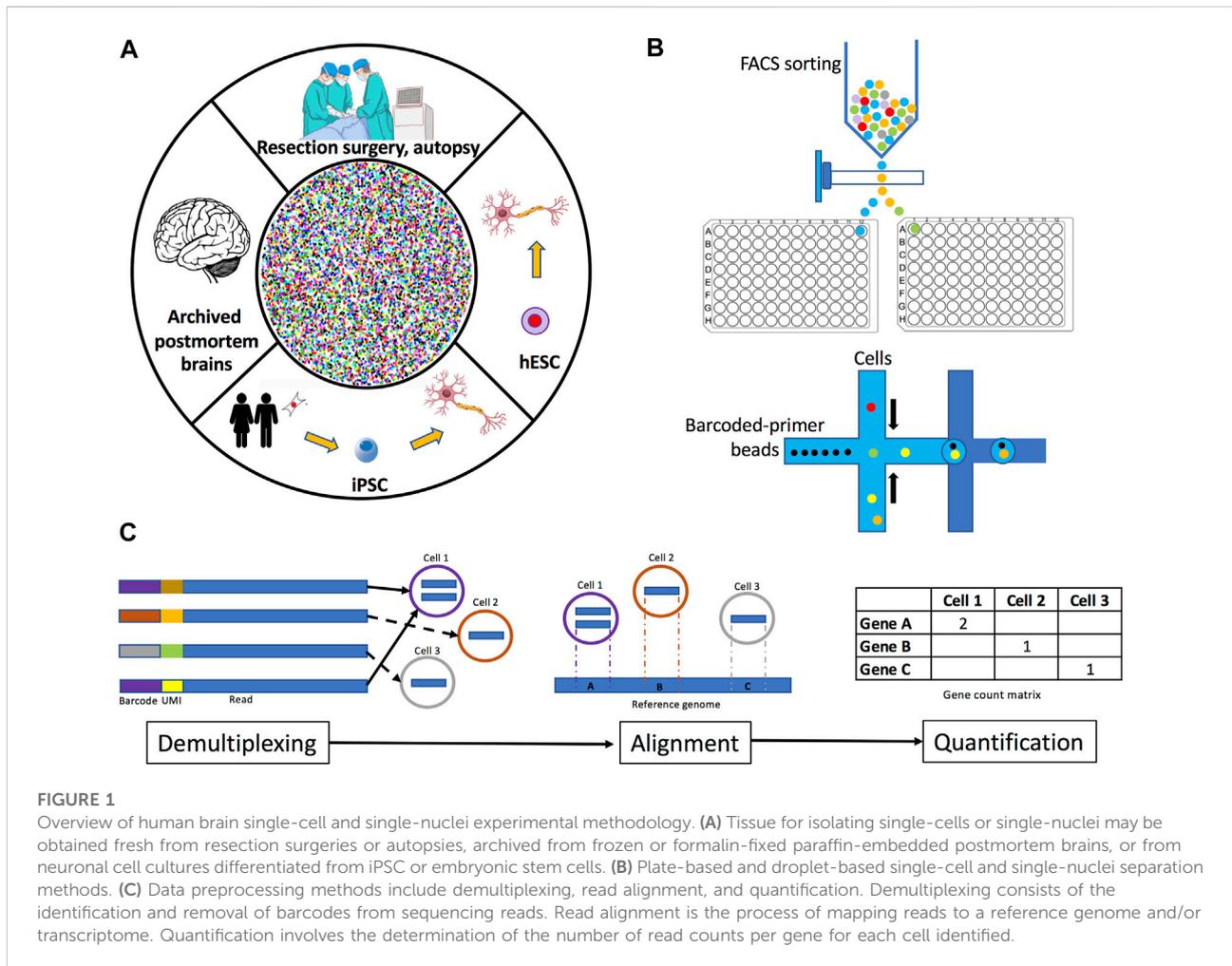
High throughput single-cell and single-nuclei mRNA sequencing (scRNA-seq/snRNA-seq) technologies have pioneered a new era of exploration into the diversity of brain cell types at the molecular level. Previously, the study of brain cells was limited to bulk assays in which the amount of RNA is averaged among all cells masking cell-type-specific gene expression (Colantuoni et al., 2011; Hawrylycz et al., 2012; Miller et al., 2014; Zhang et al., 2016). However, brain cells are heterogenous and functionally complex. ScRNA-seq/snRNA-seq technologies have provided an unprecedented opportunity to systematically investigate the expression of thousands of genes and cells, identifying cell subpopulations, and reconstructing temporal and spatial dynamics of gene expression. To this end, a comprehensive understanding of the brain cellular subtypes and their expression profiles in healthy physiological context and neurodegenerative conditions is required. Currently, scRNA-seq and snRNA-seq studies of mouse and human brains depict yields ranging from thousands of cells up to 1.3 million cells (Yao and van Velthoven, 2021). However, the scope of these studies can still be improved in the future given that the adult human brain has approximately 170 billion cells including a similar proportion of neuronal and non-neuronal cells (Azevedo et al., 2009). Exciting insights have been achieved with scRNA-seq and snRNA-seq in the field of human oligodendrocyte cell diversity (Marques et al., 2016), and functional states of human microglia (Masuda et al., 2019) among others. Furthermore, recent advances allow the correlation of molecular features with cellular processes such as proliferation leading to neurogenesis in the adult mouse hippocampus (Habib et al., 2016), as well as anatomical location and electrophysiological characteristics in the murine thalamus (Li et al., 2020).

Herein we will review the recent progress in understanding neurodegenerative diseases using single-cell sequencing. We will

particularly focus on AD and PD. We will discuss the fundamental principles and implications of scRNA-seq and snRNA-seq of the human brain. Furthermore, we will briefly describe some examples of the computational and analytical tools used to interpret the extensive amount of data generated from these assays. Also, we will focus on pioneering human transcriptomic studies which have addressed the cellular heterogeneity of the brain regions mainly affected in AD and PD. Finally, we will highlight challenges and limitations in the application of these technologies in the study of AD and PD.

## Single-cell/nuclei transcriptomic experimentation

The first step in performing scRNA-seq/snRNA-seq is to identify the brain region of interest. Human brain cells or cell nuclei may be isolated from several sources: fresh tissue, archived organs, or from cells differentiated *in vitro* (Figure 1A). Samples from fresh tissue are collected from resection surgeries, biopsies, or autopsies (Gradišnik et al., 2021), while samples from archived organs are obtained from postmortem frozen or formalin-fixed paraffin-embedded brain sections (Wang et al., 2013). If frozen postmortem brains are selected, the best option is to collect the transcripts to be sequenced from the nucleus (snRNA-seq) because freezing ruptures the plasma membrane, thus complicating the isolation of intact cells and therefore increasing the probability of capturing degraded transcripts from the cytoplasm (Krishnaswami et al., 2016; Slyper et al., 2020; Maitra et al., 2021). Similarly, it is highly recommended to perform snRNA-seq instead of scRNA-seq on samples obtained from formalin-fixed tissues since the fixation process often damages the integrity of various cellular structures, leading to the detection of heavily degraded cytoplasmic RNA (Esteve-Codina et al., 2017). Furthermore, very recent protocols have optimized the extraction of cell nuclei for sequencing from formalin-fixed tissues, including brain (Chung et al., 2022; Vallejo et al., 2022). An alternative sample source is *in vitro* differentiation of either embryonic stem cells (ESCs) or reprogrammed induced pluripotent stem cells (iPSCs), which are a reliable source of biological material since the processes of artificial induction to the desired cell lineage highly recapitulates the molecular processes that occur in the organism during the development and maturation of cells (Marei et al., 2017). Moreover, some studies have already compared through scRNA-seq the transcriptomes of differentiated cells (e.g., dopaminergic neurons) from iPSCs and ESCs to the transcriptomes of cells obtained from human brain tissue to assess the fidelity of *in vitro*-derived cells, observing that these cells retain the characteristics from their *in vivo* counterparts (La Manno et al., 2016; Fernandes et al., 2020). Additionally, some protocols optimized for scRNA-seq/snRNA-seq of *in vitro* differentiated cells recommend collecting the cells of interest



by FACS sorting to increase the reliability of the procedure (Cuomo et al., 2020).

As mentioned above, the selection of the sequencing method may depend on the conservation status of the sample, but also on the type of cells to be analyzed. Single-cell isolation from brain represents a challenge because neurons are very sensitive to enzymatic and mechanical dissociation methods. Brain samples can be meticulously collected, for example using laser microdissection, and then carefully processed with an enzymatic or mechanical digestion method to obtain a single-cell suspension. However, as mentioned, there are at least two challenges when performing brain cell dissociation. First, since neuronal axons, dendrites, and synapses form an intricate and highly connected network in the brain, their dissociation may affect cellular integrity and even modify transcriptional profiles by upregulation of stress-induced artifacts (Denisenko et al., 2020). Second, neuronal transcripts residing in distal cellular compartments will most likely be lost upon dissociation (Cajigas et al., 2012; Tushev et al., 2018).

These issues may lead to an underrepresentation of neuronal populations with respect to glial cells, which are more resistant to dissociation procedures as has been observed in single-cell suspensions obtained from human brain cortex (Darmanis et al., 2015). Therefore, careful consideration must be taken when selecting a cell dissociation method. Isolating single nuclei instead of single cells is a better option to analyze tissues that cannot be easily dissociated. Thus, an advantage of snRNA-seq is that it can capture transcripts from cells that are more susceptible to death in the process of dissociation for example, neurons from the brain cortex (Tasic et al., 2018). Moreover, nuclei isolation protocols are fast and do not require protease digestion or heating, reducing the probability of aberrant transcription (Lacar et al., 2016); also, nuclei protocols allow the profiling of large cells (<40  $\mu\text{m}$ ) that do not fit through microfluidics sequencing methods (Denisenko et al., 2020) and can be combined with gene regulatory studies, for example the assay for transposase-accessible chromatin sequencing (ATAC-seq). Despite all these advantages, an important concern for performing

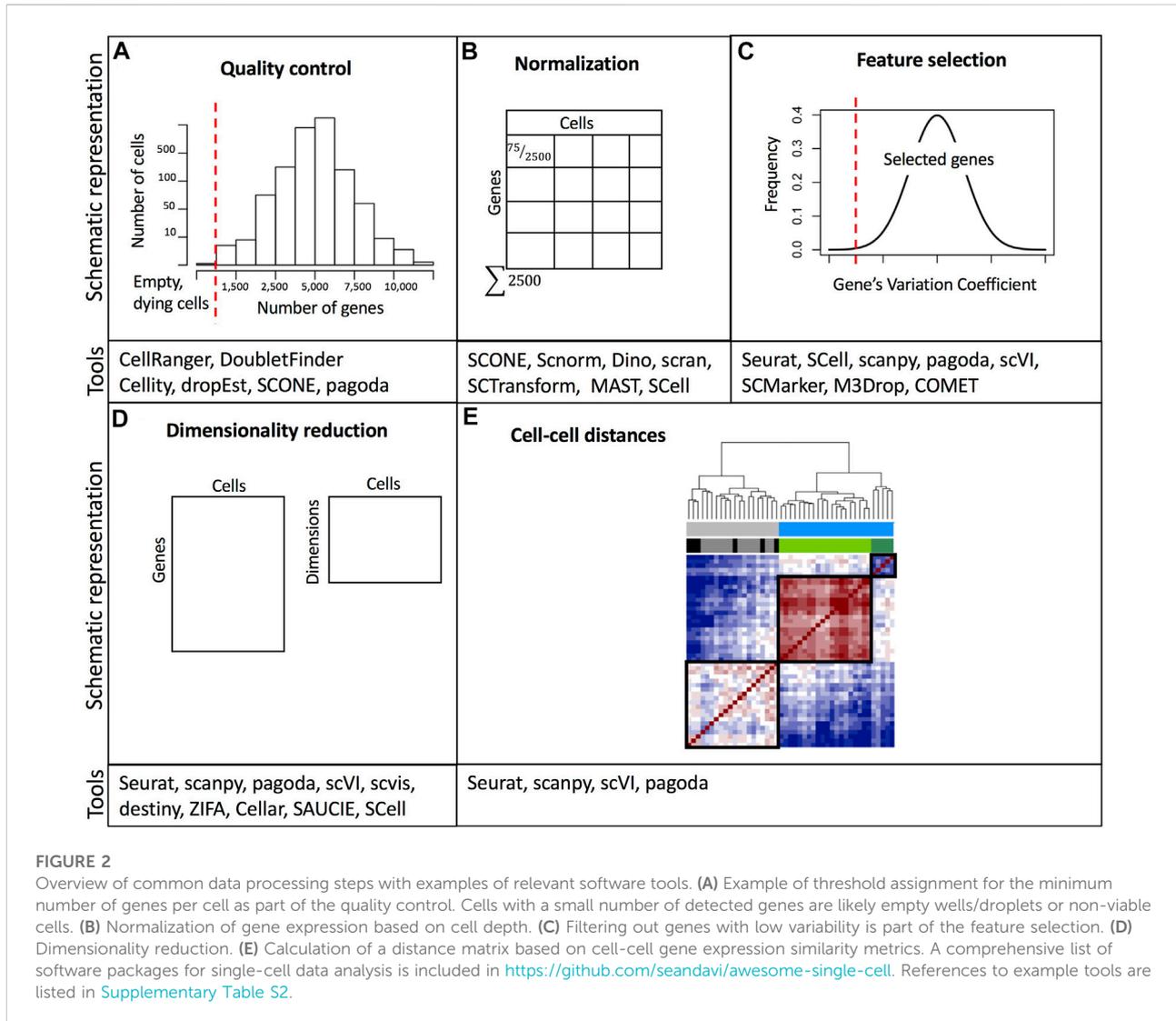
snRNA-seq is that nuclear transcripts represent only a percentage of the total cell's mRNA, as demonstrated through a study of large and small pyramidal mouse neurons (Bakken et al., 2018). Notwithstanding, numerous studies (mostly in mice) have demonstrated that snRNA-seq is comparable to scRNA-seq; for example, although more transcripts were detected in individual whole cells than nuclei isolated from the primary visual cortex of mice, the independent analysis of the scRNA-seq and snRNA-seq datasets demonstrated that the same neuronal subtypes can be similarly discriminated with both sequencing methods (Bakken et al., 2018). Similar results were observed in another study where both sequencing methods were compared using samples isolated from mouse whole brain cortex (Ding et al., 2020), suggesting that snRNA-seq is useful and reliable for cellular diversity characterization of brain tissues. Hence, this method has been used extensively to profile and classify mouse brain cells (Yao and Liu, 2021). Regarding transcriptome coverage, researchers have demonstrated that snRNA-seq of mouse neural stem cells and hippocampal neurons detects 66% and 82% of the protein coding genes respectively, while the amount of gene expression variation was similar when measured between single nuclei and single cells (Grindberg et al., 2013). Furthermore, in human organs outside the nervous system such as kidney and pancreas, single-cell and single-nucleus methods yielded equivalent gene detection sensitivity with a significant percentage of transcripts detected in scRNA-seq and not detected in snRNA-seq corresponding to mitochondrial and stress-induced genes (Wu et al., 2019; Basile et al., 2021).

Once a single-cell/nucleus suspension is obtained, a method to physically separate them is used. Two common separation methods are plate-based cell sorting and droplet-based microfluidics (Figure 1B). In plate-based methods, cells or nuclei are isolated by fluorescence activated cell sorting (FACS) or nuclei sorting (FANS) into wells (Tasic et al., 2016) where lysis and library preparation are performed. With this method, samples can be stained to enable the exclusion of dead cells or enrichment of cells labeled with specific antibodies against cell surface or intracellular markers (Baran-Gale, Chandra and Kirschner, 2018). Droplet-based methods separate cells into nanoliter-size lipid droplets, where cells are lysed and library preparation takes place (Macosko et al., 2015). Library construction consists of capturing transcripts, labeling them with cell or nuclei-specific barcodes, reverse-transcribing them into cDNA, and amplifying them. Additionally, protocols may also label captured transcripts with unique molecular identifiers (UMIs) before amplification, this approach results in more accurate quantification of counts (Zheng et al., 2017). UMIs are random oligonucleotide barcodes of a fixed length and they are used to tag original transcripts and distinguish them from

PCR duplicates (Islam et al., 2014). Droplet-based sequencing methods use lower reaction volumes and yield higher throughputs compared to plate-based methods (Forsberg et al., 2018). However, due to methodological differences, these technologies differ in the way they quantify transcripts. Droplet-based methods incorporate UMIs to either 5' or 3' ends, thus, they do not distinguish between gene isoforms. Contrastingly, plate-based methods capture the full-length transcripts including exons and splice junctions (Tasic et al., 2016; Gupta et al., 2018). Another decision to consider is the addition of spike-in RNA. Spike-ins are non-biological RNA molecules of known sequence which are added to each cell's lysate in the same known concentration (Jiang et al., 2011). Spike-ins undergo all library preparation processes and thus they are very useful for normalization under the assumption that they are present in every cell in the same amount (Stegle, Teichmann and Marioni, 2015). However, the incorporation of spike-ins is not easy. For example, the concentration of the added spike-ins must be precisely calibrated to obtain optimal results, since small variations may lead to biased estimation. Spike-ins should be considered to contribute to 1%–5% of the total number of mRNA molecules in the sample (Robinson and Oshlack 2010; Grün and van Oudenaarden 2015). Furthermore, spike-ins are prone to degradation and may be captured less efficiently than endogenous transcripts in scRNA-seq/snRNA-seq experiments. Additionally, the inclusion of spike-ins is often complicated when using droplet-based separation methods (Haque et al., 2017; Svensson et al., 2017). Therefore, careful consideration should be taken when selecting the separation method.

## Key computational steps of single-cell/single-nuclei sequencing data analysis

Sequencing datasets, whether derived from single cells or nuclei are represented by matrixes in which rows correspond to features (e.g. thousands of genes or transcripts) and columns to barcodes (cells or nuclei). Data analysis may be challenging and requires the implementation of sophisticated computational methods some of which have different statistical assumptions (Kharchenko, 2021). Moreover, computational standards for single-cell/nucleus sequencing data analysis are still lacking, whereby researchers must be very careful when selecting the appropriate computational analysis methods. Excellent reviews focused on detailing the best practices for computational data analysis and interpretation have been published (Luecken and Theis, 2019; Andrews et al., 2021; Slovin et al., 2021). In this section, we will describe some of the more general steps required for analyzing



scRNA-seq or snRNA-seq datasets (referred to as single-cell datasets because the same type of computational analysis is applied for both sequencing methods) and their downstream analysis.

## Data preprocessing

In this first data analysis step, sequenced reads are summarized into a count matrix. Generally, preprocessing pipelines include demultiplexing and barcode correction, alignment to reference genome or transcriptome, quantification, and quality control (Figure 1C). Some common preprocessing pipelines are Cell Ranger (Zheng et al., 2017) and Optimus (<https://data.humancellatlas.org/pipelines/optimus-workflow>) supporting 10x Chromium datasets, dropEst (Petukhov et al., 2018) and DropSeqTools (<https://github.com/>

[broadinstitute/Drop-seq](#)) developed for droplet-based protocols, and zUMIs (Parekh et al., 2018) compatible with all UMI-based protocols. A benchmark of common preprocessing methods is described in (You et al., 2021). Demultiplexing consists on the identification and removal of barcodes from sequencing reads. In this process, reads are grouped together by barcode similarity. However, since synthetic techniques of barcode generation are prone to deletion errors, the next process is barcode correction. Corrupted barcodes undergo an error correction by comparing their sequences to a set of known barcodes, provided in the library preparation kit (Macosko et al., 2015). If the assay was performed including UMIs, demultiplexing will attach the UMI sequence to the read name. The number of different UMIs per cell is an important metric to distinguish empty droplets/wells and outliers (Islam et al., 2014). Finally, PCR duplicates are removed, and read counts are assigned to individual genes and cells.

The last step in preprocessing is determining the quality of the captured cells/nuclei. Quality control considers three metrics: the number of counts per cell (barcode) also referred to as count depth, the number of genes per cell, and the fraction of counts per cell which mapped to mitochondrial genes (Ilicic et al., 2016; Griffiths, Scialdone and Marioni, 2018). Users assign thresholds to these metrics to filter out potentially non-viable cells or doublets (Figure 2A). It is important to note that threshold assignment for filtering cells should be selected cautiously since the behavior of these metrics may be of biological relevance. Cells with an elevated number of genes with a high number of counts are potential doublets. Doublets may occur when two or more cells/nuclei are captured in the same well or droplet. Computational methods for detecting doublets have been developed: Doublet Finder (McGinnis, Murrow and Gartner, 2019), Scrublet (Wolock, Lopez and Klein, 2019), DoubletDecon (DePasquale et al., 2019), and scds (Bais and Kostka, 2020). Another relevant aspect of quality control unique to droplet-based methods is the probability that some droplets will not be able to capture a viable cell or nucleus, but rather capture ambient RNA that was present in the cell suspension (Hong et al., 2022). These empty droplets must be filtered out for downstream analysis and some tools such as EmptyDrops or scCB2 have been developed to discriminate real cells from background barcodes (Lun et al., 2019; Ni et al., 2020). Finally, it is also important to consider that ambient RNA can also be present in droplets containing cells and can be detected in combination with endogenous cellular RNA, thus contaminating the single-cell dataset (Hong et al., 2022). For this purpose, computational tools including DecontX and SoupX can be used to estimate contamination levels and to separate endogenous RNA reads from ambient RNA reads (Yang et al., 2020; Young and Behjati, 2020).

## Normalization

Variations in cell lysis, reverse transcription efficiency, and stochastic molecular sampling during sequencing may generate differences in identical cells' sequencing depths (number of detected genes per cell) (Hicks et al., 2018). Furthermore, single-cell datasets suffer from increased sparsity, which means a high proportion of zero read counts derived from both biological and technical reasons (Kharchenko, Silberstein and Scadden, 2014; Lun, Bach and Marioni, 2016). Thus, normalization methods aim at removing technical biases while preserving real biological variation. After normalization accurate relative gene expression abundances between cells are obtained and gene counts are comparable between cells (Figure 2B). There are two broad categories of normalization methods depending on whether spike-ins were added during library preparation. Data sets with spike-ins are generally normalized using the counts of these RNAs as a reference to scale cell counts assuming that the

same number of these molecules was added to each cell. However, adding the same quantity of spike-ins to all cells is technically challenging, especially in droplet-based protocols. Thus, other normalization methods have been implemented. These other methods rely on the assumption that the majority of genes are not differentially expressed. In this sense, a commonly used global linear normalization method adopted from bulk RNA-seq analysis is count depth scaling. This method calculates counts per million (CPM) assuming that all sampled cells have an equal number of mRNA molecules. Nevertheless, this is insufficient and may bias comparisons when differentially expressed genes (DEGs) are present (Robinson and Oshlack, 2010). Instead, methods based on count ratios between cells are more robust to DEGs, for example, the median of ratios and the trimmed mean of M values (TMM) implemented in DESeq and EdgeR respectively (Robinson and Oshlack, 2010; Love, Huber and Anders, 2014).

Algorithms based on quantile normalization define scaling factors based on each cell's count distribution (e.g. upper quartile) or by fitting to a reference distribution (e.g. full quartile). Still, these methods are biased due to the high proportion of zero counts affecting the count ratios and gene expression distribution (Vallejos et al., 2017). A method developed specifically for single-cell RNA-seq was proposed by Lun et al. to account for cellular heterogeneity and zero counts (Lun, Bach and Marioni, 2016). In this method, the authors used a strategy where expression values are summed across pools of cells for normalization, then cell-specific size factors are deconvolved from pool factors. This method is implemented in the *scrn* R library (<http://bioconductor.org/packages/scrn>) and it has proven to yield more accurate scaling factors than others because it reduces the incidence of zeros by counts summation across pooled cells (Lun, Bach and Marioni, 2016; Vieth et al., 2019). Other normalization methods not relying on global scaling factors have been proposed, for example, SCnorm (Bacher et al., 2017). In this approach, zero values are filtered out and then two stages of quantile regressions are used for normalization, one to group genes based on their dependence on sequencing depth and the other to estimate scale factors within each group (Bacher et al., 2017). However, it is important to mention that removing or reducing zeros from datasets may lead to biases (especially where a low-abundance gene is expressed in many cells), thereby ignoring potential information (Linderman et al., 2022). For this reason, some authors have proposed to evaluate relative gene expression by determining the percentage of cells with non-zero expression for a specific gene instead of using normalized gene counts where zeros were reduced or removed from the dataset (Booeshaghi and Pachter, 2021).

Non-linear normalization methods have also been developed to account for the multiple sources of variation in single-cell data (Cole et al., 2019). The majority of these algorithms attempt to model read counts according to a parametric distribution.

Technical covariates such as read depth and counts per gene can be used to adjust the model. A commonly used non-linear normalization method was implemented in the R package SCTransform (Hafemeister and Satija, 2019). In this approach, authors fit cell counts to a negative binomial distribution and use sequencing depth as a covariate in their linear regression model. SCTransform outperforms global scaling methods and it has recently been included in Seurat, an R toolkit for single-cell genomics (<https://satijalab.org/seurat/>) (Satija et al., 2015). Another recently developed method based on negative binomial distribution is the R package Dino (Brown et al., 2021). In this method (specially designed for UMI-based protocols), normalization is performed by correcting the full expression distribution of each gene for library size dependent variation, rather than only correcting mean expression as most of the existing methods do, resulting in a robust approach for sample heterogeneity maintenance (Brown et al., 2021).

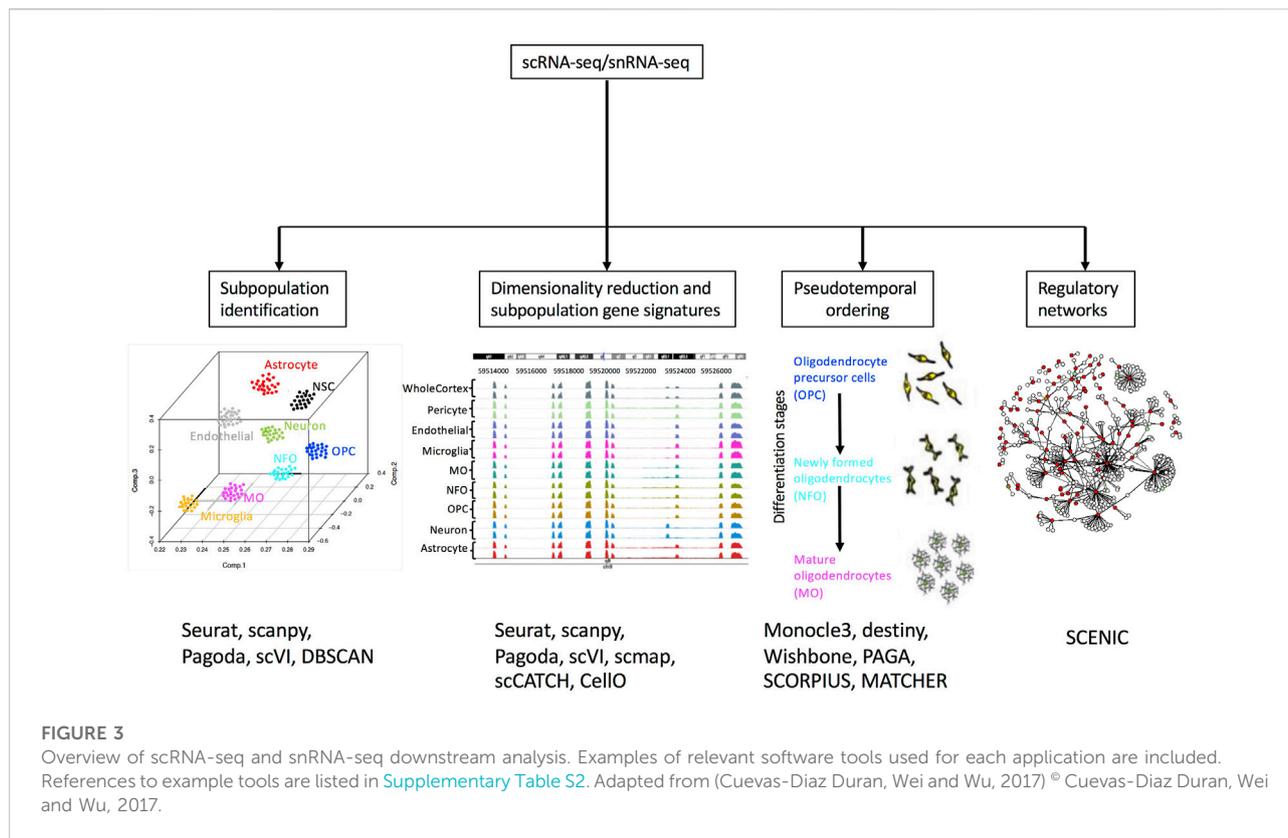
The distribution of gene counts is dramatically skewed with both an elevated fraction of zero counts and a reduced number of genes with high read counts. Thus, normalizing for gene counts (z-score) aims at improving gene comparisons within and between cells placing them in a similar scale. After normalization, a common practice is to transform data counts using a  $\log(x+1)$  function. Although data counts do not follow a log-normal distribution (Vieth et al., 2017), this transformation is useful for downstream analysis, for example, differential expression. However, it has been demonstrated that log-transformation may introduce systematic errors leading to spurious differential expression effects (Lun, 2018). These artifacts are more dramatic when there are big differences between cells' size factors and the read depth is low. Therefore, caution is suggested when interpreting clusters and trajectories derived from datasets with these conditions. Furthermore, it is also relevant to mention that a growing number of computational methods such as MAGIC, SAVER, or kNN-smoothing (van Dijk et al., 2018; Huang et al., 2018; Almanjahie et al., 2021) have been proposed to resolve the increased sparsity observed in single-cell datasets by imputing data (thereby named as imputation methods) to values that are missing or unobservable, thus improving the analysis of datasets with an elevated fraction of zero counts (Patrino et al., 2021). Some imputation methods directly address the sparsity of the single-cell datasets by using probabilistic models to distinguish biological from technical zeros and then adding values only to the technical ones. Meanwhile, other methods adjust all zero and non-zero values by smoothing or diffusing gene expression values in cells with similar expression profiles (Hou et al., 2020). Albeit imputation methods represent a potential alternative for single-cell analysis, they have the challenge of imputing data accurately while preserving true biological zeros. Also, although to date there are no quantitative benchmarks to evaluate such methods, recent studies has shown that some imputation methods can outperform non-imputation methods

in recovering gene expression, however, imputation methods did not improve performance in downstream analysis, particularly in clustering and trajectory analysis, therefore these methods should be used with caution (Hou et al., 2020; Patrino et al., 2021; Linderman et al., 2022).

Typically, toolkits implement more than one normalization method. For example, Seurat's LogNormalize function, uses a global-scaling normalization method by dividing counts for each cell by the total counts for that cell, multiplying by a scale factor (10,000 by default), and transforming data with a natural-log function (Satija et al., 2015). An overview of common normalization methods, their characteristics and limitations are depicted in Supplementary Table S1. The selection of the normalization method to use should be considered carefully since it highly impacts downstream analysis.

## Feature selection, dimensionality reduction, and visualization

Single-cell datasets have a high number of features (genes or transcripts). Many of these features or genes contain zero counts ("dropouts") for the majority of cells or their count values are not informative towards explaining the biological variability. Dropout events do not signify a lack of expression, instead, they potentially represent a failure in the detection of a transcript (Qiu, 2020). Therefore, in the feature selection step, the researcher must decide which genes to keep (Figure 2C). Generally, the top most variable genes (1,000–5,000) are the ones that may be useful in explaining cell-to-cell variability (Luecken and Theis, 2019; Yip, Sham and Wang, 2019). After selecting the most variable features, the dimensionality of expression matrices (number of rows or genes) can further be reduced from thousands to less than 100 using linear or non-linear feature projections (Figure 2D). Several dimensionality reduction techniques are available, however, a popular one is principal component analysis (PCA). PCA performs linear combinations of genes that maximize the residual variance of such combinations. The top principal components (PC) explain the majority of the data variation. Each PC, also referred to as "metagene" is a linear combination of several genes. PCA is useful for downstream analysis (e.g., clustering), however since single-cell data is inherently non-linear, PCA is not the best option for data visualization, even though it is very efficient in reducing dimensions. PCA depicts limitations derived from the fact that the computation of the PCs is not related to the underlying statistical structure of single-cell data. Since single cells can have more zero read counts than others (due to technical factors), PCA may identify this difference as one of the top PCs (Kharchenko, 2021). Instead, manifold learning methods are preferred for dimensionality reduction and data visualization. A manifold is a topological or geometric space that locally resembles a Euclidean space. Moreover, the manifold



hypothesis states that high-dimensional data lie in low-dimensional manifolds embedded in high-dimensional space (Fefferman, Mitter and Narayanan, 2016). Two commonly used manifold-learning methods for dimensionality reduction and visualization include t-distributed stochastic neighbor embedding (t-SNE) (Bonin-Font, Ortiz and Oliver, 2008) and Uniform Approximation and Projection Method (UMAP) (McInnes et al., 2018). In t-SNE, high-dimensional data are mapped into two dimensions allowing neighboring similar cells to remain close and distant cells to remain far in the low-dimensional space. t-SNE requires setting the parameter denominated “perplexity” to control the width of the Gaussian function used to determine the similarity between cells. Differing perplexity parameters may yield different number of clusters (Wattenberg, Viégas and Johnson, 2016). A caveat of t-SNE is that it fails to capture the global geometry of the data, fragmenting natural progressions or trajectories (Kobak and Berens, 2019). This is especially problematic when single-cell datasets were obtained from cell classes with a meaningful hierarchy, e.g. progenitor and differentiated cell subpopulations, and induction time points. Another commonly used alternative is the UMAP. UMAPs create a fuzzy graph from the data matrix to reflect its topology and then build a low-dimensional graph using the weight of the edges. UMAP has been demonstrated to provide the fastest run times as

well as the ability to process large numbers of cells (Becht et al., 2018), and it is recommended as a best practice for exploratory data visualization since some studies have shown that it better preserves the global structure of the data (Becht et al., 2018; Luecken and Theis, 2019), albeit other studies indicate that UMAP has the same performance as t-SNE in preserving the global structure, and is only superior to this latter method in the initial implementations for data embedding (Kobak and Linderman, 2021). Finally, it is important to emphasize that both t-SNE and UMAP are only used for dimension reduction and data visualization and not for downstream analysis such as clustering.

## Clustering

One of the most popular applications of single-cell transcriptomics is clustering or grouping cells based on transcriptional similarity (Figure 3). Clustering allows researchers to determine the number of subpopulations with distinct molecular signatures in a sample. Furthermore, based on expression profiles, researchers can infer cluster identities and arrange them in a hierarchy (Cuevas-Diaz Duran et al., 2017; Zeng and Sanes, 2017). Grouping cells based on similarity is performed using two approaches: clustering algorithms and

community detection methods. Clustering algorithms group cells based on a distance metric (Figure 2E), however, due to the high dimensionality of data, distances between cells can be very similar. This is known as the “curse of dimensionality” (Bellman, 1961) and it may be overcome by feature selection and dimensionality reduction. Since t-SNE and UMAP methods highly compress multidimensional data in two dimensions, they are not recommended for clustering analysis. Moreover, it has been suggested that t-SNE and UMAP may introduce significant distortion to the data (Chari et al., 2021), although efforts have been made to combine PCA with t-SNE to make it more suitable for clustering (Kobak and Berens, 2019). Therefore, PCA or some of its variants like weighted PCA (WPCA) or SpatialPCA are highly used for dimensionality reduction prior to clustering (Tsuyuzaki et al., 2020; Liu et al., 2022). Clustering algorithms, based on unsupervised machine learning methods use a distance calculation, for example, Euclidean, cosine similarity, or Pearson’s and Spearman’s correlation, to determine intracluster distances. Cells are assigned to clusters where similarity distances are minimized. A common clustering method used is *k*-means, which iteratively assigns cells into *k* clusters where the distance from a cell to a cluster centroid is the lowest. Numerous applications of *k*-means clustering have been developed with different distance metrics (Grün et al., 2015; Kiselev et al., 2017; Hicks et al., 2021). *K*-means clustering requires the number of clusters as input, thus users may need to run several scenarios and determine the best choice. A shortcoming of *k*-means clustering methods is that they tend to identify round clusters of equal size, missing out on rare cell subpopulations (Kiselev, Andrews and Hemberg, 2019). To overcome the drawback of *k*-means clustering, community detection algorithms have been implemented successfully.

Community detection algorithms rely on representing single-cell data as graphs with nodes and edges. These methods construct *K*-nearest neighbor graphs (KNN) from a low-dimensional space. Cells are represented as nodes and they are connected by edges to their *K* most similar nodes using a distance metric, typically Euclidean. Densely connected nodes are identified as clusters through community detection methods, which are faster than other clustering methods and can scale up to millions of cells (Rosvall and Bergstrom, 2008). Community detection-based methods outperform *k*-means clustering methods in large-scale datasets, and one of the most commonly used is the Louvain community detection algorithm (Blondel et al., 2008). This method identifies communities under the premise that cells in each group will have more links between them than what is expected from the total number of links. In this way, the algorithm optimizes a modularity function (number of links) by iteratively assigning nodes to different communities (Blondel et al., 2008). Louvain community detection is implemented in SCANPY (Wolf, Angerer and Theis, 2018) and Seurat toolkit (Satija et al., 2015). Although the Louvain algorithm is very popular, some

authors have identified flaws, resulting in badly connected communities, so the use of the Leiden algorithm has been proposed as an alternative (Traag et al., 2019). In contrast to Louvain, the Leiden algorithm can split clusters instead of only merging them; furthermore, by relying on a fast local move approach, the Leiden algorithm runs faster than the Louvain algorithm (Anuar et al., 2021) and it is also included in Seurat toolkit (Satija et al., 2015).

Density-based clustering methods identify cell clusters by recognizing contiguous regions of high-density of cells. These methods can identify cell clusters with arbitrary shapes and sizes. The most popular density-based clustering method is Density Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996). These methods do not need an *a priori* number of clusters, however, density parameters must be provided. A caveat of these methods is that they assume that all clusters have similar densities. Nonetheless, DBSCAN was included in early versions of Seurat (Satija et al., 2015) and GiniClust (Jiang et al., 2016) for the detection of rare cell types.

Alternative approaches based on matrix decomposition have also been proposed, for example, Nonnegative Matrix Factorization (NMF). NMF has conventionally been used to decompose high-dimensional transcriptional data (genes) into interpretable features (meta-genes). Nevertheless, NMF has been successfully implemented to find clusters of cells from a samples-by-genes matrix (Shao and Höfer, 2017; Duren et al., 2018; Kotliar et al., 2019; Wu et al., 2020).

## Cluster characterization

Cluster annotation consists of finding a set of genes denoted as the “gene signature” for each cluster (Figure 3). It is important to note that the transcriptional differences found do not always correspond to cell types since other variables such as cell-state and cell-cycle can have a greater influence on transcription (Buettner et al., 2015). Furthermore, differences in the cellular state may cause cells of the same type to be assigned to different clusters. Thus, interpreting cell clusters is not trivial and researchers must rely on external reference databases, for example, the Human Cell Atlas (Regev et al., 2017) or, on the expression of marker genes derived from literature. DEGs may be obtained by comparing the gene expression of cells belonging to a cluster to all other cells. Due to the nature of unsupervised clustering methods, clusters will always depict DEGs (Luecken and Theis, 2019). Upregulated genes are ranked using statistical tests such as t-tests or Wilcoxon rank-sum test and the top-ranking genes are referred to as “marker genes” (genes that have a high expression only in one cluster). Then, these marker genes are manually compared to available information in the literature or in reference databases to annotate cell type labels for each detected cluster (Zhang et al., 2019). Nevertheless, the manual characterization of clusters is prone to error since marker genes

are often expressed in multiple clusters and correspond to multiple cell types. In addition, negative marker genes should also be considered for the annotation of cell types (Ianevski et al., 2022). In this context, other approaches for cluster characterization rely on an automatic annotation of cell types (Nguyen and Griss, 2022). These computational methods are based on different principles. For example, scmap is a method that assigns cell type labels to clusters by projecting cells to cell type landmarks, then inferring unknown cells by proximity to already known cell types in the embedded space (Kiselev et al., 2018). Another approach is to correlate marker genes in annotated clusters of cells with unannotated clusters, this is implemented in the methods scCATCH (Shao et al., 2020), CIPR (Ekiz et al., 2020), ScType (Ianevski et al., 2022) and clustifyr (Fu et al., 2020). Meanwhile, other methods such as Cello use machine learning for cell type classification of cell clusters by considering the rich hierarchical structure of known cell types (Bernstein et al., 2020). A caveat on the use of marker genes in automated methods is that performance heavily relies on gene lists provided as markers for specific cell types that in some cases were manually constructed, or on marker databases that are still suboptimal both in coverage and specificity (Ianevski et al., 2022). More extensive reviews about methods for cluster characterization and cell type annotation providing guidelines and recommendations for their use are available (Clarke et al., 2021; Pasquini et al., 2021; Nguyen and Griss, 2022).

Finally, it is relevant to mention that DEGs are also used for functional enrichment analysis to identify biologically relevant functions. Importantly, marker genes representing each detected cluster should be used for experimental validation, for example, RT-qPCR, cytometry, or *in situ* imaging.

## Trajectory inference

An application of single-cell transcriptomics is trajectory inference. These computational methods are based on the premise that single-cell data is a snapshot of a continuous biological process that drives the observed heterogeneity (Trapnell, 2015). Cells are assigned a pseudo time or numerical value that measures the distance of a cell in a dynamic biological process, for example, a differentiation process (Cannoodt, Saelens and Saeys, 2016). Transition stages can be defined when cells are ordered according to their pseudo time, this is known as pseudo temporal ordering (Figure 3). Trajectory inference methods work by implementing a dimensionality reduction and a trajectory modeling step (Cannoodt, Saelens and Saeys, 2016). Trajectory modeling uses a graph representation of the data and finds a path that connects nodes (cells or groups of cells) (Moon et al., 2018). Different path-finding algorithms have been implemented and several of them rely on knowing *a priori* the location of the “root cell” (Bendall et al., 2014; Shin

et al., 2015; Matsumoto and Kiryu, 2016; Setty et al., 2016; Welch, Hartemink and Prins, 2016). There are also methods that do not require *a priori* knowledge and generate the longest path connecting all nodes in the graph (Trapnell et al., 2014; Ji and Ji, 2016). The resulting graphs can be represented using linear trajectories, showing for example the differentiation trajectory of a progenitor cell towards an intermediate state, and then ending in a differentiated stage (Deconinck et al., 2021). Methods such as SCORPIUS (Cannoodt et al., 2016) and MATCHER (Welch et al., 2017) are specific to linear trajectories. Likewise, cyclical trajectories (for cell cycle modeling, for example) can be inferred from methods like ELPiGraph (Albergante et al., 2020). Nevertheless, cell differentiation in some developmental events is not linear, and may include branching points (Deconinck et al., 2021); therefore, Slingshot (Street et al., 2018) and Monocle (Trapnell et al., 2014) are commonly used methods for modeling this type of events. Meanwhile, recent methods such as PAGA (Wolf et al., 2019) or TinGa (Todorov et al., 2020) are able to extend the modeling trajectory to allow the inclusion of loops or even separated trajectories. Finally, other approaches that model developmental processes in a probabilistic manner to infer trajectories are also available. One of them is Ouija, an approach that uses Bayesian probability to learn pseudo time from a small set of known or suspected marker genes (Campbell and Yau, 2019). SCORPIUS, MATCHER, ELPiGraph, Slingshot, Monocle, PAGA, and Ouija methods are included in the collection of R packages dynverse (<https://dynverse.org/>), which is also useful for comparing trajectory inference methods to select the most suitable method (Saelens et al., 2019).

## Kinetics of transcription

Additional information can be obtained from cells' mRNA lifecycle kinetics to infer the state of a cell before and after the time of measurement. Researchers have demonstrated in bulk RNA-seq that transcriptional data can be used to quantify the relative abundance of mRNA molecules at different lifecycle stages (Zeisel et al., 2011). Authors applied this idea to scRNA-seq in an mRNA kinetic model called RNA velocity, where the time derivative of gene expression is calculated (La Manno et al., 2018). RNA velocity distinguishes between nascent pre-mRNAs (unspliced) and mature mRNAs (spliced) from single-cell datasets. Thus, RNA velocity can quantify cellular transitions and reveal transient cellular states in a heterogeneous population. Moreover, it has been demonstrated that RNA velocity information can be also useful for trajectory inference (Zhang and Zhang, 2021). An overview of the challenges of RNA velocity modeling is discussed in (Bergen et al., 2021).

## Gene regulatory networks

Numerous computational methods have been developed to infer gene regulatory networks (GRN) using single-cell datasets (Pratapa et al., 2020). GRN have been used to identify cell types and states (Moignard et al., 2015; Aibar et al., 2017; Chan, Stumpf and Babbitt, 2017; van Dijk et al., 2018). The underlying principle of GRN is that the coordinated function of a limited number of transcription factors and co-factors regulates numerous downstream targets determining the transcriptional state of a cell. Graphs are used for the visual representation of GRNs where nodes are genes and edges are interactions (activation or inhibition). These graphs are undirected since edges are bidirectional. However, given the inherent stochastic variations (e.g. transcriptional bursting), sparsity of gene expression at single-cell resolution, drop-outs, and technical variations, deriving GRN from single-cell datasets is challenging (Raj and van Oudenaarden, 2008; Grün, Kester and van Oudenaarden, 2014; Stegle, Teichmann and Marioni, 2015). One of the main approaches of these methods is inferring networks through gene co-expression. For example, Single-Cell rEgulatory Network Inference and Clustering (SCENIC) finds modules of genes co-expressed with transcription factors and filters them to retain only those modules in which target genes have significant motif enrichment (Aibar et al., 2017). Pratapa et al. recently performed a systematic evaluation of 12 algorithms for inferring GRNs using synthetic networks, datasets from curated models from literature, and datasets from experimental single-cell transcriptomics (Pratapa et al., 2020). The authors published a series of recommendations and implemented a framework (<https://github.com/murali-group/BEELINE>) to evaluate the selected GRN methods.

## Single-cell/nuclei sequencing for understanding Alzheimer's disease

Alzheimer's disease (AD) is the most common cause of dementia worldwide with a prevalence expected to reach 113 million by 2050 (Wu et al., 2017). AD is clinically characterized by a progressive decline in memory, language, problem-solving and cognitive skills (Dubois et al., 2010). The core pathological hallmarks of AD are the extracellular deposition of  $\beta$ -amyloid (A $\beta$ ) plaques and the accumulation of neuronal fibrillary tangles (NFTs). NFTs are aggregates of Tau protein and Apolipoprotein E. The accumulation of A $\beta$  plaques and NFTs is cytotoxic and causes synapse and neuron loss. Consequently, another pathological hallmark of AD is a widespread neuronal loss specifically in the cortex and hippocampus (McKhann et al., 1984). Increasing evidence suggests that neuroinflammation is also a major contributor to the pathogenesis of AD (Heneka et al., 2015). Pro-inflammatory cytokines, markers of reactive astrocytes and

activated microglia, have been found in AD brains of both postmortem humans and transgenic mouse models (Bamberger et al., 2003; Olabarria et al., 2010; Medeiros and LaFerla, 2013). However, the use of anti-inflammatory drugs to treat AD has been unsuccessful mainly because both microglial activation and reactive astrogliosis are complex multi-stage processes that yield diverse phenotypes (Monterey et al., 2021). Thus, it is evident that neuroinflammation is a complex response that involves diverse cell types (neurons, astrocytes, and microglia) with different molecular alterations as well as the crosstalk between them.

Microarray and bulk RNA-seq studies of AD brains have revealed downregulation of neuronal functions and upregulation of immune responses (Heneka et al., 2015; Canter, Penney and Tsai, 2016; De Strooper and Karran, 2016; Nativio et al., 2018). Thus, interactions between the triad (neurons, astrocytes, and microglia) are an important part of the pathophysiology of AD. Other pathways found to be dysregulated are energy metabolism, synaptic transmission, myelin-axon interactions, cytoskeletal dynamics, and protein misfolding (Ginsberg et al., 2000; Colangelo et al., 2002; Blalock et al., 2004; Miller, Oldham and Geschwind, 2008). It is important to note that AD pathology starts in brain regions involved in learning, memory, perception, and emotion (entorhinal cortex, amygdala, and hippocampus) and progresses throughout the cortex (Braak and Braak, 1991, 1996). Interestingly, the hippocampus depicts regional vulnerability with CA1 pyramidal neurons more severely affected than CA2, CA3, and CA4 neighbors (Padurariu et al., 2012). Dissecting the cellular identities and correlating with their vulnerability to degeneration and their contribution to neuroinflammation is fundamental for understanding AD pathogenesis and it can open new avenues for the development of therapeutics. Here, we review recent pioneering single-cell and single-nuclei human transcriptomic studies aimed at studying the cellular heterogeneity involved in Alzheimer's disease (Table 1).

There are several limitations to performing scRNA-seq/snRNA-seq from frozen postmortem brains, for example, the difficulty of acquiring postmortem brain samples and the low quality due to the degradation of mRNA. The quality of RNA is typically evaluated using the RNA integrity number (RIN), which is the result of numerous factors, including postmortem interval (PMI: time between death and brain processing), patient's medical condition previous to death, storage conditions, as well as storage time (White et al., 2018). The mRNA degradation is tissue-specific, transcript-specific, and molecule-specific (Nagy et al., 2015; Sobue et al., 2016; Zhu et al., 2017). As expected, transcripts with a lower expression will degrade faster; this is why performing snRNA-seq on postmortem brains is more challenging than bulk RNA-seq. Studies have also demonstrated that miRNAs have a higher PMI-dependent resistance to degradation than other biomolecules (Nagy et al., 2015). There is no clear correlation between the PMI and RIN

**TABLE 1** Selected examples of recent scRNA-seq or snRNA-seq studies in AD. Prefrontal cortex (PFC), Entorhinal cortex (ECtx), Dorsolateral prefrontal cortex (DLPFC), temporal neocortex (TNC), parietal lobe (PL), and superior frontal gyrus (SFG). Oligodendrocyte precursor cells (OPC), oligodendrocytes (Oligo), astrocytes (Astro), microglia (Mgl), inhibitory neurons (InN), excitatory neurons (ExN), and endothelial cells (EC).

Source	Brain region	Cell type addressed	Condition	Number of cells (nuclei) analyzed	Subpopulations found	Single nuclei or cell	Major findings	Ref.
Frozen postmortem brain	PFC	All cells	24 controls and 24 age-matched individuals with varying degrees of AD	80,660	41 clusters: 13 ExN, 12 InN, 4 Astro, 5 Oligo, 3 OPC, 4 Mgl	Nuclei	Gene expression in ExN, InN, Astro, Oligo and Mgl is dysregulated in AD brains	<a href="#">Mathys et al. (2019)</a>
Frozen postmortem brain	ECtx	All cells	6 AD individuals and 6 sex and age-matched controls	13,214	8 clusters: MGL, Astro, neurons, Oligo, OPC, EC, hybrid, and unidentified	Nuclei	Astro, EC, and Mgl showed the highest gene expression differences between controls and AD	<a href="#">Grubman et al. (2019)</a>
Frozen postmortem brain	ECtx, SFG	All cells	10 individuals with different Braak stages (0, 2, and 6)	106,136	7 broad clusters (ExN, InN, Astro, Oligo, OPC, Mgl, and EC) with 9 and 11 specific subpopulations of ExN in the ECtx and SFG	Nuclei	Some subpopulations of ExN in the ECtx and SFG are more susceptible to degeneration	<a href="#">Leng et al. (2021)</a>
Autopsy brains and surgical resection	DLPFC and TNC	Mgl	3 controls, 4 MCI, and 10 AD individuals	16,242	9 Mgl subpopulations	Cell	Mgl subtypes are altered in AD brain	<a href="#">Olah et al. (2020)</a>
Frozen postmortem brain	PL	All cells	2 individuals with sporadic AD and 1 carrying a mutation in PSEN1	30,000	14 clusters of cells: 6 ExN, 2 InN, Astro, Oligo, OPC, EC, Mgl	Nuclei	ExN proportion is reduced in samples of sporadic AD	<a href="#">Del-Aguila et al. (2019)</a>

values obtained from postmortem brain samples. Thus, the general recommendation when requesting tissue from a brain bank is to require similar RIN values, rather than PMI.

Severely affected brain regions in AD are characterized by massive neuronal loss. Therefore, sequencing regions with a high neuronal deficit may bias cell proportions since they rely on assigning cells to clusters that might not have a correct representation. This problem was solved by Mathys et al. in one of the first studies that provided insights into the heterogeneity of AD in cortical regions with single-cell resolution ([Mathys et al., 2019](#)). Researchers profiled 80,660 nuclei isolated from prefrontal cortices obtained from 48 human postmortem brains with varying degrees of AD pathology, including controls. Interestingly, the analysis of cell profiles from all combined samples yielded cell types, markers, and cell-type proportions highly consistent with another human cortex snRNA-seq study of neurotypical controls ([Lake et al., 2018](#)). Mathys et al. found a total of 8 major transcriptional clusters (each with subclusters) including excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, microglia, oligodendrocyte precursor cells (OPC), endothelial cells, and pericytes. When comparing samples with AD-pathology against no-pathology, 1031 DEGs were found in total in different cell types ([Mathys et al., 2019](#)). DEGs specific to excitatory and inhibitory neurons were mostly downregulated whereas oligodendrocytes, astrocytes, and microglia depicted

more than 50% of upregulated DEGs. Notably, authors found transcriptional dysregulation due to AD pathology in all cell types, with sex-dependent differences. Dysregulation was highest in the early AD stages, and it was mostly cell-type-specific. Conversely, in late AD stages, upregulated DEGs were common to several cell types denoting a global response ([Mathys et al., 2019](#)).

One of the first cortical regions that depict neurofibrillary inclusions and neuronal loss at the early stages of AD is the entorhinal cortex. Grubman et al. characterized the heterogeneity underlying this early affected region using postmortem human brains ([Grubman et al., 2019](#)). Authors sequenced 13,214 nuclei isolated from entorhinal cortex tissue, dissected from 6 AD individuals and 6 sex and age-matched controls. Their results demonstrated the presence of all 6 known brain cell types (microglia, astrocytes, neurons, OPC, oligodendrocytes, and endothelial cells) of which astrocytes, endothelial cells, and microglia showed the highest gene expression differences between controls and AD. Researchers also demonstrated that specific transcription factors have alterations in opposite directions in different cell subpopulations. For example, in AD, APOE was found downregulated in OPC and upregulated in astrocytes and microglia ([Grubman et al., 2019](#)). This is an example of the advantage of using snRNA-seq over bulk RNA-seq; it is possible to detect transcripts with opposite regulation.

Damage to the parietal lobe is common in the early stages of AD, however, its role in the development of AD remains elusive. Del-Aguila et al. transcriptionally profiled single-nuclei isolated from the parietal lobes of an individual carrying a known mutation in PSEN1 gene (p.A79V) and 2 relatives with sporadic AD (Del-Aguila et al., 2019). Samples were obtained from postmortem brains and a total of 30,000 nuclei were sequenced. Authors identified and annotated 14 clusters including 6 subclasses of excitatory neurons, 2 inhibitory neurons, oligodendrocytes, astrocytes, microglia, OPC, and endothelial cells. A reduced proportion of excitatory neurons was observed after comparing the proportion of cells in each cluster between the mutation carrier sample against the sporadic AD samples (Del-Aguila et al., 2019). Furthermore, Del-Aguila et al. performed a pseudo time gene expression reconstruction using TSCAN (Ji and Ji, 2016). This method was implemented using the transcriptomic profile of microglial cells and it allowed authors to capture the sequence of activation and transition to disease-associated microglia (DAM) cells. The expression profiles of DAM cells were obtained from a previous study (Keren-Shaul et al., 2017). Interestingly, the authors found that 79 genes of the DAM markers were significantly associated with a temporal trajectory. This is the first study that analyzed DAM markers in single-nuclei microglia profiles obtained from AD brains carrying a known mutation.

A recent pioneering study adopted an interesting strategy to delineate the progression of AD in specific cell types (Leng et al., 2021). Leng et al. isolated 42,528 and 63,608 single nuclei from the entorhinal cortex and superior frontal gyrus respectively; samples were obtained from 10 human postmortem brains with varying degrees of AD-tau neuropathological progression (inferred from the Braak stage). The entorhinal cortex and superior frontal gyrus regions were selected because they are affected in the early and late stages of AD correspondingly (Braak and Braak, 1991). Leng et al. found 9 and 11 subpopulations of excitatory neurons in the entorhinal cortex and the superior frontal gyrus respectively depicting region-specific genes. Authors found subpopulations of neurons in both regions which were more vulnerable to degeneration in the early and late stages. Furthermore, marker genes potentially responsible for these selective vulnerabilities were derived and validated. For example, the relative abundance of a subpopulation of excitatory neurons in the entorhinal cortex showed a striking decrease in Braak stage 2 samples compared to Braak stage 0, suggesting their susceptibility to neurodegeneration in early stages (Leng et al., 2021). By correlating the degree of degeneration with the region, cell subpopulations, transcriptional profiles, and relative abundances, researchers were able to determine subpopulation selective vulnerabilities.

Given that mounting evidence suggests a microglial role in aging and AD pathology, Olah et al. profiled 13,368 single live cells isolated from the dorsolateral prefrontal cortices of autopsy samples from individuals with mild cognitive impairment or AD

pathology (Olah et al., 2020). Additionally, 2,874 cells used as controls were isolated from temporal neocortices of individuals undergoing intractable epilepsy surgeries. The authors used a previously published protocol involving FACS to purify microglial cells using antibodies for CD11b and CD45 (Olah et al., 2018). Single cells were isolated using droplet technology (Olah et al., 2020). Interestingly, researchers found 9 distinct microglial clusters and validated 4 of them histologically using marker genes. Since samples were obtained from individuals without dementia, the authors used a gene set enrichment approach to link microglial subclusters to diseases. One microglial cluster was found to be altered in AD and it was validated in the single nucleus RNA-seq study by Mathys et al. (Mathys et al., 2019).

## Single-cell/nuclei sequencing for understanding Parkinson's disease

Parkinson's disease (PD) is one of the most common slowly progressive neurodegenerative movement disorders but is very challenging at advanced stages. A histological hallmark of PD is the accumulation of fibrillar aggregates called Lewy bodies, enriched in  $\alpha$ -Synuclein misfolded protein (Wakabayashi et al., 2007). Another important characteristic of PD is the degeneration of dopaminergic neurons (DaNs) in the *substantia nigra pars compacta* (SNpc). Loss of these neurons leads to PD motor symptoms such as rigidity, resting tremor, slowness in movement (bradykinesia), and postural instability (Bloem, Okun and Klein, 2021).

Midbrain DaNs have important roles in the regulation of voluntary movement, reward, and emotion. These brain cells are highly heterogeneous even though they share a common neurotransmitter phenotype and lie in close proximity within the ventral midbrain. The current classification of DaNs is based on topographical features, for example, anatomical location and axonal innervation targets. Traditionally, three distinct types of midbrain DaNs are considered: A8, A9, and A10 located in the retrorubral field (RRF), SNpc, and ventral tegmental (VTA) areas respectively (Bentivoglio and Morelli, 2005). DaNs depict heterogeneous susceptibilities to neurodegeneration, specifically to PD. A9 DaNs project their axons into the dorsal striatum through the nigrostriatal pathway, and they are involved in the control of involuntary movement. Although A9 DaNs are the primarily degenerated cell type in PD (Lees, Hardy and Revesz, 2009), other cell types such as astrocytes and microglia have also been implicated in neurodegeneration and PD pathogenesis [reviewed in (Brück et al., 2016; Booth, Hirst and Wade-Martins, 2017)]. A8 and A10 DaNs depict projections into the ventral striatum and the prefrontal cortex, and they are involved in the regulation of emotion and reward. Degeneration of A8 and A10 DaNs is associated with schizophrenia, drug addiction, and depression

**TABLE 2** Selected examples of recent scRNA-seq or snRNA-seq studies in PD. Middle frontal gyrus (MFG), and substantia nigra (SN). Oligodendrocyte precursor cells (OPC), oligodendrocytes (Oligo), astrocytes (Astro), microglia (Mgl), macrophages (Macro), inhibitory neurons (InN), excitatory neurons (ExN), endothelial cells (EC), dopaminergic neurons (DaNs), fibroblasts (Fibro), mural cells (MC), pericytes (Peri), neuroblasts (NB) and young neurons (yNeu).

Source	Brain region	Cell type addressed	Condition	Number of cells (nuclei) analyzed	Number of subpopulations found	Single cell or nuclei	Major findings	Ref.
Frozen postmortem brains	MFG and SN	All cells	5 controls	16,649	10 clusters in SN: 2 Astro, 3 Oligo, EC, Mgl, OPC, DaNs, and GABAergic neurons	Nuclei	Gene expression of DaNs and Oligo from the SN are associated to PD genetic risks	<a href="#">Agarwal et al. (2020)</a>
Frozen postmortem brains	SN	All cells	7 controls	44,274	24 clusters: 3 Astro, Fibro, MC, EC, 3 Mgl, 7 neurons including DaNs, 5 Oligo, 2 OPC	Nuclei	Cell-type-specific expression patterns in mouse SN are highly similar to human SN	<a href="#">Welch et al. (2019)</a>
Frozen postmortem brains	Mid-brain	All cells	6 idiopathic PD individuals and 5 age and sex-matched controls	41,435	12 clusters: 5 neurons (InN, ExN, GABAergic, DaNs, CADPS2 high), OPCs, Oligo, Astro, ependymal, Peri, EC, and Mgl	Nuclei	Dysregulation of gene expression is observed in neuronal and glial cells of brains with idiopathic PD	<a href="#">Smajić et al. (2021)</a>
Frozen postmortem brains	Mid-brain	All cells	8 controls and 10 age-matched and PMI-matched individuals with PD or Lewy body dementia	387,483	10 clusters of DaNs in controls, and 68 in PD including DaNs, ExN, InN, Astro, OPC, Oligo, EC/Peri, and Mgl/Macro	Nuclei	Some DaNs subtypes are more susceptible to degeneration than others	<a href="#">Kamath et al. (2022)</a>
iPSC	None	DaNs	WT and SNCA-A53T mutant iPSC differentiated into DaNs	15,325	6 clusters: 2 neuron progenitors and 4 DaNs	Cells	DaNs have different degrees of sensitivity	<a href="#">Fernandes et al. (2020)</a>
iPSC	None	TH + DaNs	3 WT and 2 PD individuals with GBA-N370S mutation	146	Not applicable	Cells	HDAC4 is the common early repressor of downregulated genes in PD	<a href="#">Lang et al. (2019)</a>
iPSC midbrain organoids	None	All cells	1 WT and a cell line with isogenic LRRK2-p.Gly2019Ser insertion	10,475	8 clusters: NB, yNeu, DaNs, non-DaNs, glia, progenitors, Peri, EC	Cells	PD-related mutations disrupt midbrain development	<a href="#">Zagare et al. (2022)</a>

(Tzschentke and Schmidt, 2000), highlighting the different functions of dopamine in specific brain regions. An important characteristic of human DaNs is that they have up to one million axon terminals, many of which reach long distances from the soma, with projections to the putamen, and the caudate nuclei, compared to neocortical neurons which have up to tens of thousands of synapses (Bolam and Pissadaki, 2012). Thus, this uniquely massive, unmyelinated axonal branching renders ventral midbrain DaNs under a high energy demand compared to neurons from other regions contributing to their vulnerability (Pissadaki and Bolam, 2013). The molecular mechanisms that underlie the phenotypic and functional differences between ventral midbrain DaNs are largely unknown. Limited access to all these brain regions makes single-cell studies experimentally challenging. Here, we review recent pioneering single-cell and single-nuclei human transcriptomic studies aimed at elucidating the cellular heterogeneity involved in Parkinson's disease (Table 2).

Researchers have contributed to profiling the cellular heterogeneity with single-cell resolution of the SN using postmortem brains of both healthy and PD individuals. One of the earliest single-nuclei transcriptomic studies of the SN was published by Agarwal et al. (Agarwal et al., 2020) who sequenced 10,706 and 5,943 nuclei from the middle frontal gyrus and SN, respectively. Researchers obtained 12 region-matched samples from 5 human postmortem brains without neurological disease. A transcriptomic cellular atlas was compiled from the identification of cell-type-specific gene expression patterns in 10 distinct cell subpopulations found in SN: 2 types of astrocytes, 3 subtypes of oligodendrocytes, endothelial cells, microglia cells, oligodendrocyte precursor cells, DaNs, and GABAergic neurons. Significant associations were found between PD genetic risks (obtained from GWAS) and specific SN subpopulation gene profiles (DaNs and oligodendrocytes) (Agarwal et al., 2020).

In another study, researchers developed a computational method called LIGER (linked inference of genomic experimental relationships) to enable the combination and

analysis of single-cell datasets from different individuals, species, regions, conditions, and molecular origin (genomic, epigenomic, and spatial) (Welch et al., 2019). To test their method, Welch et al. generated snRNA-seq from 44,274 single-nuclei isolated from the SN of 7 frozen postmortem brains (healthy controls). A total of 24 cell clusters were identified: 3 subtypes of astrocytes, fibroblasts, mural cells, endothelial cells, 3 microglia types, 7 neuron classes including DaNs, 5 oligodendrocytes, and 2 oligodendrocyte precursor cell subtypes (Welch et al., 2019). Next, the authors compared their SN dataset with a published single-cell dataset obtained from the SN of healthy mice (Saunders et al., 2018) and found a high correlation in the cell-type-specific expression patterns between these species. Gene ontology (GO) analysis of highly correlated human-mouse gene pairs yielded significantly enriched gene sets related to brain cell identity and molecular functions (Welch et al., 2019). Conversely, gene pairs with the lowest correlation were enriched in DNA repair and chromatin remodeling functions, suggesting species differences in epigenetic regulatory functions (Welch et al., 2019). Although this comparison between species was performed on healthy individuals, the approach proposed by this study can be implemented in future studies to compare samples obtained from rodent models of PD with samples obtained from PD human patients (Duty and Jenner, 2011). Moreover, regarding the comparison between species, Geirsdottir et al. demonstrated through single-cell analysis and by using ortholog conjectures that human microglial cells depict significant heterogeneity when compared among other vertebrate species, and in addition, microglia-specific gene expression profiles depicted significant susceptibility for PD and AD (derived from human GWAS) in primates and humans (Geirsdottir et al., 2019).

Additionally, in a recent study, researchers profiled 41,435 single nuclei from postmortem midbrains of 6 individuals diagnosed with idiopathic PD (IPD) and 5 age and sex-matched controls (Smajić et al., 2021). In this study Smajić et al. demonstrated that the dysregulation in IPD is not exclusive to DaNs of the SN, but also different cell types show alterations in other brain regions. Authors found a neuronal cell cluster present only in IPD midbrains, concomitant to reduced numbers of oligodendrocytes in IPD samples. Moreover, an increased microglial subpopulation of the SN in IPD depicted an amoeboid shape, suggesting an activated state (Smajić et al., 2021).

Another comprehensive study recently available was performed by Kamath et al., 2021, (Kamath et al., 2022). This is the first study that profiled hundreds of thousands of nuclei (184,673 and 202,810) obtained from postmortem midbrains of both controls and individuals diagnosed with PD or Lewy body dementia (LBD). Researchers purified nuclei using fluorescence-activated nuclei sorting (FANS) and a nuclear receptor (NURR1 or NR4A2) previously found to be essential for the survival of DaNs in mice (Zetterström et al., 1997; Saunders et al.,

2018). The deep sequencing strategy and the data analysis method (LIGER) used allowed researchers to identify 10 clusters of DaNs in both controls and PD/LBD individuals. Interestingly, the authors used a method to accurately determine cell proportions thus, eliminating technical confounders, for example, batch effects and individual variations (MASC: mixed effect association of single cells), and found that one of the DaNs clusters was significantly reduced in PD samples, while another one was increased. These observations were validated through single-molecule fluorescence *in situ* hybridization (smFISH). Overall, these results confirm the existence of diverse subpopulations of DaNs with different degeneration susceptibilities.

The development of iPSCs and their application for disease modeling has resulted in powerful methodologies to study human complex pathologies. Reprogramming somatic cells from individuals carrying known PD-related mutations to iPSCs has been used to differentiate DaNs *in vitro*, generating a valuable source of cells that would otherwise be accessible only from postmortem brains, where the progression of PD is already at its endpoint. Thus, modeling PD *in vitro* through the differentiation of iPSCs to form DaNs, allows the discovery of biomarkers at the onset of disease, and facilitates drug testing (Laperle et al., 2020). However, one limitation of such studies is the high degree of heterogeneity and cellular variability. Variability has been reported in several aspects including iPSC differentiation potential, DNA methylation, transcriptional profiles, and even cell morphologies due to differences between donors, experiments, and genetic stability (Kilpinen et al., 2017; Volpato and Webber, 2020). To overcome these limitations, scRNA-seq and snRNA-seq are currently being used first, to compare the similarity of cells differentiated from iPSCs with those found *in vivo*, and then to elucidate the cellular heterogeneity at the onset of PD by using dopaminergic neurons differentiated from iPSCs. In this context, La Manno et al. demonstrated through scRNA-seq that dopaminergic neurons generated from human iPSC recapitulated key stages of *in vivo* midbrain development, retained the expression of markers genes, and conserved the cellular heterogeneity observed *in vivo* (La Manno et al., 2016).

In a recent study by Fernandes et al., researchers generated a common PD mutation (A53T in SNCA gene) in a human iPSC line (Human induced pluripotent stem cell initiative, Sanger Institute) and induced both WT and A53T-SNCA iPSC into DaNs using a widely accepted induction protocol (Fernandes et al., 2020). After 6 weeks, more than 15,000 cells were profiled from scRNA-seq, and using the expression of known marker genes, 6 clusters were found: 2 neuronal progenitors, and 4 dopaminergic neuron subpopulations. Also, authors found overlapping clusters between their clusters and those found in a study of substantia nigra from 7 human adult postmortem brains (Welch et al., 2019), particularly in clusters characterized by the positive expression of Tyrosine Hydroxylase, thus

suggesting that *in vitro* DaNs are representative of corresponding *in vivo* DaNs. Finally, induced cells were tested with cytotoxic drugs and genetic stressors and the proportion of cells in clusters changed, suggesting that clusters presented varying degrees of sensitivity (Fernandes et al., 2020).

To overcome the difficulties inherent to cellular heterogeneity of iPSC neuronal induction, Lang et al. sorted iPSC-derived DaNs using FACS with the Tyrosine Hydroxylase intracellular marker (Lang et al., 2019). The authors derived iPSCs and performed both bulk and single-cell RNA-seq from two individuals carrying a known PD mutation (GBA-N370S) and three controls. Although the number of profiled cells after sorting was reduced (146 cells), authors were able to identify DEGs. Notably, cells segregated between control and PD along the second component of the PCA plot. By combining DEGs found with bulk and scRNA-seq and clustering single cells with the Single-cell consensus clustering (SC3) method (Kiselev et al., 2017) authors obtained a core set of 60 DEGs, with the majority downregulated in PD. Using this small set of core genes, the authors implemented the Ouija method to infer single-cell pseudo times (Campbell and Yau, 2019). As a result, authors suggested a “continuous disease axis” of gene expression variation with 60 core genes mostly downregulated. The authors used ingenuity pathway analysis (IPA) (Krämer et al., 2014) to build regulatory gene networks and found that the Histone deacetylase HDAC4 was the common early repressor of the core set of genes. Pharmacological modulation of HDAC4 rescued PD-related phenotypes, including ER stress (Lang et al., 2019). These results show that even with a low number of cells, it is possible to find molecular drivers of a specific PD phenotype, through purification before sequencing, and a comprehensive data analysis using the appropriate computational methods.

Cerebral organoids have emerged as a useful 3D model for neural differentiation studies since they recapitulate certain aspects of brain development and disease-associated phenotypes (Lancaster et al., 2013). For example, by comparing neural precursors differentiated in 2D vs. 3D by scRNA-seq, it was recently established that midbrain organoids presented lower levels of cellular senescence and mitochondrial stress, which correlated with resistance to toxic challenges, robust synaptic contacts, and functionality of DaNs (Kim et al., 2021). Furthermore, another study found by single-cell analysis demonstrated that midbrain organoids maintain the same neuronal heterogeneity observed *in vivo*, indicating that organoids are a useful system for the *in vitro* study of neurodegenerative diseases (Smits et al., 2020). Thus, pathological features have been analyzed in brain organoids generated from isogenic iPSCs that carry a PD-related mutation in LRRK2. In such context, scRNA-seq experiments showed clear differences between the normal and mutated iPSCs. The LRRK2 p.Gly2019Ser mutation disrupted normal development, resulting in incomplete differentiation and

reduced viability (Zagare et al., 2022). These studies show that midbrain organoids and scRNA-seq constitute an excellent system to study different cellular and molecular aspects related to PD.

## Discussion

In the last decade, impressive progress has been achieved through the use of scRNA-seq/snRNA-seq to elucidate the transcriptional and regional heterogeneity of brain cells. These developments have transformed our understanding of brain cell diversity and their interplay in neurodegenerative disease conditions. Even though methods for scRNA-seq/snRNA-seq have been well developed, several limitations and challenges remain. One example is the source of brain cells. Fresh brain samples mainly from resection surgeries are scarce and the best alternatives are either postmortem brains or differentiated iPSCs. The integrity of the nuclear transcripts obtained from postmortem brains must be carefully determined and brain samples should have similar RIN values. Also, it is important to have access to brain donors' clinical information including the degree of neurodegeneration to adequately form groups or model covariables. A well-designed experiment including age, sex, and degree of degeneration-matched controls is very important to reduce the number of variables affecting gene expression.

Additionally, since AD and PD are characterized by progressive neurodegeneration and loss of neurons in different regions, identifying the cell subtypes whose proportions are altered in AD or PD is challenging. Researchers rely on algorithms that automatically preprocess and normalize the datasets assigning cells to clusters. However, different algorithms or parameter choices can lead to divergent results. Careful consideration must be employed when selecting which methods to use for data analysis and experimental validation may be needed. Furthermore, the preparation of single cells or nuclei in suspension is also a major challenge when sequencing transcripts obtained from brain cells or nuclei. Particularly for single-cell suspensions, it is difficult to isolate intact brain cells because they are embedded in a complex and interconnected network. Also, neurons are very sensitive to cell dissociation methods, therefore scRNA-seq datasets may show an underrepresentation of neuronal types with respect to glial cells (Darmanis et al., 2015). Therefore, since nuclei isolation protocols are faster and they represent the best alternative when the tissue is degraded, frozen or formalin-fixed (Lacar et al., 2016), snRNA-seq is an excellent alternative for profiling of brain cells. In addition, studies have demonstrated that snRNA-seq results are comparable to scRNA-seq results (Bakken et al., 2018).

Another important caveat of scRNA-seq and snRNA-seq is that by isolating cells or nuclei, their spatial information and connectivity are lost. Transcriptional profiles can convey information about a cell's identity and the interplay with

other cells, however, the information about neighboring cells or their proximity to pathologies (e.g. tau fibrillary tangles, Lewy bodies) is unknown. To address this deficiency, an emerging high-throughput alternative is spatial transcriptomics (Stahl et al., 2016). Spatial transcriptomic methods are now being used to resolve the location of cells, for example, smFISH and Slide-Seq (Raj et al., 2008; Rodrigues et al., 2019). However, it is important to consider that through spatial transcriptomics it is difficult to obtain conclusions with single-cell resolution because these methods consist of profiling spots distributed throughout a histological sample and each spot covers various cells (Noel et al., 2022). Furthermore, the analysis of spatial transcriptomics data can be challenging since detailed computational methods for their study do not exist or are still in development (Atta and Fan, 2021). Thus, in the near future, scRNA-seq/snRNA-seq technologies should be combined with other multi-omics and with spatial transcriptomics to provide a holistic understanding.

Overall, AD and PD brain scRNA-seq and snRNA-seq are providing new insights, for example, identifying cell subpopulations that are more vulnerable to degeneration and cellular transcriptional profiles specific to the affected regions. Particularly, the identification of cell subpopulations and the overall characterization of cell heterogeneity in brain regions affected by AD or PD is valuable for understanding why some neuronal or glial cell types are more susceptible to degeneration than others, as well to better understand how gene expression is regulated at the single-cell level in a pathological state of the brain. Besides, the comparison of the information obtained from single-cell datasets with data obtained, for example, from GWAS, can provide insights for the identification of specific cell subtypes or genes that may provide the risk of neurodegenerative diseases development. The construction of cell trajectories can show the progression of cellular degeneration as the disease progresses, and infer cell types that are lost while others are enriched. Thus, future advancements in this area will facilitate the discovery of potential and novel therapeutic targets.

## References

- Agarwal, D., Sandor, C., Volpato, V., Caffrey, T. M., Monzon-Sandoval, J., Bowden, R., et al. (2020). A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nat. Commun.* 11 (1), 4183. doi:10.1038/s41467-020-17876-0
- Aibar, S., Gonzalez-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., et al. (2017). Scenic: Single-cell regulatory network inference and clustering. *Nat. Methods* 14 (11), 1083–1086. doi:10.1038/nmeth.4463
- Albergante, L., Mirkes, E., Bac, J., Chen, H., Martin, A., Faure, L., et al. (2020). Robust and scalable learning of complex intrinsic dataset geometry via ELPIGraph. *Entropy* 22 (3), 296. doi:10.3390/e22030296
- Almanjahie, I., Alahmari, W., Laksaci, A., and Rachdi, M. (2021). Computational aspects of the kNN local linear smoothing for some conditional models in high dimensional statistics. *Commun. Statistics - Simul. Comput.* doi:10.1080/03610918.2021.1923745
- Andrews, T., Kiselev, V. Y., McCarthy, D., and Hemberg, M. (2021). Tutorial: Guidelines for the computational analysis of single-cell RNA sequencing data. *Nat. Protoc.* 16 (1), 1–9. doi:10.1038/s41596-020-00409-w
- Anuar, S. H. H., Abas, Z. A., Yunus, N. M., Mohd Zaki, N. H., Hashim, N. A., Mokhtar, M. F., et al. (2021). Comparison between Louvain and leiden algorithm for network structure: A review. *J. Phys. Conf. Ser.* 2129, 012028. doi:10.1088/1742-6596/2129/1/012028
- Atta, L., and Fan, J. (2021). Computational challenges and opportunities in spatially resolved transcriptomic data analysis. *Nat. Commun.* 12 (1), 5283. doi:10.1038/s41467-021-25557-9
- Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., et al. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* 513 (5), 532–541. doi:10.1002/cne.21974

## Author contributions

RC-DD and JQW contributed to the idea conception and overall review design. RC-DD and JCG-O wrote the manuscript. All authors revised and approved the final manuscript.

## Funding

This work was supported by grants from the National Institutes of Health R01 NS088353, R21EY028647-01, and Amy and Edward Knight Fund-the UTHSC Senator Lloyd Bentsen Stroke Center. Research in the laboratory of IV is supported by UNAM-PAPIIT IN219122 and Conacyt 64382.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2022.884748/full#supplementary-material>

- Bacher, R., Chu, L. F., Leng, N., Gasch, A. P., Thomson, J. A., Stewart, R. M., et al. (2017). SCnorm: Robust normalization of single-cell RNA-seq data. *Nat. Methods* 14 (6), 584–586. doi:10.1038/nmeth.4263
- Bais, A. S., and Kostka, D. (2020). scds: computational annotation of doublets in single-cell RNA sequencing data. *Bioinforma. Oxf. Engl.* 36 (4), 1150–1158. doi:10.1093/bioinformatics/btz698
- Bakken, T. E., Hodge, R. D., Miller, J. A., Yao, Z., Nguyen, T. N., Aevermann, B., et al. (2018). Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS one* 13 (12), e0209648. doi:10.1371/journal.pone.0209648
- Bamberger, M. E., Harris, M. E., McDonald, D. R., Husemann, J., and Landreth, G. E. (2003). A cell surface receptor complex for fibrillar beta-amyloid mediates microglial activation. *J. Neurosci.* 23 (7), 2665–2674. doi:10.1523/jneurosci.23-07-02665.2003
- Baran-Gale, J., Chandra, T., and Kirschner, K. (2018). Experimental design for single-cell RNA sequencing. *Brief. Funct. Genomics* 17 (4), 233–239. doi:10.1093/bfgp/elix035
- Basile, G., Kahraman, S., Dirice, E., Pan, H., Dreyfuss, J. M., and Kulkarni, R. N. (2021). Using single-nucleus RNA-sequencing to interrogate transcriptomic profiles of archived human pancreatic islets. *Genome Med.* 13 (1), 128. doi:10.1186/s13073-021-00941-8
- Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I., Ng, L. G., et al. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37, 38–44. doi:10.1038/nbt.4314
- Bellman, R. E. (1961). *Adaptive control processes: A guided tour*. Princeton University Press.
- Bendall, S. C., Davis, K. L., Amir, E. A. D., Tadmor, M. D., Simonds, E. F., Chen, T. J., et al. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 157 (3), 714–725. doi:10.1016/j.cell.2014.04.005
- Bentivoglio, M., and Morelli, M. (2005). Chapter I the organization and circuits of mesencephalic dopaminergic neurons and the distribution of dopamine receptors in the brain. *Handb. Chem. Neuroanat.* 21, 1–107. doi:10.1016/S0924-8196(05)80005-3
- Bergen, V., Soldatov, R. A., Kharchenko, P. V., and Theis, F. J. (2021). RNA velocity-current challenges and future perspectives. *Mol. Syst. Biol.* 17 (8), e10282. doi:10.15252/msb.202110282
- Bernstein, M., Ma, Z., Gleicher, M., and Dewey, C. N. (2020). CellO: Comprehensive and hierarchical cell type classification of human cells with the cell ontology. *iScience* 24 (1), 101913. doi:10.1016/j.isci.2020.101913
- Blalock, E. M., Geddes, J. W., Chen, K. C., Porter, N. M., Markesbery, W. R., and Landfield, P. W. (2004). Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc. Natl. Acad. Sci. U. S. A.* 101 (7), 2173–2178. doi:10.1073/pnas.0308512100
- Bloem, B. R., Okun, M. S., and Klein, C. (2021). Parkinson's disease. *Lancet (London, Engl.)* 397, 2284–2303. doi:10.1016/S0140-6736(21)00218-X
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008 (10), P10008. doi:10.1088/1742-5468/2008/10/P10008
- Bolam, J. P., and Pissadaki, E. K. (2012). Living on the edge with too many mouths to feed: Why dopamine neurons die. *Mov. Disord.* 27 (12), 1478–1483. doi:10.1002/mds.25135
- Bonin-Font, F., Ortiz, A., and Oliver, G. (2008). Visual navigation for mobile robots: A survey. *J. Intell. Robot. Syst.* 53 (3), 263–296. doi:10.1007/s10846-008-9235-4
- Boeshaghi, A., and Pachter, L. (2021). Normalization of single-cell RNA-seq counts by  $\log(x + 1)^*$  or  $\log(1 + x)$ . *Bioinformatics* 37 (15), 2223–2224. doi:10.1093/bioinformatics/btab085
- Booth, H. D. E., Hirst, W. D., and Wade-Martins, R. (2017). The role of astrocyte dysfunction in Parkinson's disease pathogenesis. *Trends Neurosci.* 40 (6), 358–370. doi:10.1016/j.tins.2017.04.001
- Braak, H., and Braak, E. (1996). Development of Alzheimer-related neurofibrillary changes in the neocortex inversely recapitulates cortical myelogenesis. *Acta Neuropathol.* 92 (2), 197–201. doi:10.1007/s004010050508
- Braak, H., and Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 82 (4), 239–259. doi:10.1007/BF00308809
- Brown, J., Ni, Z., Mohanty, C., Bacher, R., and Kendzierski, C. (2021). Normalization by distributional resampling of high throughput single-cell RNA-sequencing data. *Bioinformatics* 37, 4123–4128. doi:10.1093/bioinformatics/btab450
- Brück, D., Wenning, G. K., Stefanova, N., and Fellner, L. (2016). Glia and alpha-synuclein in neurodegeneration: A complex interaction. *Neurobiol. Dis.* 85, 262–274. doi:10.1016/j.nbd.2015.03.003
- Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., et al. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33 (2), 155–160. doi:10.1038/nbt.3102
- Cajigas, I. J., Tushev, G., Will, T. J., tom Dieck, S., Fuerst, N., and Schuman, E. M. (2012). The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron* 74 (3), 453–466. doi:10.1016/j.neuron.2012.02.036
- Campbell, K., and Yau, C. (2019). A descriptive marker gene approach to single-cell pseudotime inference. *Bioinformatics* 35 (1), 28–35. doi:10.1093/bioinformatics/bty498
- Cannoodt, R., Saelens, W., and Saeys, Y. (2016). Computational methods for trajectory inference from single-cell transcriptomics. *Eur. J. Immunol.* 46 (11), 2496–2506. doi:10.1002/eji.201646347
- Canter, R. G., Penney, J., and Tsai, L.-H. (2016). The road to restoring neural circuits for the treatment of Alzheimer's disease. *Nature* 539 (7628), 187–196. doi:10.1038/nature20412
- Chan, T. E., Stumpf, M. P. H., and Bachtie, A. C. (2017). Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* 5 (3), 251–267. e3. doi:10.1016/j.cels.2017.08.014
- Chari, T., Banerjee, J., and Pachter, L. (2021). The specious art of single-cell genomics. *bioRxiv (Preprint)*. doi:10.1101/2021.08.25.457696
- Chung, H., Melnikov, A., McCabe, C., Drokhyansky, E., Van Wittenberghe, N., Magee, E. M., et al. (2022). SnFFPE-Seq: Towards scalable single nucleus RNA-seq of formalin-fixed paraffin-embedded (FFPE) tissue. *bioRxiv, Prepr.* (Accessed September 22, 2022). doi:10.1101/2022.08.25.505257
- Clarke, Z., Andrews, T. S., Atif, J., Pouyababar, D., Innes, B. T., MacParland, S. A., et al. (2021). Tutorial: Guidelines for annotating single-cell transcriptomic maps using automated and manual methods. *Nat. Protoc.* 16 (6), 2749–2764. doi:10.1038/s41596-021-00534-0
- Colangelo, V., Schurr, J., Ball, M. J., Pelaez, R. P., Bazan, N. G., and Lukiw, W. J. (2002). Gene expression profiling of 12633 genes in alzheimer hippocampal CA1: Transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. *J. Neurosci. Res.* 70 (3), 462–473. doi:10.1002/jnr.10351
- Colantuoni, C., Lipska, B. K., Ye, T., Hyde, T. M., Tao, R., Leek, J. T., et al. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* 478 (7370), 519–523. doi:10.1038/nature10524
- Cole, M. B., Risso, D., Wagner, A., DeTomaso, D., Ngai, J., Purdom, E., et al. (2019). Performance assessment and selection of normalization procedures for single-cell RNA-seq. *Cell Syst.* 8 (4), 315–328. e8. doi:10.1016/j.cels.2019.03.010
- Cuevas-Diaz Duran, R., Wei, H., and Wu, J. Q. (2017). Single-cell RNA-sequencing of the brain. *Clin. Transl. Med.* 6 (1), 20. doi:10.1186/s40169-017-0150-9
- Cuomo, A., Seaton, D. D., McCarthy, D. J., Martinez, I., Bonder, M. J., Garcia-Bernardo, J., et al. (2020). Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.* 11 (1), 810. doi:10.1038/s41467-020-14457-z
- Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., et al. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U. S. A.* 112 (23), 7285–7290. doi:10.1073/pnas.1507125112
- De Strooper, B., and Karran, E. (2016). The cellular phase of Alzheimer's disease. *Cell* 164 (4), 603–615. doi:10.1016/j.cell.2015.12.056
- Deconinck, L., Cannoodt, R., Saelens, W., Deplancke, B., and Saeys, Y. (2021). Recent advances in trajectory inference from single-cell omics data. *Curr. Opin. Syst. Biol.* 27, 100344. doi:10.1016/j.coisb.2021.05.005
- Del-Aguila, J. L., Li, Z., Dube, U., Mihindukulasuriya, K. A., Budde, J. P., Fernandez, M. V., et al. (2019). A single-nuclei RNA sequencing study of Mendelian and sporadic AD in the human brain. *Alzheimers Res. Ther.* 11 (1), 71. doi:10.1186/s13195-019-0524-x
- Denisenko, E., Guo, B. B., Jones, M., Hou, R., de Kock, L., Lassmann, T., et al. (2020). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* 21 (1), 130. doi:10.1186/s13059-020-02048-6
- DePasquale, E. A. K., Schnell, D. J., Van Camp, P. J., Valiente-Alandi, I., Blaxall, B. C., Grimes, H. L., et al. (2019). DoubletDecon: Deconvoluting doublets from single-cell RNA-sequencing data. *Cell Rep.* 29 (6), 1718–1727. e8. doi:10.1016/j.celrep.2019.09.082
- Ding, J., Adiconis, X., Simmons, S. K., Kowalczyk, M. S., Hession, C. C., Marjanovic, N. D., et al. (2020). Systematic comparison of single-cell and

- single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* 38 (6), 737–746. doi:10.1038/s41587-020-0465-8
- Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., Dekosky, S. T., Barberger-Gateau, P., et al. (2010). Revising the definition of Alzheimer's disease: A new lexicon. *Lancet. Neurol.* 9 (11), 1118–1127. doi:10.1016/S1474-4422(10)70223-4
- Duren, Z., Chen, X., Zamanighomi, M., Zeng, W., Satpathy, A. T., Chang, H. Y., et al. (2018). Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations. *Proc. Natl. Acad. Sci. U. S. A.* 115 (30), 7723–7728. doi:10.1073/pnas.1805681115
- Duty, S., and Jenner, P. (2011). Animal models of Parkinson's disease: A source of novel treatments and clues to the cause of the disease. *Br. J. Pharmacol.* 164 (4), 1357–1391. doi:10.1111/j.1476-5381.2011.01426.x
- Ekiz, H., Conley, C. J., Stephens, W. Z., and O'Connell, R. M. (2020). Cipr: A web-based R/shiny app and R package to annotate cell clusters in single cell RNA sequencing experiments. *BMC Bioinforma.* 21 (1), 191. doi:10.1186/s12859-020-3538-2
- Ester, M., Kriegel, H. P., Sander, J., and Xu, X. (1996). "A density-based algorithm for discovering clusters in large spatial databases with Noise," in *proceedings of the second international conference on knowledge discovery and data mining (KDD'96)* (Munich, Germany: AAAI Press), 226–231.
- Esteve-Codina, A., Arpi, O., Martinez-Garcia, M., Pineda, E., Mallo, M., Gut, M., et al. (2017). A comparison of RNA-seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. *PLoS one* 12 (1), e0170632. doi:10.1371/journal.pone.0170632
- Fefferman, C., Mitter, S., and Narayanan, H. (2016). Testing the manifold hypothesis. *J. Amer. Math. Soc.* 29 (4), 983–1049. doi:10.1090/jams/852
- Fernandes, H. J. R., Patikas, N., Foskolou, S., Field, S. F., Park, J. E., Byrne, M. L., et al. (2020). Single-cell transcriptomics of Parkinson's disease human *in vitro* models reveals dopamine neuron-specific stress responses. *Cell Rep.* 33 (2), 108263. doi:10.1016/j.celrep.2020.108263
- Forsberg, E. M., Huan, T., Rinehart, D., Benton, H. P., Warth, B., Hilmers, B., et al. (2018). Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. *Nat. Protoc.* 13 (4), 633–651. doi:10.1038/nprot.2017.151
- Fu, R., Gillen, A. E., Sheridan, R. M., Tian, C., Daya, M., Hao, Y., et al. (2020). clustifyr: an R package for automated single-cell RNA sequencing cluster classification. *F1000Res.* 9, 223. doi:10.12688/f1000research.22969.1
- GBD (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019. *Lancet (London, Engl.)* 396 (10258), 1204–1222. doi:10.1016/S0140-6736(20)30925-9
- Geirsdottir, L., David, E., Keren-Shaul, H., Weiner, A., Bohlen, S. C., Neuber, J., et al. (2019). Cross-species single-cell analysis reveals divergence of the primate microglia program. *Cell* 179 (7), 1609–1622. e16. doi:10.1016/j.cell.2019.11.010
- Ginsberg, S. D., Hemby, S. E., Lee, V. M., Eberwine, J. H., and Trojanowski, J. Q. (2000). Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons. *Ann. Neurol.* 48 (1), 77–87. doi:10.1002/1531-8249(200007)48:1<77:aid-ana12>3.0.co;2-a
- Gradišnik, L., Bosnjak, R., Bunc, G., Ravnik, J., Maver, T., and Velnar, T. (2021). Neurosurgical approaches to brain tissue harvesting for the establishment of cell cultures in neural experimental cell models. *Mater. (Basel, Switz.)* 14 (22), 6857. doi:10.3390/ma14226857
- Griffiths, J. A., Scialdone, A., and Marioni, J. C. (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. *Mol. Syst. Biol.* 14 (4), e8046. doi:10.15252/msb.20178046
- Grindberg, R., Yee-Greenbaum, J. L., McConnell, M. J., Novotny, M., O'Shaughnessy, A. L., Lambert, G. M., et al. (2013). RNA-sequencing from single nuclei. *Proc. Natl. Acad. Sci. U. S. A.* 110 (49), 19802–19807. doi:10.1073/pnas.1319700110
- Grubman, A., Chew, G., Ouyang, J. F., Sun, G., Choo, X. Y., McLean, C., et al. (2019). A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* 22 (12), 2087–2097. doi:10.1038/s41593-019-0539-4
- Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat. Methods* 11 (6), 637–640. doi:10.1038/nmeth.2930
- Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., et al. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525 (7568), 251–255. doi:10.1038/nature14966
- Grün, D., and van Oudenaarden, A. (2015). Design and analysis of single-cell sequencing experiments. *Cell* 163 (4), 799–810. doi:10.1016/j.cell.2015.10.039
- Gupta, I., Collier, P. G., Haase, B., Mahfouz, A., Joglekar, A., Floyd, T., et al. (2018). Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* 36, 1197–1202. doi:10.1038/nbt.4259
- Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J. J., et al. (2016). Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Sci. (New York, N.Y.)* 353 (6302), 925–928. doi:10.1126/science.aad7038
- Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 20 (1), 296. doi:10.1186/s13059-019-1874-1
- Haque, A., Engel, J., Teichmann, S. A., and Lonnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* 9 (1), 75. doi:10.1186/s13073-017-0467-4
- Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489 (7416), 391–399. doi:10.1038/nature11405
- Heemels, M.-T. (2016). Neurodegenerative diseases. *Nature* 539 (7628), 179. doi:10.1038/539179a
- Heneka, M. T., Carson, M. J., El Khoury, J., Landreth, G. E., Brosseron, F., Feinstein, D. L., et al. (2015). Neuroinflammation in Alzheimer's disease. *Lancet. Neurol.* 14 (4), 388–405. doi:10.1016/S1474-4422(15)70016-5
- Hicks, S. C., Liu, R., Ni, Y., Purdom, E., and Risso, D. (2021). mbkmeans: Fast clustering for single cell data using mini-batch k-means. *PLoS Comput. Biol.* 17 (1), e1008625. doi:10.1371/journal.pcbi.1008625
- Hicks, S. C., Townes, F. W., Teng, M., and Irizarry, R. A. (2018). Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostat. Oxf. Engl.* 19 (4), 562–578. doi:10.1093/biostatistics/kxx053
- Hong, R., Koga, Y., Bandyadka, S., Leshchyk, A., Wang, Y., Akavoor, V., et al. (2022). Comprehensive generation, visualization, and reporting of quality control metrics for single-cell RNA sequencing data. *Nat. Commun.* 13 (1), 1688. doi:10.1038/s41467-022-29212-9
- Hou, W., Ji, Z., Ji, H., and Hicks, S. C. (2020). A systematic evaluation of single-cell RNA-sequencing imputation methods. *Genome Biol.* 21 (1), 218. doi:10.1186/s13059-020-02132-x
- Huang, M., Wang, J., Torre, E., Dueck, H., Shaffer, S., Bonasio, R., et al. (2018). Saver: Gene expression recovery for single-cell RNA sequencing. *Nat. Methods* 15 (7), 539–542. doi:10.1038/s41592-018-0033-z
- Ianevski, A., Giri, A. K., and Aittokallio, T. (2022). Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat. Commun.* 13 (1), 1246. doi:10.1038/s41467-022-28803-w
- Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C., et al. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* 17, 29. doi:10.1186/s13059-016-0888-1
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., et al. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11 (2), 163–166. doi:10.1038/nmeth.2772
- Ji, Z., and Ji, H. (2016). Tscan: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* 44 (13), e117. doi:10.1093/nar/gkw430
- Jiang, L., Chen, H., Pinello, L., and Yuan, G. C. (2016). GiniClust: Detecting rare cell types from single-cell gene expression data with gini index. *Genome Biol.* 17 (1), 144. doi:10.1186/s13059-016-1010-4
- Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R., Salit, M., et al. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* 21 (9), 1543–1551. doi:10.1101/gr.121095.111
- Kamath, T., Abdullaouf, A., Burris, S., Gazestani, V., Nadaf, N., Vanderburg, C., et al. (2021). bioRxiv. Available at: doi:10.1101/2021.06.16.448661A molecular census of midbrain dopaminergic neurons in Parkinson's disease448661
- Kamath, T., Abdullaouf, A., Burris, S. J., Langlieb, J., Gazestani, V., Nadaf, N. M., et al. (2022). Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease. *Nat. Neurosci.* 25 (5), 588–595. doi:10.1038/s41593-022-01061-1
- Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T. K., et al. (2017). A unique microglia type Associated with restricting development of Alzheimer's disease. *Cell* 169 (7), 1276–1290. e17. doi:10.1016/j.cell.2017.05.018
- Kharchenko, P. V., Silberstein, L., and Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* 11 (7), 740–742. doi:10.1038/nmeth.2967
- Kharchenko, P. V. (2021). The triumphs and limitations of computational methods for scRNA-seq. *Nat. Methods* 18 (7), 723–732. doi:10.1038/s41592-021-01171-x

- Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., et al. (2017). Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* 546 (7658), 370–375. doi:10.1038/nature22403
- Kim, S. W., Woo, H. J., Kim, E. H., Kim, H. S., Suh, H. N., Kim, S. H., et al. (2021). Neural stem cells derived from human midbrain organoids as a stable source for treating Parkinson's disease: Midbrain organoid-NSCs (Og-NSC) as a stable source for PD treatment. *Prog. Neurobiol.* 204, 102086. doi:10.1016/j.pneurobio.2021.102086
- Kiselev, V. Y., Andrews, T. S., and Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* 20 (5), 273–282. doi:10.1038/s41576-018-0088-9
- Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., et al. (2017). SC3: Consensus clustering of single-cell RNA-seq data. *Nat. Methods* 14 (5), 483–486. doi:10.1038/nmeth.4236
- Kiselev, V. Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* 15 (5), 359–362. doi:10.1038/nmeth.4644
- Kobak, D., and Berens, P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* 10 (1), 5416. doi:10.1038/s41467-019-13056-x
- Kobak, D., and Linderman, G. C. (2021). Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nat. Biotechnol.* 39 (2), 156–157. doi:10.1038/s41587-020-00809-z
- Kotliar, D., Veres, A., Nagy, M. A., Tabrizi, S., Hodis, E., Melton, D. A., et al. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* 8, e43803. doi:10.7554/eLife.43803
- Krämer, A., Green, J., Pollard, J., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30 (4), 523–530. doi:10.1093/bioinformatics/btt703
- Krishnaswami, S. R., Grindberg, R. V., Novotny, M., Venepally, P., Lacar, B., Bhutani, K., et al. (2016). Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* 11 (3), 499–524. doi:10.1038/nprot.2016.015
- La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., et al. (2016). Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* 167 (2), 566–580. e19. doi:10.1016/j.cell.2016.09.027
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., et al. (2018). RNA velocity of single cells. *Nature* 560 (7719), 494–498. doi:10.1038/s41586-018-0414-6
- Lacar, B., Linker, S. B., Jaeger, B. N., Krishnaswami, S. R., Barron, J. J., Kelder, M. J. E., et al. (2016). Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat. Commun.* 7, 11022. doi:10.1038/ncomms11022
- Lake, B. B., Chen, S., Sos, B. C., Fan, J., Kaeser, G. E., Yung, Y. C., et al. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* 36 (1), 70–80. doi:10.1038/nbt.4038
- Lancaster, M. A., Renner, M., Martin, C. A., Wenzel, D., Bicknell, L. S., Hurles, M. E., et al. (2013). Cerebral organoids model human brain development and microcephaly. *Nature* 501 (7467), 373–379. doi:10.1038/nature12517
- Lang, C., Campbell, K. R., Ryan, B. J., Carling, P., Attar, M., Vowles, J., et al. (2019). Single-cell sequencing of iPSC-dopamine neurons reconstructs disease progression and identifies HDAC4 as a regulator of Parkinson cell phenotypes. *Cell Stem Cell* 24 (1), 93–106. e6. doi:10.1016/j.stem.2018.10.023
- Laperle, A. H., SanceS, S., Yucer, N., Dardov, V. J., Garcia, V. J., Ho, R., et al. (2020). iPSC modeling of young-onset Parkinson's disease reveals a molecular signature of disease and novel therapeutic candidates. *Nat. Med.* 26 (2), 289–299. doi:10.1038/s41591-019-0739-1
- Lees, A. J., Hardy, J., and Revesz, T. (2009). Parkinson's disease. *Lancet (London, Engl.)* 373 (9680), 2055–2066. doi:10.1016/S0140-6736(09)60492-X
- Leng, K., Li, E., Eser, R., Piergies, A., Sit, R., Tan, M., et al. (2021). Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat. Neurosci.* 24 (2), 276–287. doi:10.1038/s41593-020-00764-7
- Li, Y., Lopez-Huerta, V. G., Adiconis, X., Levandowski, K., Choi, S., Simmons, S. K., et al. (2020). Distinct subnetworks of the thalamic reticular nucleus. *Nature* 583 (7818), 819–824. doi:10.1038/s41586-020-2504-5
- Linderman, G., Zhao, J., Roulis, M., Bielecki, P., Flavell, R. A., Nadler, B., et al. (2022). Zero-preserving imputation of single-cell RNA-seq data. *Nat. Commun.* 13 (1), 192. doi:10.1038/s41467-021-27729-z
- Liu, W., Liao, X., Yang, Y., Lin, H., Yeong, J., Zhou, X., et al. (2022). Joint dimension reduction and clustering analysis of single-cell RNA-seq and spatial transcriptomics data. *Nucleic Acids Res.* 50 (12), e72. doi:10.1093/nar/gkac219
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Luecken, M. D., and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: A tutorial. *Mol. Syst. Biol.* 15 (6), e8746. doi:10.15252/msb.20188746
- Lun, A. (2018). bioRxiv, 404962. doi:10.1101/404962 Overcoming systematic errors caused by log-transformation of normalized single-cell RNA sequencing data
- Lun, A., Riesenfeld, S., Andrews, T., Dao, T. P., Gomes, T.; participants in the 1st Human Cell Atlas Jamboree, et al. (2019). EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* 20 (1), 63. doi:10.1186/s13059-019-1662-y
- Lun, A. T. L., Bach, K., and Marioni, J. C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* 17, 75. doi:10.1186/s13059-016-0947-7
- Macosko, E. Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161 (5), 1202–1214. doi:10.1016/j.cell.2015.05.002
- Maitra, M., Nagy, C., Chawla, A., Wang, Y. C., Nascimento, C., Suderman, M., et al. (2021). Extraction of nuclei from archived postmortem tissues for single-nucleus sequencing applications. *Nat. Protoc.* 16 (6), 2788–2801. doi:10.1038/s41596-021-00514-4
- Marei, H., Althani, A., LaShen, S., CenCiarelli, C., and Hasan, A. (2017). Genetically unmatched human iPSC and ESC exhibit equivalent gene expression and neuronal differentiation potential. *Sci. Rep.* 7 (1), 17504. doi:10.1038/s41598-017-17882-1
- Marques, S., Zeisel, A., Codeluppi, S., van Bruggen, D., Mendanha Falcao, A., Xiao, L., et al. (2016). Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Sci. (New York, N.Y.)* 352 (6291), 1326–1329. doi:10.1126/science.aaf6463
- Masuda, T., Sankowski, R., Staszewski, O., Böttcher, C., Amann, L., Sagar, et al. (2019). Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature* 566 (7744), 388–392. doi:10.1038/s41586-019-0924-x
- Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* 570 (7761), 332–337. doi:10.1038/s41586-019-1195-2
- Matsumoto, H., and Kiryu, H. (2016). SCOUP: A probabilistic model based on the ornstein-uhlenbeck process to analyze single-cell expression data during differentiation. *BMC Bioinforma.* 17 (1), 232. doi:10.1186/s12859-016-1109-3
- McGinnis, C. S., Murrow, L. M., and Gartner, Z. J. (2019). DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* 8 (4), 329–337. e4. doi:10.1016/j.cels.2019.03.003
- McInnes, L., Healy, J., Saul, N., and GroBberger, L. (2018). Umap: Uniform manifold approximation and projection. *J. Open Source Softw.* 3 (29), 861. doi:10.21105/joss.00861
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of department of Health and human services task force on Alzheimer's disease. *Neurology* 34 (7), 939–944. doi:10.1212/wnl.34.7.939
- Medeiros, R., and LaFerla, F. M. (2013). Astrocytes: Conductors of the alzheimer disease neuroinflammatory symphony. *Exp. Neurol.* 239, 133–138. doi:10.1016/j.expneurol.2012.10.007
- Miller, J. A., Ding, S. L., Sunkin, S. M., Smith, K. A., Ng, L., Szafer, A., et al. (2014). Transcriptional landscape of the prenatal human brain. *Nature* 508 (7495), 199–206. doi:10.1038/nature13185
- Miller, J. A., Oldham, M. C., and Geschwind, D. H. (2008). A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J. Neurosci.* 28 (6), 1410–1420. doi:10.1523/JNEUROSCI.4098-07.2008
- Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A. J., Tanaka, Y., Wilkinson, A. C., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* 33 (3), 269–276. doi:10.1038/nbt.3154
- Monterey, M. D., Wei, H., Wu, X., and Wu, J. Q. (2021). The many faces of astrocytes in Alzheimer's disease. *Front. Neurol.* 12, 619626. doi:10.3389/fneur.2021.619626
- Moon, K. R., Stanley, J. S., Burkhardt, D., van Dijk, D., Wolf, G., and Krishnaswamy, S. (2018). Manifold learning-based methods for analyzing single-cell RNA-sequencing data. *Curr. Opin. Syst. Biol.* 7, 36–46. doi:10.1016/j.coisb.2017.12.008
- Nagy, C., Maheu, M., Lopez, J. P., Vaillancourt, K., Cruceanu, C., Gross, J. A., et al. (2015). Effects of postmortem interval on biomolecular integrity in the brain. *J. Neuropathol. Exp. Neurol.* 74 (5), 459–469. doi:10.1097/NEN.0000000000000190

- Nativio, R., Donahue, G., Berson, A., Lan, Y., Amlie-Wolf, A., Tuzer, F., et al. (2018). Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease. *Nat. Neurosci.* 21 (4), 497–505. doi:10.1038/s41593-018-0101-9
- Nguyen, V., and Griss, J. (2022). scAnnotatR: framework to accurately classify cell types in single-cell RNA-sequencing data. *BMC Bioinforma.* 23 (1), 44. doi:10.1186/s12859-022-04574-5
- Ni, Z., Chen, S., Brown, J., and Kendziorzi, C. (2020). CB2 improves power of cell detection in droplet-based single-cell RNA sequencing data. *Genome Biol.* 21 (1), 137. doi:10.1186/s13059-020-02054-8
- Noel, T., Wang, Q. S., Greka, A., and Marshall, J. L. (2022). Principles of spatial transcriptomics analysis: A practical walk-through in kidney tissue. *Front. Physiol.* 12, 809346. doi:10.3389/fphys.2021.809346
- Olabarria, M., Noristani, H. N., Verkhatsky, A., and Rodriguez, J. J. (2010). Concomitant astroglial atrophy and astrogliosis in a triple transgenic animal model of Alzheimer's disease. *Glia* 58 (7), 831–838. doi:10.1002/glia.20967
- Olah, M., Menon, V., Habib, N., Taga, M. F., Ma, Y., Yung, C. J., et al. (2020). Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. *Nat. Commun.* 11 (1), 6129. doi:10.1038/s41467-020-19737-2
- Olah, M., Patrick, E., Villani, A. C., Xu, J., White, C. C., Ryan, K. J., et al. (2018). A transcriptomic atlas of aged human microglia. *Nat. Commun.* 9 (1), 539. doi:10.1038/s41467-018-02926-5
- Padurariu, M., Ciobica, A., Mavroudis, I., Fotiou, D., and Baloyannis, S. (2012). Hippocampal neuronal loss in the CA1 and CA3 areas of Alzheimer's disease patients. *Psychiatr. Danub.* 24 (2), 152–158.
- Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2018). zUMIs - a fast and flexible pipeline to process RNA sequencing data with UMIs. *GigaScience* 7 (6). doi:10.1093/gigascience/giy059
- Pasquini, G., Rojo Arias, J. E., Schafer, P., and Busskamp, V. (2021). Automated methods for cell type annotation on scRNA-seq data. *Comput. Struct. Biotechnol. J.* 19, 961–969. doi:10.1016/j.csbj.2021.01.015
- Patrino, L., Maspero, D., Craighero, F., Angaroni, F., Antoniotti, M., and Graudenzi, A. (2021). A review of computational strategies for denoising and imputation of single-cell transcriptomic data. *Brief. Bioinform.* 22 (4), bbaa222. doi:10.1093/bib/bbaa222
- Petukhov, V., Guo, J., Baryawno, N., Severe, N., Scadden, D. T., Samsonova, M. G., et al. (2018). dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biol.* 19 (1), 78. doi:10.1186/s13059-018-1449-6
- Pissadaki, E. K., and Bolam, J. P. (2013). The energy cost of action potential propagation in dopamine neurons: Clues to susceptibility in Parkinson's disease. *Front. Comput. Neurosci.* 7, 13. doi:10.3389/fncom.2013.00013
- Pratap, A., Jaliha, A. P., Law, J. N., Bharadwaj, A., and Murali, T. M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* 17 (2), 147–154. doi:10.1038/s41592-019-0690-6
- Qiu, P. (2020). Embracing the dropouts in single-cell RNA-seq analysis. *Nat. Commun.* 11 (1), 1169. doi:10.1038/s41467-020-14976-9
- Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* 5 (10), 877–879. doi:10.1038/nmeth.1253
- Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* 135 (2), 216–226. doi:10.1016/j.cell.2008.09.050
- Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., et al. (2017). The human cell atlas. *eLife* 6, e27041. doi:10.7554/eLife.27041
- Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11 (3), R25. doi:10.1186/gb-2010-11-3-r25
- Rodrigues, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., et al. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Sci. (New York, N.Y.)* 363 (6434), 1463–1467. doi:10.1126/science.aaw1219
- Rosvall, M., and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U. S. A.* 105 (4), 1118–1123. doi:10.1073/pnas.0706851105
- Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* 37 (5), 547–554. doi:10.1038/s41587-019-0071-9
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33 (5), 495–502. doi:10.1038/nbt.3192
- Saunders, A., Macosko, E. Z., Wysoker, A., Goldman, M., Krienen, F. M., de Rivera, H., et al. (2018). Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* 174 (4), 1015–1030. e16. doi:10.1016/j.cell.2018.07.028
- Setty, M., Tadmor, M. D., Reich-Zeliger, S., Angel, O., Salame, T. M., Kathail, P., et al. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* 34 (6), 637–645. doi:10.1038/nbt.3569
- Shao, C., and Höfer, T. (2017). Robust classification of single-cell transcriptome data by nonnegative matrix factorization. *Bioinforma. Oxf. Engl.* 33 (2), 235–242. doi:10.1093/bioinformatics/btw607
- Shao, X., Liao, J., Lu, X., Xue, R., Ai, N., and Fan, X. (2020). scCATCH: automatic annotation on cell types of clusters from single-cell RNA sequencing data. *iScience* 23, 100882. doi:10.1016/j.isci.2020.100882
- Shin, J., Berg, D. A., Zhu, Y., Shin, J. Y., Song, J., Bonaguidi, M. A., et al. (2015). Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* 17 (3), 360–72. doi:10.1016/j.stem.2015.07.013
- Slovin, S., Carissimo, A., Panariello, F., Grimaldi, A., Bouche, V., Gambardella, G., et al. (2021). Single-cell RNA sequencing analysis: A step-by-step overview. *Methods Mol. Biol.* 2284, 343–365. doi:10.1007/978-1-0716-1307-8\_19
- Slyper, M., Porter, C. B. M., Ashenberg, O., Waldman, J., Drokhyansky, E., Wakiro, I., et al. (2020). A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* 26 (5), 792–802. doi:10.1038/s41591-020-0844-1
- Smajić, S., Prada-Medina, C. A., Landoulsi, Z., Ghelfi, J., Delcambre, S., Dietrich, C., et al. (2021). Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state. *Brain*. 145, 964–978. doi:10.1093/brain/awab446
- Smits, L., Magni, S., Kinugawa, K., Grzyb, K., Luginbuhl, J., Sabate-Soler, S., et al. (2020). Single-cell transcriptomics reveals multiple neuronal cell types in human midbrain-specific organoids. *Cell Tissue Res.* 382 (3), 463–476. doi:10.1007/s00441-020-03249-y
- Sobue, S., Sakata, K., Sekijima, Y., Qiao, S., Murate, T., and Ichihara, M. (2016). Characterization of gene expression profiling of mouse tissues obtained during the postmortem interval. *Exp. Mol. Pathol.* 100 (3), 482–492. doi:10.1016/j.yexmp.2016.05.007
- Stahl, P. L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Sci. (New York, N.Y.)* 353 (6294), 78–82. doi:10.1126/science.aaf2403
- Stegle, O., Teichmann, S. A., and Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16 (3), 133–145. doi:10.1038/nrg3833
- Street, K., Rizzo, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., et al. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 19 (1), 477. doi:10.1186/s12864-018-4772-0
- Svensson, V., Natarajan, K. N., Ly, L. H., Miragaia, R. J., Labalette, C., Macaulay, I. C., et al. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* 14 (4), 381–387. doi:10.1038/nmeth.4220
- Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* 19 (2), 335–346. doi:10.1038/nn.4216
- Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563 (7729), 72–78. doi:10.1038/s41586-018-0654-5
- Todorov, H., Cannoodt, R., Saelens, W., and Saeys, Y. (2020). TinGa: Fast and flexible trajectory inference with growing neural gas. *Bioinformatics* 36 (1), 66–74. doi:10.1093/bioinformatics/btaa463
- Traag, V., Waltman, L., and van Eck, N. J. (2019). From Louvain to leiden: Guaranteeing well-connected communities. *Sci. Rep.* 9 (1), 5233. doi:10.1038/s41598-019-41695-z
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32 (4), 381–386. doi:10.1038/nbt.2859
- Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome Res.* 25 (10), 1491–1498. doi:10.1101/gr.190595.115
- Tsuyuzaki, K., Sato, H., Sato, K., and Nikaido, I. (2020). Benchmarking principal component analysis for large-scale single-cell RNA-sequencing. *Genome Biol.* 21 (1), 9. doi:10.1186/s13059-019-1900-3
- Tushev, G., Glock, C., Heumuller, M., Biever, A., Jovanovic, M., and Schuman, E. M. (2018). Alternative 3' UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments. *Neuron* 98 (3), 495–511. e6. doi:10.1016/j.neuron.2018.03.030
- Tzschentke, T. M., and Schmidt, W. J. (2000). Functional relationship among medial prefrontal cortex, nucleus accumbens, and ventral tegmental area in

- locomotion and reward. *Crit. Rev. Neurobiol.* 14 (2), 12–42. doi:10.1615/critrevneurobiol.v14.i2.20
- Vallejo, A. F., Harvey, K., Wang, T., Wise, K., Butler, L. M., Polo, J., et al. (2022). snPATHO-seq: Unlocking the FFPE archives for single nucleus RNA profiling. *bioRxiv, Prepr.* (Accessed September 21, 2022). doi:10.1101/2022.08.23.505054
- Vallejos, C. A., Risso, D., Scialdone, A., Dudoit, S., and Marioni, J. C. (2017). Normalizing single-cell RNA sequencing data: Challenges and opportunities. *Nat. Methods* 14 (6), 565–571. doi:10.1038/nmeth.4292
- van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* 174 (3), 716–729. e27. doi:10.1016/j.cell.2018.05.061
- Vieth, B., Parekh, S., Ziegenhain, C., Enard, W., and Hellmann, I. (2019). A systematic evaluation of single cell RNA-seq analysis pipelines. *Nat. Commun.* 10 (1), 4667. doi:10.1038/s41467-019-12266-7
- Vieth, B., Ziegenhain, C., Parekh, S., Enard, W., and Hellmann, I. (2017). powsimR: power analysis for bulk and single cell RNA-seq experiments. *Bioinforma. Oxf. Engl.* 33 (21), 3486–3488. doi:10.1093/bioinformatics/btx435
- Volpato, V., and Webber, C. (2020). Addressing variability in iPSC-derived models of human disease: Guidelines to promote reproducibility. *Dis. Model. Mech.* 13 (1), dmm042317. doi:10.1242/dmm.042317
- Wakabayashi, K., Tanji, K., Mori, F., and Takahashi, H. (2007). The Lewy body in Parkinson's disease: Molecules implicated in the formation and degradation of alpha-synuclein aggregates. *Neuropathology* 27 (5), 494–506. doi:10.1111/j.1440-1789.2007.00803.x
- Wang, J., Gouda-Vossos, A., Dzamko, N., Halliday, G., and Huang, Y. (2013). DNA extraction from fresh-frozen and formalin-fixed, paraffin-embedded human brain tissue. *Neurosci. Bull.* 29 (5), 649–654. doi:10.1007/s12264-013-1379-y
- Wattenberg, M., Viégas, F., and Johnson, I. (2016). How to use t-SNE effectively. *Distill* 1 (10). doi:10.23915/distill.00002
- Welch, J. D., Hartemink, A. J., and Prins, J. F. (2017). MATCHER: manifold alignment reveals correspondence between single cell transcriptome and epigenome dynamics. *Genome Biol.* 18 (1), 138. doi:10.1186/s13059-017-1269-0
- Welch, J. D., Hartemink, A. J., and Prins, J. F. (2016). Slicer: Inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. *Genome Biol.* 17 (1), 106. doi:10.1186/s13059-016-0975-3
- Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E. Z. (2019). Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* 177 (7), 1873–1887. e17. doi:10.1016/j.cell.2019.05.006
- White, K., Yang, P., Li, L., Farshori, A., Medina, A. E., and Zielke, H. R. (2018). Effect of postmortem interval and years in storage on RNA quality of tissue at a repository of the NIH NeuroBioBank. *Biopreserv. Biobank.* 16 (2), 148–157. doi:10.1089/bio.2017.0099
- Wolf, F. A., Angerer, P., and Theis, F. J. (2018). Scanpy: Large-scale single-cell gene expression data analysis. *Genome Biol.* 19 (1), 15. doi:10.1186/s13059-017-1382-0
- Wolf, F., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Gottgens, B., et al. (2019). Paga: Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* 20 (1), 59–9. doi:10.1186/s13059-019-1663-x
- Wolock, S. L., Lopez, R., and Klein, A. M. (2019). Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* 8 (4), 281–291. e9. doi:10.1016/j.cels.2018.11.005
- Wu, H., Kirita, Y., Donnelly, E. L., and Humphreys, B. D. (2019). Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: Rare cell types and novel cell states revealed in fibrosis. *J. Am. Soc. Nephrol.* 30 (1), 23–32. doi:10.1681/ASN.2018090912
- Wu, P., An, M., Zou, H. R., Zhong, C. Y., Wang, W., and Wu, C. P. (2020). A robust semi-supervised NMF model for single cell RNA-seq data. *PeerJ* 8, e10091. doi:10.7717/peerj.10091
- Wu, Y.-T., Beiser, A. S., Breteler, M. M. B., Fratiglioni, L., Helmer, C., Hendrie, H. C., et al. (2017). The changing prevalence and incidence of dementia over time - current evidence. *Nat. Rev. Neurol.* 13 (6), 327–339. doi:10.1038/nrnneurol.2017.63
- Yang, S., Corbett, S. E., Koga, Y., Wang, Z., Johnson, W. E., Yajima, M., et al. (2020). Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol.* 21 (1), 57. doi:10.1186/s13059-020-1950-6
- Yao, Z., Liu, H., Xie, F., Fischer, S., Adkins, R. S., Aldridge, A. I., et al. (2021). A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* 598 (7879), 103–110. doi:10.1038/s41586-021-03500-8
- Yao, Z., van Velthoven, C. T. J., Nguyen, T. N., Goldy, J., Sedeno-Cortes, A. E., Baftizadeh, F., et al. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* 184 (12), 3222–3241.e26. e26. doi:10.1016/j.cell.2021.04.021
- Yip, S. H., Sham, P. C., and Wang, J. (2019). Evaluation of tools for highly variable gene discovery from single-cell RNA-seq data. *Brief. Bioinform.* 20 (4), 1583–1589. doi:10.1093/bib/bby011
- You, Y., Tian, L., Su, S., Dong, X., Jabbari, J. S., Hickey, P. F., et al. (2021). Benchmarking UMI-based single-cell RNA-seq preprocessing workflows. *Genome Biol.* 22 (1), 339. doi:10.1186/s13059-021-02552-3
- Young, M., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* 9 (12), gaa151. doi:10.1093/gigascience/giaa151
- Zagare, A., Barmppa, K., Smajic, S., Smits, L. M., Grzyb, K., Grunewald, A., et al. (2022). Midbrain organoids mimic early embryonic neurodevelopment and recapitulate LRRK2-p.Gly2019Ser-associated gene expression. *Am. J. Hum. Genet.* 109 (2), 311–327. doi:10.1016/j.ajhg.2021.12.009
- Zeisel, A., Kostler, W. J., Molotski, N., Tsai, J. M., Krauthgamer, R., Jacob-Hirsch, J., et al. (2011). Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Mol. Syst. Biol.* 7, 529. doi:10.1038/msb.2011.62
- Zeng, H., and Sanes, J. R. (2017). Neuronal cell-type classification: Challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* 18 (9), 530–546. doi:10.1038/nrn.2017.85
- Zetterström, R. H., SoLomin, L., Jansson, L., Hoffer, B. J., Olson, L., and Perlmann, T. (1997). Dopamine neuron agenesis in Nurr1-deficient mice. *Sci. (New York, N.Y.)* 276 (5310), 248–250. doi:10.1126/science.276.5310.248
- Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., et al. (2019). CellMarker: A manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* 47 (1), D721–D728. doi:10.1093/nar/gky900
- Zhang, Y., Sloan, S. A., Clarke, L. E., Caneda, C., Plaza, C. A., Blumenthal, P. D., et al. (2016). Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* 89 (1), 37–53. doi:10.1016/j.neuron.2015.11.013
- Zhang, Z., and Zhang, X. (2021). Inference of high-resolution trajectories in single-cell RNA-seq data by using RNA velocity. *Cell Rep. Methods* 1 (6), 100095. doi:10.1016/j.crmeth.2021.100095
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049. doi:10.1038/ncomms14049
- Zhu, Y., Wang, L., Yin, Y., and Yang, E. (2017). Systematic analysis of gene expression patterns associated with postmortem interval in human tissues. *Sci. Rep.* 7 (1), 5435. doi:10.1038/s41598-017-05882-0