Introduction to Single-Cell RNA Sequencing

Thale Kristin Olsen[1] and Ninib Baryawno[1]

[1]Childhood Cancer Research Unit, Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

During the last decade, high-throughput sequencing methods have revolutionized the entire field of biology. The opportunity to study entire transcriptomes in great detail using RNA sequencing (RNA-seq) has fueled many important discoveries and is now a routine method in biomedical research. However, RNA-seq is typically performed in "bulk," and the data represent an average of gene expression patterns across thousands to millions of cells; this might obscure biologically relevant differences between cells. Single-cell RNA-seq (scRNA-seq) represents an approach to overcome this problem. By isolating single cells, capturing their transcripts, and generating sequencing libraries in which the transcripts are mapped to individual cells, scRNA-seq allows assessment of fundamental biological properties of cell populations and biological systems at unprecedented resolution. Here, we present the most common scRNA-seq protocols in use today and the basics of data analysis and discuss factors that are important to consider before planning and designing an scRNA-seq project. © 2018 by John Wiley & Sons, Inc.

Keywords: gene expression profiling • RNA sequencing • single-cell analysis

How to cite this article:
Olsen, T. K., & Baryawno, N. (2018). Introduction to single-cell RNA sequencing. *Current Protocols in Molecular Biology*, *122*, e57. doi: 10.1002/cpmb.57

INTRODUCTION

The term "transcriptome" refers to the complete set of transcripts in a cell. This includes the cell's total gene expression and consequently its implied phenotype and functional state (Mortazavi, Williams, McCue, Schaeffer, & Wold, 2008). The transcriptome encompasses protein-coding messenger RNA (mRNA) as well as key players in ribosomal function (ribosomal RNA and transfer RNA, or rRNA and tRNA) and noncoding RNAs with regulatory functions. The emergence of high-throughput sequencing technology in 2005 (Margulies et al., 2005) enabled study of the transcriptome in great detail by analyzing RNA at the nucleotide level.

Bulk RNA sequencing (RNA-seq) is performed with RNA extracted from homogenized tissue or from large cell populations. This allows for the study of gene expression and alternative splicing, the discovery of variants, and the detection of chimeric gene fusions caused by genomic rearrangements (Ozsolak & Milos, 2011). However, as the RNA extracted and studied represents an "average" of thousands to millions of individual cell transcriptomes present in the sample, potentially significant and biologically important differences between cells may be obscured.

In 2009, only four years after high-throughput sequencing first revolutionized the field of biology, the first protocol for performing single-cell RNA-seq (scRNA-seq) was published (Tang et al., 2009). This technology enables the study of biological properties of individual cells at unprecedented resolution. In the first scRNA-seq studies, between 10 and 100 cells were analyzed and characterized (Kumar et al., 2014; Picelli et al., 2013; Shalek et al., 2013). Since then, the method has gradually been refined and improved. Costs have been reduced, throughput has increased, and

Current Protocols in Molecular Biology e57, April 2018
Published online April 2018 in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cpmb.57

several commercial platforms are now available to the research community. Researchers are currently able to analyze and sequence the transcriptomes of up to tens of thousands of individual cells for a single project (Klein et al., 2015; Macosko et al., 2015).

Here, we provide an overview of the most common scRNA-seq protocols in use today and the basics of scRNA-seq data analysis. We also discuss the biological applications as well as some of the challenges related to emerging scRNA-seq technologies. Finally, we present a practical guide for researchers planning their first scRNA-seq experiments.

OVERVIEW OF METHODOLOGY

A wide range of protocols and approaches for scRNA-seq have been developed and refined in recent years. Although there are variations, all these protocols follow the same basic principles. Single cells must first be isolated from each other for individual processing; this often includes enzymatic dissociation of tissue into a liquid cell suspension. After isolation, single cells are lysed, and RNA molecules within individual cells are captured and converted to cDNA by means of reverse transcription. Often, this step also adds cell-specific DNA barcodes onto the cDNA molecules. The cDNA is then amplified. Finally, libraries for high-throughput RNA-seq are prepared.

TISSUE DISSOCIATION

Many scRNA-seq protocols begin by isolating single cells from a cell suspension. This means that the cells must be released from their tissue surroundings, from cell-to-cell adhesions, and from the extracellular matrix, which is typically achieved by mechanical dissociation of tissue as well as enzymatic dissociation using trypsin, collagenase, and/or papain. This initial step is not trivial because enzymatic treatment may affect cell viability or lead to transcriptional changes (Huang et al., 2010; Kolodziejczyk, Kim, Svensson, Marioni, & Teichmann, 2015). Mechanical stress at this step may also alter the cells' transcriptional programs (Poulin, Tasic, Hjerling-Leffler, Trimarchi, & Awatramani, 2016).

Single-Cell Capture

Once a single-cell suspension is obtained, the cells must be isolated individually so that downstream biochemical reactions can occur independently in each cell. The most commonly used approaches for isolating single cells from suspension are fluorescence-activated cell sorting (FACS) and microfluidics-based techniques. Both of these enable automated, high-throughput, cost-effective single-cell isolation and are described in more detail below. Cells may also be isolated manually from suspension/culture by mechanical micromanipulation (micropipetting). Finally, it is also possible to isolate single cells from fixed tissue on glass slides; this can be done by means of laser capture microdissection (LCM).

Fluorescence-activated cell sorting

FACS is a type of flow cytometry technique in which cells are sorted according to size, morphology, and/or fluorescence properties (Fig. 33.2.1A). After tissue dissociation, cells in suspension are labeled using mixtures of fluorescently conjugated antibodies targeting specific cell markers. The sample is run as a stream of liquid through a narrow channel where cells are exposed to a laser beam. Fluorescence detectors then identify cells with the desired staining pattern, and these cells are collected into 96- or 384-well plates containing lysis buffer, with one cell per well.

FACS is a commonly used method due to its high-throughput, cost-efficient, automated isolation of thousands of individual cells. Antibody labeling also enables the enrichment of cells of interest. However, the method requires very large numbers of cells (typically tens of thousands) as starting material (Hu, Zhang, Xin, & Deng, 2016). Thus, single-cell isolation from samples with low starting amounts, such as fine-needle aspirates, may be difficult (Saliba, Westermann, Gorski, & Vogel, 2014). Issues to consider with this approach are that contamination from laboratory surroundings may occur and that sorting errors can occasionally result in empty or doublet wells.

Microfluidics

Microfluidic techniques enable the manipulation of ultra-low liquid volumes (on the nanoliter scale). The liquid flow is tightly controlled using pre-fabricated chips with microchannels. Several platforms are available, allowing not only automated single-cell isolation but also downstream biochemical reactions such as cDNA synthesis and transcriptome amplification (Picelli, 2016). These reactions are performed in a parallel, automated fashion, which allows for high throughput and use of very low volumes of reagents. One such example is the Fluidigm C1 platform. These arrays allow the capture of up to 800 cells per run, with reverse transcription and cDNA

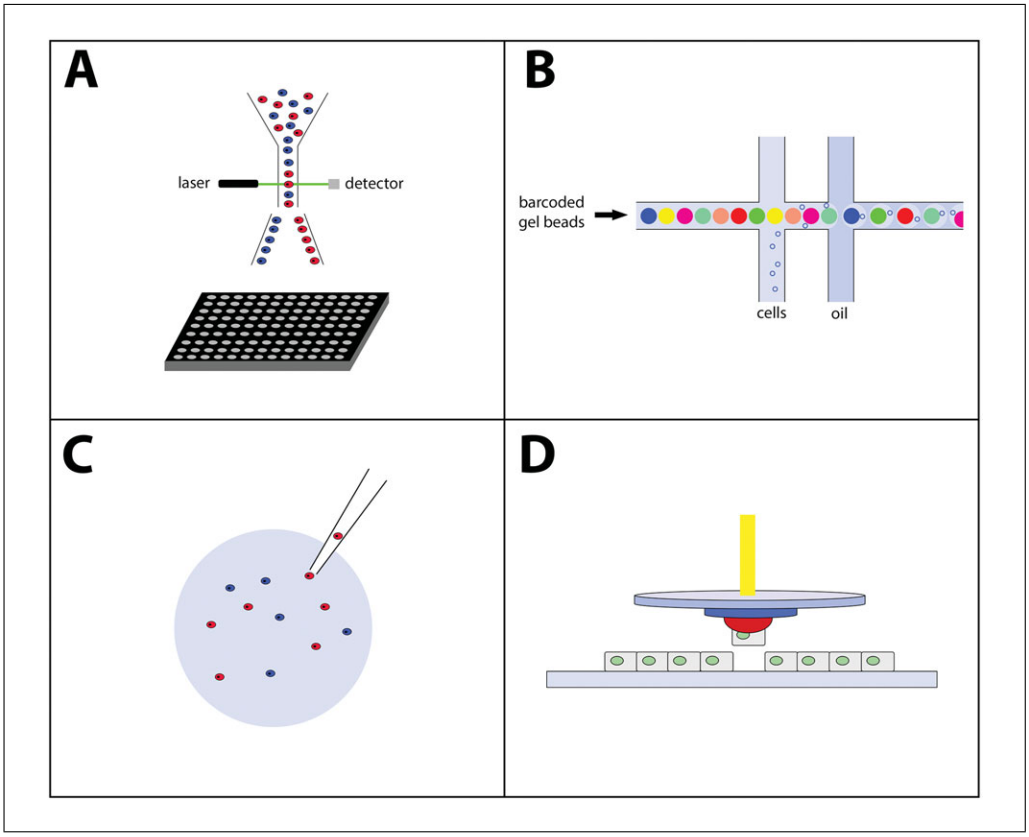

Figure 33.2.1 Single-cell isolation methods. (A) Fluorescence-activated cell sorting (FACS). Fluorescently labeled cells are exposed to a laser beam. Detectors identify cells with the desired fluorescence pattern, and the cells are sorted into plates, with one cell in each well. (B) Microdroplets. Cells and beads containing primers and reagents are enclosed in microdroplets. Downstream biochemical reactions occur within these droplets. (C) Micropipetting. Cells of interest are manually picked with a glass pipet under a microscope. (D) Laser capture microdissection (LCM). Cells on a glass slide are attached to a polymer by means of a laser beam. The polymer and cells of interest are then lifted from the slide and transferred to a microcentrifuge tube.

amplification occurring in individual reaction chambers surrounding each cell. The platform also allows visual inspection of the captured cells (Kolodziejczyk et al., 2015).

In droplet-based microfluidic platforms (Fig. 33.2.1B), the cell suspension is mixed with beads, with each bead containing a unique barcoded primer. The beads and cells are then encapsulated into nanoliter-size droplets in an oil emulsion. In the Drop-seq protocol (Macosko et al., 2015), the cells within each droplet are lysed, and the unique barcoded primers are hybridized to the polyadenylated mRNA in each droplet. Thus, every cell is marked with its own unique barcode. Furthermore, each transcript within each cell is also marked with a unique molecular identifier (UMI; see below). The barcoded transcripts are then pooled, and library preparation is performed in a single tube.

inDrop (Klein et al., 2015) is another droplet protocol. Within each droplet, after photoactivation, mRNA fragments are barcoded, and cDNA is synthesized. The barcoded cDNA is pooled and then amplified using *in vitro* transcription (IVT; see below). A similar principle is also used in the commercially available Chromium Single Cell platform by 10× Genomics (Zheng et al., 2017).

Mechanical micromanipulation/ micropipetting

Here, a capillary glass pipet is used to aspirate single cells from a suspension (Fig. 33.2.1C) or from tissue, thereby providing spatial information (Silberstein et al., 2016a). This procedure is performed manually under a microscope. Micromanipulation may be used to study samples with very few or very fragile cells, such as early embryos or niche cells in the bone-marrow microenvironment (Silberstein et al., 2016a; Xue et al., 2013). However, this is a technically challenging, time-consuming, and low-throughput method that can also cause cellular injury due to mechanical shearing (Brehm-Stecher & Johnson,

2004; Ishii, Tago, & Senoo, 2010; Silberstein et al., 2016a; Yilmaz & Singh, 2012).

Laser capture microdissection

LCM (Fig. 33.2.1D and *UNIT 25A.3* Chandran et al., 2015) enables the isolation of single cells from tissue sections under a microscope. After visualization, a laser beam is focused on cells of interest, which fuses the cells onto a thin, transparent film. When this film is lifted, the selected single cells remain attached to the film, whereas surrounding tissue is left behind on the slide. The film is then transferred to a microcentrifuge tube containing appropriate buffer solutions.

Because cells are isolated from tissue slides, LCM has the advantage of preserving the spatial relationship between the isolated cells. It also does not require enzymatic tissue dissociation. However, the tissue must be very thinly sectioned, which may cause loss of material if the cell diameter is larger than the thickness of the section. The method is also low throughput and labor intensive and requires specialized equipment (Datta et al., 2015; Emmert-Buck et al., 1996; Frumkin et al., 2008; Hu et al., 2016; Okuducu, Hahne, Von Deimling, & Wernert, 2005).

REVERSE TRANSCRIPTION AND CDNA AMPLIFICATION

As current high-throughput sequencing platforms are only able to sequence DNA molecules, reverse transcription (from mRNA to cDNA) followed by cDNA amplification is necessary before sequencing can be performed. The cDNA must also be amplified in order to generate sufficient amounts of material for preparing the sequencing libraries. The two main strategies used are template switching (TS) and IVT.

Tang et al. (2009) described the first scRNA-seq protocol. A primer sequence containing an oligo-dT sequence and a sequencing adapter was used to synthesize the first complementary cDNA strand (Fig. 33.2.2A). A poly-A tail was then added to this first strand by means of a terminal transferase. A second primer, also an oligo-dT primer, containing another adapter sequence, was then used to synthesize the second cDNA strand. The resulting cDNA molecule thus carried known adapter sequences at each end. These were then amplified by PCR, and libraries were generated for sequencing on the SOLiD sequencing platform.

TS is an intrinsic property of the Moloney murine leukemia virus (MMLV) reverse transcriptase (Tang et al., 2013). TS is employed in the widely used Smart-seq and Smart-seq2 protocols (Picelli et al., 2013; Ramsköld et al., 2012), which are designed for use with Illumina's sequencers, as well as in the single-cell tagged reverse transcription (STRT)-seq protocol (Islam et al., 2011, 2014). In TS, the reverse transcriptase introduces untemplated nucleosides (2 to 5 cytosines) when it reaches the 5′ end of the template. Briefly, mRNA is initially primed with oligo-dT primers containing an adapter sequence (Fig. 33.2.2A). The first cDNA strand is synthesized, and then, by means of the MMLV transcriptase, some cytosines (not present in the original template) are added to the 3′ end of the newly synthesized strand. These cytosine residues serve to dock a second helper oligo (TS oligonucleotide, or TSO), which carries three guanine residues at its 3′ end. Then, the "switch" occurs: the reverse transcriptase uses the DNA of the helper oligo as a template and continues to transcribe until reaching the end of the primer (Picelli, 2016). This provides full-length cDNA molecules containing both the 5′ end of the original mRNA and an anchor sequence that serves as a priming site for second-strand synthesis. The known sequences at both ends of the cDNA molecule allow for efficient PCR amplification.

IVT represents another approach and is used in the CEL-seq protocol (Hashimshony, Wagner, Sher, & Yanai, 2012). The use of linear amplification instead of PCR reduces so-called amplification bias (Grün & van Oudenaarden, 2015). At the first step of reverse transcription, a primer that contains a poly-T sequence, a unique barcode, an Illumina adapter sequence, and a T7 polymerase promoter sequence is attached (Fig. 33.2.2B). The primer hybridizes onto the poly-A tail of mRNA, and the first and second cDNA strands are synthesized. The cells are then pooled, and IVT is performed from the T7 promoter in order to amplify the cDNA. The resulting antisense RNA (aRNA) molecules are fragmented, and another Illumina 5′ adapter is ligated. Another round of reverse transcription results in double-stranded cDNA fragments. Finally, the fragments containing barcodes and both Illumina adapters, *i.e.,* the fragments closest to the 3′ end of the transcripts, are selected. When subjected to paired-end sequencing, read 1 contains the barcode identifying each cell, and read 2 contains the mRNA transcript sequence.

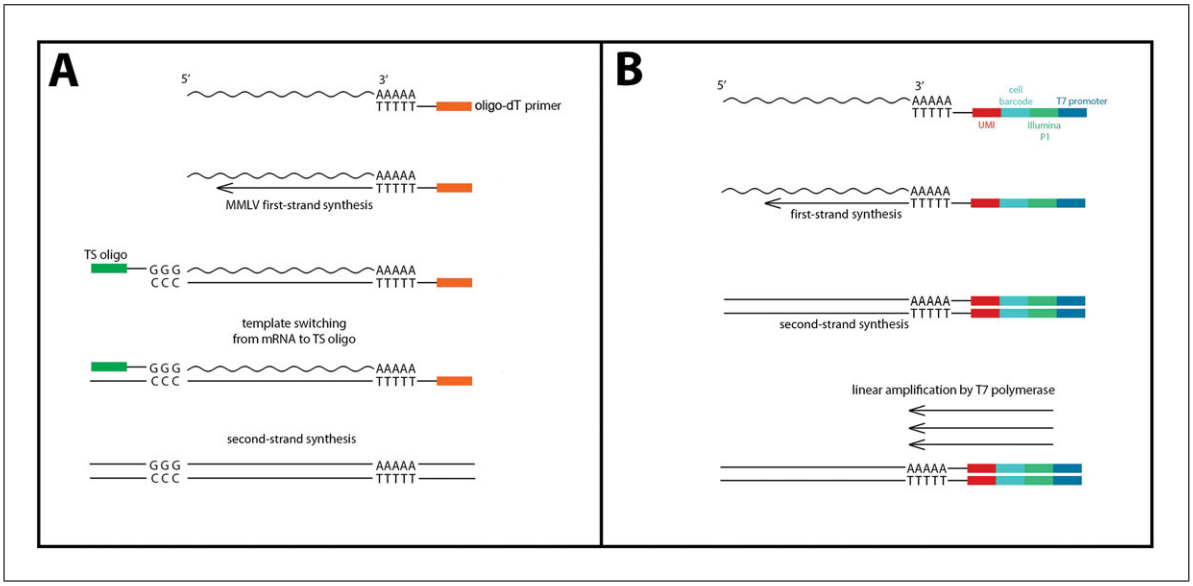

Figure 33.2.2 Strategies for cDNA amplification. (A) Template switching (TS). Reverse transcription is performed using polyadenylated mRNA as template and an oligo-dT/PCR primer. After first-strand synthesis, the MMLV reverse transcriptase introduces untemplated cytosine residues when it reaches the 5′ end of the template. A second primer anneals to these cytosines. The transcriptase then switches template from the mRNA to the TS oligo. As a result, full-length cDNA with known primer sequences at both ends is obtained. The cDNA is then amplified by PCR. (B) *In vitro* transcription (IVT). An oligo-dT primer containing 1) a UMI, 2) a unique cell barcode, 3) an Illumina adapter, and 4) a T7 promoter is attached to polyadenylated mRNA. After first- and second-strand synthesis, the uniquely barcoded cDNA fragments are pooled and amplified using T7 polymerase. The resulting cDNA will have a strong 3′ bias.

Thus, there is an inherent strong 3′ bias in CEL-seq compared to full-length approaches such as Smart-seq2, which obtains read coverage along the entire transcript (Grün & van Oudenaarden, 2015).

The CEL-seq protocol was recently modified. In the CEL-seq2 protocol (Hashimshony et al., 2016), the initial primers also contain UMIs. UMIs are sequences of 4 to 10 random nucleotides that are used to identify each individual mRNA molecule in each cell. These UMIs are amplified along with the cDNA and sequenced in the final Illumina libraries. Sequencing reads with identical UMIs represent reads that are amplified from the same original mRNA fragment. By counting each UMI only once and counting the number of UMIs per gene, it is possible to estimate the number of transcripts from a given gene that are present in a cell.

DATA ANALYSIS

After single-cell capture, reverse transcription, and cDNA amplification, the samples are ready for library preparation and sequencing. The resulting raw data may seem similar to bulk RNA-seq data but are highly complex and unique in many aspects. Thus, scRNA-seq data analysis involves computational challenges that are still being addressed, and algorithms and statistical methods are under continuous development. In any scRNA-seq project, it is therefore crucial to collaborate with computational biologists with experience and expertise in analyzing these kinds of data. The main principles of scRNA-seq data analysis will be outlined below. For a more detailed review, we refer to other recent publications (Bacher & Kendziorski, 2016; Grün & van Oudenaarden, 2015; Stegle, Teichmann, & Marioni, 2015).

scRNA-seq data analysis involves rigorous filtering and quality control. This is particularly important because scRNA-seq data contain a lot of technical variance (*i.e.,* "noise"). Several aspects contribute to this variation, which may be introduced at several stages of the scRNA-seq process, from cell capture and lysis to mRNA capture and reverse transcription. Many expressed transcripts will remain undetected (false negatives or "dropout effect") (Wagner, Regev, & Yosef, 2016). Amplification bias, meaning that the distribution of amplified fragments is uneven across the transcriptome, is another common problem (Bacher & Kendziorski, 2016).

In raw sequencing data, each nucleotide is associated with a quality score, describing the probability that the base call is correct.

This information is used to assess the raw read and overall sequencing quality, for instance by using existing bulk RNA-seq tools such as FastQC (*https://www.bioinformatics.babraham.ac.uk/projects/fastqc/*). Cells containing an unusually high number of low-quality reads may be excluded from further analysis. Reads with a low-quality cell barcode or UMI sequences may also be discarded.

The next step typically involves aligning the reads to the transcriptome. This is performed using specialized RNA-seq tools such as STAR (Dobin et al., 2013), MapSplice (Wang et al., 2010), or Salmon (Patro, Duggal, Love, Irizarry, & Kingsford, 2017). After alignment, another important level of quality control is performed. Some common metrics are the total number of mapped reads and the proportion of uniquely mapping reads. High proportions of mitochondrial reads might indicate cell damage (Bacher & Kendziorski, 2016).

The gene expression patterns of each cell are obtained by quantifying the number of reads mapping to each gene, for instance by using tools such as htseq-count (Anders, Pyl, & Huber, 2015). If UMIs are used, the numbers of transcripts per gene per cell are quantified, with each UMI representing one individual transcript. Reads with identical UMIs will be the result of amplification during library preparation, and not a consequence of transcription in the cell. Thus, by counting the numbers of UMIs per gene instead of the number of reads, much of the amplification bias can be eliminated (Grün & van Oudenaarden, 2015).

After gene expression quantification, some sort of normalization must be performed. The aim is to adjust for variation caused by technical artifacts. The normalization strategy depends on the sample preparation protocol and whether UMIs and/or spike-ins are used. Spike-in transcripts are synthetic transcripts that are added at known concentrations to lysed cells. After reverse transcription, amplification, and sequencing, the number of reads corresponding to the spike-in transcripts is quantified. Theoretically, all cells should contain the same amount of spike-in transcripts. By comparing the expected and observed spike-in expression in each cell, it is possible to adjust for some of the technical variation (Stegle et al., 2015).

Finally, cells can be compared based on their gene expression signatures. This can be used to identify cell types or subpopulations of cells and to perform differential gene expression analysis across cell types or experimental conditions. Some methods for cell subtyping include hierarchical clustering, principal component analysis (PCA), and t-distributed stochastic neighbor embedding (t-SNE) (Bacher & Kendziorski, 2016). Tools that were originally developed for bulk RNA-seq, such as DESeq2 (Love, Huber, & Anders, 2014) or edgeR (Robinson, McCarthy, & Smyth, 2010), can be used as well. However, more recent methods developed specifically for scRNA-seq, such as scde and pagoda (Fan et al., 2016; Kharchenko, Silberstein, & Scadden, 2014), are also available.

APPLICATIONS IN BIOLOGY

scRNA-seq is a powerful method that, with its ability to identify and characterize different cell types within tissues, has been successfully used in several fields of research. Some examples of such studies are outlined below.

Cancer

Cancer is an inherently heterogeneous disease, as there is considerable genetic variation between patients, between primary tumors and metastases, and within tumors (Burrell, McGranahan, Bartek, & Swanton, 2013). scRNA-seq is well suited to delving into inter-tumor and intra-tumor heterogeneity, which often cannot be investigated by means of bulk sequencing. For instance, scRNA-seq has revealed considerable intra-tumor heterogeneity in human glioblastoma (Patel et al., 2014) and hepatocellular carcinoma (Hou et al., 2016).

Several cell populations or clones may exist within one tumor. Evolution of these clones is of paramount clinical importance, as it represents a driving force behind tumor development, tumor progression, drug resistance, and metastasis. For instance, Tirosh et al. (2016b) described novel cell subpopulations with stem-like characteristics in oligodendroglioma, and Kim et al. (2015) studied heterogeneity and drug-resistant clones in lung adenocarcinoma cells.

Apart from heterogeneity, scRNA-seq is also used to study other aspects of cancer biology. The cells surrounding a tumor (its microenvironment) are widely recognized to contribute to several aspects, or hallmarks, of carcinogenesis (Hanahan & Weinberg, 2011). A recent paper (Tirosh et al., 2016a) studied the melanoma microenvironment using scRNA-seq. Another example area of study is circulating tumor cells (CTCs), which are cells that have "escaped" from tumors or metastases

and have entered the bloodstream. These can be isolated from "liquid biopsies" and are associated with disease outcome (Siravegna, Marsoni, Siena, & Bardelli, 2017). Several studies have used scRNA-seq to identify and characterize CTCs (Lohr et al., 2016; Miyamoto et al., 2015).

Developmental Biology

scRNA-seq can be used to shed light on the process of stem cell differentiation. What molecular processes underlie a stem cell's "decision" to differentiate into several potential lineages? In this context, Treutlein et al. (2014) have studied lung development in mice. The processes of embryogenesis and early mammalian development can also be studied using scRNA-seq, as shown in a recent study by Scialdone et al. (2016).

Immunology

Variation and diversity within the immune system help us to develop efficient defense mechanisms against a wide variety of microorganisms present in our environment. scRNA-seq may be used to study differentiation and heterogeneity in the context of immunology. For instance, scRNA-seq has revealed heterogeneity among Th2 CD4+ T helper cells and has helped to identify novel Th2 subpopulations (Mahata et al., 2014). Another study used scRNA-seq to study mouse bone-marrow-derived dendritic cells (Shalek et al., 2013).

Neuroscience

The central nervous system is immensely complex, with anatomically distinct areas serving highly specialized functions. scRNA-seq has been used by several groups to identify and characterize neuronal subtypes in different regions of the murine central nervous system (La Manno et al., 2016; Marques et al., 2016; Poulin et al., 2016) and the human brain (Darmanis et al., 2015; Lake et al., 2016).

EXPERIMENTAL PLANNING AND DESIGN: A PRACTICAL GUIDE

Until now, most scRNA-seq studies have been performed in highly specialized laboratories, both wet and dry, with unique expertise and equipment. However, as commercial platforms are becoming increasingly available and bioinformatic methods are steadily maturing, scRNA-seq is now on the brink of reaching the wider research community.

Why Single-Cell RNA Sequencing?

As we have described in this review, scRNA-seq is well suited to addressing a wide range of research topics, such as defining tissue heterogeneity, identifying rare cell populations, detecting cellular and molecular changes in diseased tissue compared to normal tissue, and tracing cell lineage and differentiation. It is thus a powerful technique that can help to answer a wide range of research questions and advance our understanding of biology. An example of an scRNA-seq experiment is provided in Figure 33.2.3. However, the level of detail that can be obtained, such as the number of expressed genes detected in a cell and the number of transcripts per gene, depends on the question asked and the choice of protocol.

Which Protocol to Choose?

Tissue type, tissue dissociation, and cell capture

A good place to start is to consider the tissue and cell types of interest (Fig. 33.2.4). Are you studying human or mouse cells? Do you have access to multiple biological replicates, such as several mice treated under similar experimental conditions? What would be the best set of controls for your experiment? The cells of interest must also be isolated from their surroundings. For immune cells in peripheral blood or for cell lines, this is typically relatively straightforward. For excised tumors, it is also achievable. However, for other cell types, such as neurons, isolation might not be as straightforward. As a starting point, we advise researchers to optimize the collection of single, viable cells, for instance by enzymatic tissue dissociation. It is also important to consider whether banking (cryopreservation) is an option. Moving straight from single-cell isolation to cell lysis, cDNA synthesis, and amplification might represent a challenge, especially when multiple samples are processed simultaneously. Biobanking might ease logistics, albeit at the cost of cell viability.

The choice of cell capture method is closely related to the tissue and cell types of interest and the rarity of the cells. FACS and microfluidic techniques are high throughput and automated and lead to the isolation of large numbers of cells but also require high cell numbers as a starting point. If your cell type of choice is present at very low numbers (such as embryonic cells) or spatial information is important, LCM followed by Smart-seq2 may be a better choice (Nichterwitz et al., 2016). In contrast, if you want to characterize a tissue in

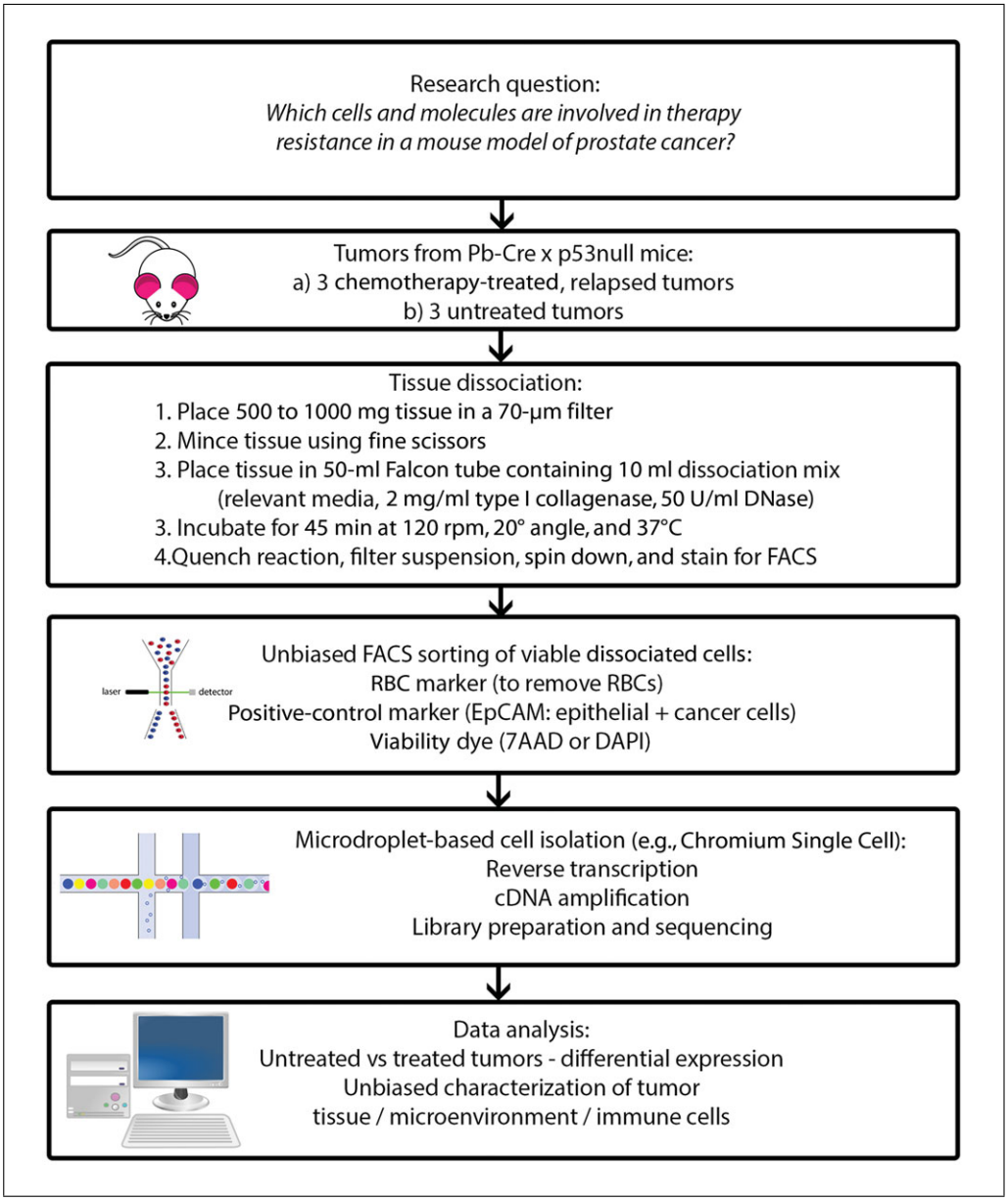

Figure 33.2.3 An example of a prostate cancer project outline. The aim is to identify cells/molecules that contribute to chemotherapy resistance in a mouse model of prostate cancer. We recommend at least three biological replicates in each group. An example of a tissue dissociation protocol is provided. During FACS sorting, red blood cells (RBCs) will be removed. Viable (7AAD- and/or DAPI-negative) cells will be kept for downstream analysis. After sorting, cells will be isolated, and sequencing libraries will be prepared using the droplet-based Chromium Single Cell platform protocol. In the subsequent data analysis, the first aim will be to compare the two groups (untreated *vs.* treated tumors) in terms of differential expression. Which markers/genes characterize the resistant tumors? An unbiased characterization of the tumor, stroma, and immune cells will also be performed (*e.g.,* do the immune cell populations present in the two tumor groups differ? Are there any distinct tumor cell populations?).

an unbiased manner, FACS isolation followed by a microfluidic approach (e.g., inDrop/Drop-seq/Chromium Single Cell/Fluidigm C1) is the better option.

Coverage

A key difference between scRNA-seq protocols is that some, such as Smart-seq2, provide full-length transcript data, whereas others, such as Drop-seq and inDrop, specifically target the 3′ end of transcripts. If a research project requires a great level of detail per cell, such as splicing variants, full-length protocols with high sensitivity, such as Smart-seq2, are probably the best option.

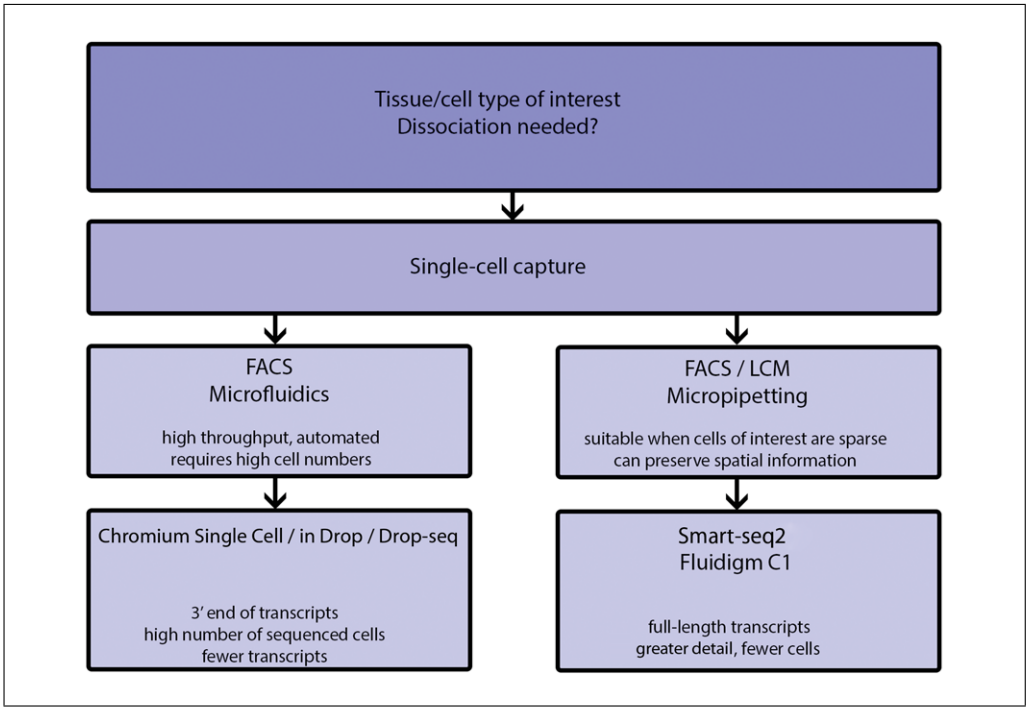

Figure 33.2.4 Overview of some important aspects to consider when planning an scRNA-seq project. What tissue and cell type(s) will be studied? Different tissue types might require varying preparation steps. If enzymatic dissociation is needed, the protocol must be optimized for the tissue in question. The choice of method for cell isolation depends on the biological question. FACS and microfluidic techniques can isolate high numbers of single cells and are cost effective and high throughput but require high starting numbers of cells. Spatial information will also be lost. LCM and micropipetting methods might be more suitable when cells of interest are sparse. The choice of protocol for reverse transcription and cDNA amplification will also depend on the research question. For unbiased tissue characterization and identification of rare cell populations, droplet-based methods are a good choice, as they enable the sequencing of many individual transcriptomes. Smart-seq2 and similar protocols are more sensitive and provide greater detail but will result in fewer sequenced single cells.

Cell numbers and sequencing depth

In any scRNA-seq experiment, it is important to consider how many cells to sequence and at what depth. These two parameters are closely related: as there is a limit to how many reads a given flow cell can produce, the number of transcripts detected for each cell will decrease as more and more cells are barcoded and added to a sequencing library. The desired number of cells will typically be higher for more complex tissues. Increasing the number of cells in these cases will give greater statistical power. If the goal is to classify various cell types in a heterogeneous population or to identify rare cells, 50,000 to 100,000 reads per cell have been suggested as an appropriate goal (Pollen et al., 2014). In such cases, droplet-based methods may be a good choice (Ziegenhain et al., 2017). However, studying more subtle gene expression changes or characterizing the transcriptome of a specific cell type at high resolution requires greater sequencing depth. In these cases, Smart-seq2 may be a better choice (Ziegenhain et al., 2017). The recommended sequencing depth for 96 single cells prepared with Smart-seq2 is typically 2 × 25 bp on one lane of Illumina HiSeq, which yields approximately 1 million aligned reads per single cell (*UNIT 4.22*; Trombetta et al., 2014).

Data Analysis

Even though scRNA-seq libraries can now be generated by more and more "first-time" users, data analysis is complex and requires bioinformatic expertise. Some companies, such as 10× and Fluidigm, offer free software that is designed to be used with their platforms. Web-based, more user-friendly interfaces are also under development from several groups. However, it is important to note that this field is rapidly evolving and that there are no "gold-standard" analysis pipelines available at the time of writing (November 2017).

Validation

In order to interpret scRNA-seq data in a biological context and to assess the relevance of these data, functional validation is an important step. For instance, if scRNA-seq data analysis results in the identification of certain cell types characterized by certain marker genes, an initial validation step could be to visualize these cell types in the original tissue. For example, one way to do this is to label the cells using immunohistochemistry gene-specific fluorescence *in situ* hybridization (FISH) probes and immunohistochemical antibody staining of sample sections to validate that the functional protein is present. These approaches also provide spatial information on the cell type of interest (Shekhar et al., 2016).

The next level of validation would be to assess the functional properties of a predicted cell type. One such approach could be to FACS sort cells based on a genetic marker identified by scRNA-seq and to perform functional assays, such as *in vitro* colony-forming assays (Guo et al., 2013) or drug sensitivity assays. scRNA-seq data may also be validated *in vivo*, for instance by using xenografts to validate the tumorigenic capability of proposed cancer stem cells or in lineage-tracing genetic mouse models.

Critical Parameters

The vast majority of transcripts in a cell are rRNA and tRNA. These are of limited biological interest compared to coding mRNA and noncoding RNAs with regulatory functions, such as micro-RNA (miRNA). Current scRNA-seq protocols are based on the positive selection of mRNA by using oligo-dT primers binding to poly-A tails. Thus, these protocols are not well suited to the study of noncoding RNAs. There are protocols for rRNA depletion that can be used to prepare bulk RNA-seq libraries, but these protocols currently require larger amounts of RNA as starting material than can be obtained from a single cell (Picelli, 2016).

The starting amount of RNA in a cell is minimal and therefore highly vulnerable to bias, which may be introduced at any step regardless of the protocol used. To some extent, current protocols and algorithms account for this bias, for instance by introducing UMIs. Nevertheless, it is of utmost importance to follow the chosen protocol closely. scRNA-seq is a young field in rapid development, and improvements and tweaks with the aim of reducing such bias, both in the wet laboratory and *in silico*, are continuously being published.

Reverse transcription and amplification represent particular challenges. Both are prone to bias and technical artifacts. Single-molecule nanopore sequencing technology represents a potential future solution to this problem (Picelli, 2016). In a recent publication, seven murine B cells were prepared according to the Smart-seq2 protocol and sequenced using long-read, single-molecule nanopore technology (Byrne et al., 2017). A protocol for direct RNA-seq in nanopores, that is, without cDNA conversion, has also recently been published (Garalde et al., 2016).

Enzymatic tissue dissociation is performed using fresh tissue samples. Access to such samples is limited and may represent a logistical challenge. Furthermore, the dissociation process may compromise the integrity of some cell types and induce a dissociation signature that can interfere with the interpretation of results (van den Brink et al., 2017). Single-nucleus RNA-seq (Habib et al., 2016; Lake et al., 2016) can be performed by extracting nuclei from fresh frozen tissue without the need for enzymatic dissociation. Archival material can be studied, which makes this a very interesting approach, for instance in the field of cancer research.

In a recent paper, Guillaumet-Adkins et al. (2017) compared full-length and 3′ scRNA-seq data from fresh and cryopreserved cells. The authors reported that cryopreservation of cells in the presence of dimethyl sulfoxide (DMSO) causes an increase in the proportion of damaged cells when compared to freshly processed material. However, when sequencing RNA from sorted viable cells, gene expression profiles did not differ between freeze-thawed and fresh cells. The possibility of cryopreserving cells for scRNA-seq experiments increases the flexibility of experimental designs.

Cell isolation using microfluidics or FACS requires using a cell suspension as the starting material. Thus, the cells' histological context will be lost. LCM has the advantage of preserving this spatial information, but it is slow and laborious. The challenge is to develop high-throughput technologies that can provide RNA sequences from cells in known spatial locations; this is sometimes referred to as "spatial transcriptomics" (Crosetto, Bienko, & van Oudenaarden, 2014). For instance, *in situ* RNA-seq is a method in which RNA from tissue sections is sequenced inside the cell without lysis; RNA is converted to cDNA and amplified. Subsequent sequencing and image analysis are performed

using a fluorescence microscope (Ke et al., 2013).

In any sequencing experiment, it is crucial to include appropriate controls. The choice of controls depends on the biological question that is being addressed. For instance, a control could be healthy corresponding tissue from the same patient and/or a healthy donor or untreated cells.

The term "batch effect" refers to systematic gene expression differences between cells that arise during sample preparation (Stegle et al., 2015). For instance, such effects may occur if different samples are prepared on different days, by different people, and/or with different reagent lots. In bulk RNA-seq experiments, batch effects can be circumvented by preparing libraries from biological replicates in parallel and then evenly and randomly distributing them across sequencing lanes. However, in scRNA-seq experiments, this is often difficult, as cells from one condition typically are captured and prepared independently of cells from another condition. In order to reduce batch effects, the experiment can be repeated (*e.g.,* run multiple biological replicates of cells from the same condition using independent microfluidic chips or microwell plates), and one sample can be distributed across several sequencing lanes (Stegle et al., 2015). We recommend a minimum of three biological replicates per condition or group.

CONCLUSIONS

Single-cell transcriptomics holds great potential for unraveling the immense biological complexity in humans as well as in the biosphere that surrounds us. Cells are the basic functional unit of life, and scRNA-seq enables us to study them at the nucleotide level, one cell at a time. Researchers are currently producing, and will continue to produce, extensive amounts of data, and these data present an analytical challenge due to technical noise. The field is young, and there is plenty of room for improvement, particularly in transcript capture efficiency and the capture of noncoding RNAs. Protocols and algorithms are thus continuously evolving. As costs drop and throughput and sensitivity increase, single-cell technology will undoubtedly have a profound impact in basic research and medicine.

ACKNOWLEDGEMENTS

The authors are supported by grants from the Swedish Childhood Cancer Foundation, Jeanssons Stiftelser, and the Wenner-Gren Foundation. Karin Ulla Gustafsson and Jimin Guo are also gratefully acknowledged for their thoughtful input and editing.

LITERATURE CITED

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*, 166–169. doi: 10.1093/bioinformatics/btu638.

Bacher, R., & Kendziorski, C. (2016). Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biology*, *17*, 63. doi: 10.1186/s13059-016-0927-y.

Brehm-Stecher, B. F., & Johnson, E. A. (2004). Single-cell microbiology: Tools, technologies, and applications. *Microbiology and Molecular Biology Reviews: MMBR*, *68*, 538–559. doi: 10.1128/MMBR.68.3.538-559.2004.

van den Brink, S. C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C. S., ... van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nature Methods*, *14*, 935–936. doi: 10.1038/nmeth.4437.

Burrell, R. A., McGranahan, N., Bartek, J., & Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, *501*, 338–345. doi: 10.1038/nature12625.

Byrne, A., Beaudin, A. E., Olsen, H. E., Jain, M., Cole, C., Palmer, T., ... Vollmers, C. (2017). Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nature Communications*, *8*, 16027. doi: 10.1038/ncomms16027.

Crosetto, N., Bienko, M., & van Oudenaarden, A. (2014). Spatially resolved transcriptomics and beyond. *Nature Reviews Genetics*, *16*, 57–66. doi: 10.1038/nrg3832.

Chandran, D., Scanlon, M. J., Ohtsu, K., Timmermans, M. C. P., Schnable, P. S., & Wildermuth, M. C. (2015). Laser microdissection-mediated isolation and in vitro transcriptional amplification of plant RNA. *Current Protocols in Molecular Biology*, *112*, 25A.3.1–25A.3.23. doi: 10.1002/0471142727.mb25a03s112.

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., ... Quake, S. R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 7285–7290. doi: 10.1073/pnas.1507125112.

Datta, S., Malhotra, L., Dickerson, R., Chaffee, S., Sen, C. K., & Roy, S. (2015). Laser capture microdissection: Big data from small samples. *Histology and Histopathology*, *30*, 1255–1269. doi: 10.14670/HH-11-622.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*, 15–21. doi: 10.1093/bioinformatics/bts635.

Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., ... Liotta, L. A. (1996). Laser capture microdissection. *Science (New York, N.Y.)*, *274*, 998–1001. doi: 10.1126/science.274.5289.998.

Fan, J., Salathia, N., Liu, R., Kaeser, G. E., Yung, Y. C., Herman, J. L., ... Kharchenko, P. V. (2016). Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nature Methods*, *13*, 241–244. doi: 10.1038/nmeth.3734.

Frumkin, D., Wasserstrom, A., Itzkovitz, S., Harmelin, A., Rechavi, G., & Shapiro, E. (2008). Amplification of multiple genomic loci from single cells isolated by laser micro-dissection of tissues. *BMC Biotechnology*, *8*, 17. doi: 10.1186/1472-6750-8-17.

Garalde, D. R., Snell, E. A., Jachimowicz, D., Heron, A. J., Bruce, M., Lloyd, J., ... Turner, D. J. (2016). Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv*, doi: 10.1101/068809.

Grün, D., & van Oudenaarden, A. (2015). Design and analysis of single-cell sequencing experiments. *Cell*, *163*, 799–810. doi: 10.1016/j.cell.2015.10.039.

Guillaumet-Adkins, A., Rodríguez-Esteban, G., Mereu, E., Mendez-Lago, M., Jaitin, D. A., Villanueva, A., ... Heyn, H. (2017). Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biology*, *18*, 45. doi: 10.1186/s13059-017-1171-9.

Guo, G., Luc, S., Marco, E., Lin, T.-W., Peng, C., Kerenyi, M. A., ... Orkin, S. H. (2013). Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell*, *13*, 492–505. doi: 10.1016/j.stem.2013.07.017.

Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J. J., ... Regev, A. (2016). Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science*, *353*, 925–928. doi: 10.1126/science.aad7038.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, *144*, 646–674. doi: 10.1016/j.cell.2011.02.013.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., ... Yanai, I. (2016). CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biology*, *17*, 77. doi: 10.1186/s13059-016-0938-8.

Hashimshony, T., Wagner, F., Sher, N., & Yanai, I. (2012). CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. *Cell Reports*, *2*, 666–673. doi: 10.1016/j.celrep.2012.08.003.

Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., ... Peng, J. (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Research*, *26*, 304–319. doi: 10.1038/cr.2016.23.

Hu, P., Zhang, W., Xin, H., & Deng, G. (2016). Single cell isolation and analysis. *Frontiers in Cell and Developmental Biology*, *4*, 116. doi: 10.3389/fcell.2016.00116.

Huang, H.-L., Hsing, H.-W., Lai, T.-C., Chen, Y.-W., Lee, T.-R., Chan, H.-T., ... Chan, H. L. (2010). Trypsin-induced proteome alteration during cell subculture in mammalian cells. *Journal of Biomedical Science*, *17*, 36. doi: 10.1186/1423-0127-17-36.

Ishii, S., Tago, K., & Senoo, K. (2010). Single-cell analysis and isolation for microbiology and biotechnology: Methods and applications. *Applied Microbiology and Biotechnology*, *86*, 1281–1292. doi: 10.1007/s00253-010-2524-4.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lönnerberg, P., & Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research*, *21*, 1160–1167. doi: 10.1101/gr.110882.110.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., ... Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods*, *11*, 163–166. Available at: Retrieved from https://www.nature.com/doifinder/10.1038/nmeth.2772 [Accessed August 6, 2017]. doi: 10.1038/nmeth.2772.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., & Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. *Nature Methods*, *10*, 857–860. doi: 10.1038/nmeth.2563.

Kharchenko, P. V., Silberstein, L., & Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nature Methods*, *11*, 740–742. doi: 10.1038/nmeth.2967.

Kim, K.-T., Lee, H. W., Lee, H.-O., Kim, S. C., Seo, Y. J., Chung, W., ... Park, W. Y. (2015). Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. *Genome Biology*, *16*, 127. doi: 10.1186/s13059-015-0692-3.

Klein, A. M. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., ... Kirschner, M. W. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, *161*, 1187–1201. doi: 10.1016/j.cell.2015.04.044.

Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The technology and biology of single-cell RNA sequencing. *Molecular Cell*, *58*, 610–620. doi: 10.1016/j.molcel.2015.04.005.

Kumar, R. M., Cahan, P., Shalek, A. K., Satija, R., Jay DaleyKeyser, A., Li, H., ... Collins, J. J. (2014). Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature*, *516*, 56–61. doi: 10.1038/nature13920.

Lake, B. B., Ai, R., Kaeser, G. E., Salathia, N. S., Yung, Y. C., Liu, R., ... Zhang, K. (2016). Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science*, *352*, 1586–1590. doi: 10.1126/science.aaf1204.

Lohr, J. G., Kim, S., Gould, J., Knoechel, B., Drier, Y., Cotton, M. J., ... Golub, T. R. (2016). Genetic interrogation of circulating multiple myeloma cells at single-cell resolution. *Science Translational Medicine*, *8*, 363ra147. doi: 10.1126/scitranslmed.aac7037.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550. doi: 10.1186/s13059-014-0550-8.

Macosko, E. Z. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., ... McCarroll, S. A. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, *161*, 1202–1214. doi: 10.1016/j.cell.2015.05.002.

Mahata, B., Zhang, X., Kolodziejczyk, A. A., Proserpio, V., Haim-Vilmovsky, L., Taylor, A. E., ... McCarroll, S. A. (2014). Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis. *Cell Reports*, *7*, 1130–1142. doi: 10.1016/j.celrep.2014.04.011.

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., ... Linnarsson, S. (2016). Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell*, *167*, 566–580.e19. doi: 10.1016/j.cell.2016.09.027.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., ... Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*, 376–380. doi: 10.1038/nature03959.

Marques, S., Zeisel, A., Codeluppi, S., van Bruggen, D., Mendanha Falcao, A., Xiao, L., ... Castelo-Branco, G. (2016). Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science*, *352*, 1326–1329. doi: 10.1126/science.aaf6463.

Miyamoto, D. T., Zheng, Y., Wittner, B. S., Lee, R. J., Zhu, H., Broderick, K. T., ... Haber, D. A. (2015). RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science*, *349*, 1351–1356. doi: 10.1126/science.aab0917.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*, 621–628. doi: 10.1038/nmeth.1226.

Nichterwitz, S., Chen, G., Aguila Benitez, J., Yilmaz, M., Storvall, H., Cao, M., ... Hedlund, E. (2016). Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nature Communications*, *7*, 12139. doi: 10.1038/ncomms12139.

Okuducu, A. F., Hahne, J. C., Von Deimling, A., & Wernert, N. (2005). Laser-assisted microdissection, techniques and applications in pathology (review). *International Journal of molecular Medicine*, *15*, 763–769. doi: 10.3892/ijmm.15.5.763.

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: Advances, challenges and opportunities. *Nature Reviews. Genetics*, *12*, 87–98. doi: 10.1038/nrg2934.

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., ... Bernstein, B. E. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (New York, N.Y.)*, *344*, 1396–1401. doi: 10.1126/science.1254257.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*, 417–419. doi: 10.1038/nmeth.4197.

Picelli, S. (2016). Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biology*, *14*, 1–14. doi: 10.1080/15476286.2016.1201618.

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., & Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods*, *10*, 1096–1098. doi: 10.1038/nmeth.2639.

Pollen, A. A., Nowakowski, T. J., Shuga, J., Wang, X., Leyrat, A. A., Lui, J. H., ... West, J. A. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nature Biotechnology*, *32*, 1053–1058. doi: 10.1038/nbt.2967.

Poulin, J.-F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J. M., & Awatramani, R. (2016). Disentangling neural cell diversity using single-cell transcriptomics. *Nature Neuroscience*, *19*, 1131–1141. doi: 10.1038/nn.4366.

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O. R., ... Sandberg, R. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*, *30*, 777–782. doi: 10.1038/nbt.2282.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*, 139–140. doi: 10.1093/bioinformatics/btp616.

Saliba, A.-E., Westermann, A. J., Gorski, S. A., & Vogel, J. (2014). Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Research*, *42*, 8845–8860. doi: 10.1093/nar/gku555.

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N. K., Macaulay, I. C., ... Göttgens, B. (2016). Resolving early mesoderm diversification through single-cell expression profiling. *Nature*, *535*, 289–293. doi: 10.1038/nature18633.

Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., ... Regev, A. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, *498*, 236–240. doi: 10.1038/nature12172.

Shekhar, K., Lapan, S. W., Whitney, I. E., Tran, N. M., Macosko, E. Z., Kowalczyk, M., ... Sanes, J. R. (2016). Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell*, *166*, 1308–1323.e30. doi: 10.1016/j.cell.2016.07.054.

Silberstein, L., Goncalves, K. A., Kharchenko, P. V., Turcotte, R., Kfoury, Y., Mercier, F., ... Scadden, D. T. (2016a). Proximity-based differential single-cell analysis of the niche to identify stem/progenitor cell regulators. *Cell Stem Cell*, *19*, 1–14. doi: 10.1016/j.stem.2016.07.004.

Siravegna, G., Marsoni, S., Siena, S., & Bardelli, A. (2017). Integrating liquid biopsies into the management of cancer. *Nature Reviews Clinical Oncology*, *14*, 531–548. doi: 10.1038/nrclinonc.2017.14.

Stegle, O., Teichmann, S. A., & Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews. Genetics*, *16*, 133–145. doi: 10.1038/nrg3833.

Tang, D. T. P., Plessy, C., Salimullah, M., Suzuki, A. M., Calligaris, R., Gustincich, S., & Carninci, P. (2013). Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. *Nucleic Acids Research*, *41*, e44. doi: 10.1093/nar/gks1128.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., ... Suraniet, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*, 377–382. doi: 10.1038/nmeth.1315.

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., ... Garraway, L. A. (2016a). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (New York, N.Y.)*, *352*, 189–196. doi: 10.1126/science.aad0501.

Tirosh, I., Venteicher, A. S., Hebert, C., Escalante, L. E., Patel, A. P., Yizhak, K., ... Suvà, M. L. (2016b). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature*, *539*, 309–313. doi: 10.1038/nature20123.

Treutlein, B., Brownfield, D. G., Wu, A. R., Neff, N. F., Mantalas, G. L., Espinoza, F. H., ... Quake, S. R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*, *509*, 371–375. doi: 10.1038/nature13173.

Trombetta, J. J., Gennert, D., Lu, D., Satija, R., Shalek, A. K., & Regev, A. (2014). Preparation of single-cell RNA-Seq libraries for next generation sequencing. *Current Protocols in Molecular Biology*, *107*, 4.22.1–17. doi: 10.1002/0471142727.mb0422s107.

Wagner, A., Regev, A., & Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. *Nature Biotechnology*, *34*, 1145–1160. doi: 10.1038/nbt.3711.

Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., ... Liu, J. (2010). MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Research*, *38*, e178. doi: 10.1093/nar/gkq622.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C., Feng, Y., ... Fan, G. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*, *500*, 593–597. doi: 10.1038/nature12364.

Yilmaz, S., & Singh, A. K. (2012). Single cell genome sequencing. *Current Opinion in Biotechnology*, *23*, 437–443. doi: 10.1016/j.copbio.2011.11.018.

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., ... Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*, 14049. doi: 10.1038/ncomms14049.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., ... Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. *Molecular Cell*, *65*, 631–643.e4. doi: 10.1016/j.molcel.2017.01.023.