Tutorial: Guidelines for the Experimental Design of Single-Cell RNA Sequencing Studies

Atefeh Lafzi[1,4], Catia Moutinho[1,4], Simone Picelli[2], and Holger Heyn[1,3]

[1] CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain

[2] Research Institute for Neurodegenerative Diseases (DZNE), Bonn, Germany

[3] Universitat Pompeu Fabra (UPF), Barcelona, Spain

[4] These authors contributed equally to this work

Correspondence should be addressed to H.H. (holger.heyn@cnag.crg.eu)

Corresponding author:
Holger Heyn
Single Cell Genomics Group
National Centre for Genomic Analysis (CNAG)
Parc Científic de Barcelona – Torre I, Baldiri Reixac, 4
08028 Barcelona, Spain
Phone: +34.934020804
Email: holger.heyn@cnag.crg.eu

Abstract

Single-cell RNA sequencing is at the forefront of high-resolution phenotyping experiments for complex samples. Although this methodology requires specialized equipment and expertise, it is now broadly applied in research. However, it is challenging to create broadly applicable experimental designs because each experiment requires the user to make informed decisions about sample preparation, RNA sequencing and data analysis. To facilitate this decision-making processes, in this Tutorial we summarize current methodological and analytical options, and discuss their suitability for a range of research scenarios. Specifically, we provide information about best practices to separate individual cells and provide an overview of current single-cell capture methods at different cellular resolutions and scales. RNA sequencing library preparation methods vary profoundly across applications and we discuss features important for an informed selection process. An erroneous or biased analysis can lead to misinterpretations or obscure biologically important information. We provide a guide to the major data processing steps and options for meaningful data interpretation. These guidelines will serve as a reference to support users in building a single-cell experimental framework from sample preparation to data interpretation, tailored to the underlying research context.

Running title

Design of Single-Cell RNA Sequencing Studies

Keywords

Single-cell genomics, transcriptomics, single-cell transcriptomics, single cell genomics, single cell transcriptomics, RNA sequencing, RNA-seq, scRNA-Seq, single-cell RNA-Seq, single cell RNA-Seq, single-cell RNA Sequencing, single cell RNA Sequencing, experimental design, practical guide, tissue digestion, sample preparation, methods, microfluidics, data analysis.

Introduction

Single-cell transcriptomics studies have dramatically improved our understanding of the complexity of tissues, organs and organisms[1]. Profiling gene expression in individual cells has revealed an unprecedented variety of cell types and subpopulations that were invisible to traditional experimental techniques. As well as providing profound insights into cell composition, single-cell studies have changed established paradigms regarding cell plasticity in dynamic processes such as development[2] and differentiation[3]. Cell states are now known to be more flexible than previously thought, and present multipotent characteristics before reaching fate decision endpoints. While various approaches are available for phenotyping individual cells (e.g. transcriptomics[4], proteomics[5] and epigenomics[6]), single-cell RNA sequencing (scRNAseq) is currently at the forefront, facilitating ever larger-scale experiments. The scalability of scRNAseq experiments has advanced rapidly through the use of automation and sophisticated microfluidics systems, producing datasets from more than a million cells[7]. As a result, experimental designs have shifted from focusing on specific cell types to unbiased analysis of entire organs[8–10] and organisms[11,12], enabling a hypothesis-free approach to explore the cellular composition of a sample.

Most scRNAseq methods are now broadly applied in both basic research and clinically translational contexts; even though they require specialized equipment and expertise in sample handling, sequencing library preparation, and data analysis. As a result, single-cell research has become one of the fastest growing fields in life science, producing fascinating new insights into tissue composition and dynamic biological processes. Large-scale scRNAseq experiments have permitted cellular maps of Caenorhabditis elegans[12], the planarian Schmidtea mediterranea[13], Drosophila[11,14] and different mouse organs[8,15] to be defined. In humans, single-cell analysis has improved our understanding of developmental processes[16], aging[17] and different diseases, such as cancer[18–21]. However, it is challenging to create generalizable designs for single-cell transcriptomic experiments because each one requires the user to make informed decisions in order to obtain interpretable results. These include selecting sample types, cell numbers and preparation methods, choosing scRNAseq techniques and sequencing parameters, and designing computational analysis strategies to generate insights from single-cell datasets. Ultimately, successful single-cell transcriptomic studies with interpretable datasets and meaningful scientific output can only be achieved using tailored experimental designs. To inform this decision-making process, in this Tutorial we provide a comprehensive description of the phases of single-cell transcriptomic studies, including 1) sample preparation, 2) single-cell RNA sequencing, 3) data processing, and 4) data analysis (as discussed further below, and see **Figure 1**). We summarize the methodological and analytical options, and highlight their suitability for distinct research scenarios to support users in designing an end-to-end experimental framework tailored to the underlying research context.

1. Sample preparation

Preparation of high-quality single-cell suspensions is key for successful single-cell studies. Irrespective of the starting material, the condition of the cells is critical to ensure efficient cell capture and optimal performance of the scRNAseq protocols (**Table 1**). While most methods use fresh viable single cells, alternatives include preserved samples[22–24] and nuclear RNA from frozen tissue[25–29]. Here, we provide common general guidelines applicable for all tissues, and optimized parameters tailored to the major tissues of interest. In principle scRNAseq applications are not restricted to specific species as long polyA-tailed RNA is present. However, some organisms might require additional processing steps to efficiently release molecules into the reactions (e.g. cell wall removal for plant material).

Good practices for sterile sample handling are recommended, including use of nuclease-free reagents and consumables. To minimize cell damage, pipetting and centrifugation should be kept to the minimum. Cell concentration and size both influence pelleting efficiency at a given centrifugation speed, time, and temperature, and a tightly-packed cell pellet may require extra pipetting, which can damage cells through shearing effects; thus, centrifugation conditions should be optimized. Sufficient volumes should be used when washing and re-suspending cells, as high concentrations can cause aggregation and clumping. Suspensions should be filtered with appropriately sized cell strainers (pore size larger than cell diameter) to remove clumps and debris. The recommended cell washing and resuspension solution is phosphate-buffered saline (calcium and magnesium free) containing bovine serum albumin to minimize cell losses and aggregation. Primary cells, stem cells, and other sensitive cell types may require washing and suspension in alternative buffers to ensure viability, which also may diminish when cells are kept in suspension for a prolonged period. Cell clumps cause automated cell counters to underestimate the effective concentration of single cells, so suspensions should be processed as soon as possible after preparation, ideally within 30 minutes. It is important to minimize cellular aggregates, dead cells, non-cellular nucleic acids, and reverse transcription inhibitors in single-cell preparations. To minimize these contaminants while maximizing the purity and the unbiased recovery of different cell types, optimization may be necessary (e.g. adjusting the number of wash steps, the composition of the wash solution, centrifugation conditions, strainer type).

Preparation of cell suspensions. To isolate single cells from suspensions (e.g. blood samples), samples are density centrifuged (for example Ficoll-Paque or Histopaque-1077)[30] and then can be used directly for single-cell capture. Solid tissues must firstly be dissociated using mechanical and enzymatic treatment. Initially, tissues are disaggregated by mechanical cutting or mincing with blades. Then, enzymatic digestion is used to separate cells, using specific enzymes and digestion times for different tissues (**Table 2**). Enzyme types include accutase, elastase or collagenases, or commercial enzymatic mixtures such as TrypLE Express or Liberase Blendzyme 3. An elevated cell lysis can lead to cell clumping, which is reduced through the treatment with DNaseI during cell separation. Finally, suspensions are cleaned by filtering samples through a mesh or strainer, before capture of single cells.

It is important to note that sample processing might introduce variation in the gene expression profile, as has been shown for the activation of stress-related genes[31]. Also, sensitive cell types might be damaged during sample preparation, so processing time should be kept to the minimum required. In contrast, too short digestion times could result in incomplete cell separation and the exclusion of tightly interconnected cells from subsequent single-cell analysis.

To avoid biases in cell type composition an alternative strategy involves disrupting cellular membranes and isolating the nuclei[25–29]. The sequencing of nuclear RNA was shown to be sufficient for deconvoluting cell types[29], although this diminished the overall resolution per cell. Single-nuclei sequencing has been performed extensively for differentiated neurons, for example, as it is largely impracticable to isolate intact cells from the highly interconnected adult neuronal tissue.

Single-cell capture. To profile the transcriptomes of single cells, most methods require the physical isolation of cells into individual reaction volumes. Cells can be isolated by microdissection or pipetting[32], although high-throughput experiments use fluorescence-activated cell sorting (FACS)[33] or microfluidics[34] to guide cells into micro- or nanoliter reaction volumes, respectively. Microfluidic systems capture cells in integrated fluidics circuits (IFC), droplets or nanowells, allowing thousands of cells to be processed simultaneously while minimizing reaction volumes and reagent use. FACS sorts cells into microtiter plates ready for library preparation by manual or automated processing, and facilitates the exclusion of dead or damaged cells, and the enrichment of target cell populations (e.g. through surface marker labelling). To reduce background and maximize assay performance, we also recommend FACS or magnetic-activated cell sorting (MACS) processing of single-cell solutions for microfluidic systems, to remove debris, damaged/dead cells and cell aggregates.

Sample size and composition. To obtain an unbiased view of the cellular composition of a sample, all cells need to be captured during the isolation process. Here, attention must be paid to very small and large cells that may be excluded during FACS isolation or captured in microfluidic systems, respectively. However, for many experiments it may be necessary to enrich for or exclude some cell types to increase the total number of cells of interest in the final scRNAseq libraries. For example, profiling specific immune responses requires enrichment of blood cell subtypes, while cancer studies might need to exclude blood cells (e.g. CD45 positive) to increase the overall number of tumour cells. Target populations can be selected using FACS and MACS using appropriate labelling (e.g. antibodies or transgenic systems). Microtiter plates and some nanowell capture systems allow index-sorting, where fluorescence intensity or cell size (FACS information) is associated to capture coordinates and subsequently to single-cell indices. Here, the FACS device records the sorting position and intensity values of a given cell, enabling the subsequent integration of transcriptome profiles with the recorded cell properties. For microfluidic systems, CITE-seq[35] provides a viable alternative to conserve the information of surface markers. Here, epitopes of interest are targeted with oligonucleotide-labelled antibodies. The antibody-specific sequences are poly-A tailed and contain barcodes that allow epitope tracking after scRNAseq library preparation and sequencing.

To define adequate cell numbers per experiment, one must consider sample heterogeneity and sub-population frequency (estimated abundance of the cell type of interest). In particular, larger cell numbers are required to resolve the structure of heterogeneous samples with many expected subpopulations. Also the total number of cells required increases when rare cell types need to be identified. The required cell numbers can be calculated by estimating both sub-population structure and low-frequency cell type abundance and defining the desired cell number per group (computational tool accessible at: https://satijalab.org/howmanycells). Since most experiments target poorly described systems, heterogeneity can only be estimated, so pilot experiments are recommended before entering large-scale data production. In line, in comparative studies across experimental conditions, patient samples or larger population cohorts, control experiments can inform about optimal cell numbers and the need for sub-population enrichment steps. Note that higher cell numbers can also be beneficial for homogenous samples, as this increases statistical power during analysis[36].

Sample preservation. All common scRNAseq methods were initially designed to use freshly isolated cells. However, in research and clinical practice, immediate sample processing can be challenging due to a lack of the required infrastructure or specialized equipment, such as FACS devices. Moreover, although samples may be collected at multiple time points, simultaneous sample processing may be preferred to avoid technical batch effects. Sample preservation is a viable solution because it disconnects the location and time of sampling from downstream processing steps. In this context, cryopreservation has been established for single-cell transcriptome analysis[22]. After sample storage at -80 ºC or in liquid nitrogen and thawing, cryopreserved cells from cell lines and primary samples show complete integrity of the RNA molecules and unchanged expression profile as compared to freshly prepared cells. Note that multiple freeze-thaw cycles should be avoided through the preparation of aliquots or by scraping out cells from frozen vials. Similarly, methanol fixation has been established as an alternative for droplet-based single-cell methods, which could also avoid technically induced variations in gene expression triggered by prolonged sample processing time[23]. Importantly, both methods allow the archiving and transport of samples and broaden the range of applications of scRNAseq methods, for example to the clinical context. However, both approaches detected a potential bias in cell type composition and it is strongly recommended to thoroughly evaluate preservation methods for new cell types that have not been tested. For previously archived samples, such as snap-frozen specimens, nuclei sequencing provides the only solution for scRNAseq[25–29]. In contrary to cryopreservation, the formation of ice crystals during snap-freezing disrupts the outer cellular membrane, however, the nuclei remain intact. Nevertheless, it is preferable to make an initial estimation of the RNA integrity to avoid biases related to the sample quality.

2. Single-cell RNA sequencing

Transcriptome profiling of individual cells can be split into four major components: RNA molecule capture, reverse transcription and transcriptome amplification, sequencing library preparation, and single-cell RNA sequencing. Various scRNAseq methods exist, but all apply the same underlying principles. Below we discuss these basic experimental design considerations, and highlight common and emerging microtiter plate-based and microfluidic scRNAseq techniques and their applications. Key features of the different scRNAseq approaches discussed below are also summarized in **Table 3**. Many of these methods have been systematically evaluated, confirming their generally high accuracy, although efficiency, scalability and costs vary significantly[37,38]. This should be taken into account when selecting methods for a given experiment.

RNA molecule capture, reverse transcription and transcriptome amplification for sequencing library preparation. Most scRNAseq methods, including those described below, capture polyA-tailed RNA, although specific protocols are available for profiling total RNA[39,40] or miRNAs[41]. After cell lysis, polyA-tailed RNA is captured by polyT-oligonucleotides, which excludes abundant RNA types such as rRNA or tRNA. Following capture, the RNA is reverse-transcribed into stable cDNA, at which point most methods add single-cell-specific barcodes within the polyT-oligonucleotides that allow cost-effective multiplexed processing of pooled samples. Moreover, random nucleotide sequence stretches in the polyT-oligonucleotide serve as unique molecule identifiers (UMI) that allow the user to correct for amplification biases and reduce technical noise[42]. Reverse transcription (RT) is a crucial step, and different protocols have been optimized in various ways using efficient enzymes and specific additives that maximize efficiency (**Box 1**). cDNA can then be amplified by PCR or through *in vitro* transcription (IVT). To enable this, adapter sequences or RNA polymerase promoter sequences are introduced during RT or second strand synthesis. Although IVT is less prone to biases through linear amplification of molecules, it requires additional downstream steps to convert the amplified RNA into cDNA and sequencing-ready libraries. PCR-based protocols however, require less hands-on time, but the exponential amplification phase leads to biases in RNA composition in the final libraries. Both approaches were shown to results in interpretable results and were successfully implemented in several scRNAseq methods (Table 3).

Full-length vs 3'- or 5'-end transcript sequencing. Single-cell transcriptome profiling can be performed using full-length transcript analysis, or by digital counting of 3´- or 5´-transcript ends[42]. The choice of sequencing method would be dictated by the goal of the experiment, i.e. prioritizing cost-effectiveness versus retaining sequence information. Digital RNA counting is a cost-effective quantification strategy, although sequence information of the transcripts is lost. Full-length transcriptome sequencing allows the detection of splice variants and alternative transcripts, as well as genetic alterations in the transcribed fraction, such as single-nucleotide variants[19,43] or fusion transcripts[44]. Moreover, T- and B-cell receptor genotypes can be obtained from full-length transcriptomes[45]. Unlike 3´- and 5´-end methods, full-length protocols do not allow the introduction of UMI and impede an early cellular barcoding and pooling, resulting in higher costs for library preparation. This limitation can be overcome using long-read sequencing technologies that do not need library fragmentation[46]. However, such technologies generate lower quantities of sequencing reads and transcriptome quantification is not yet possible.

scRNAseq methodologies: Microtiter plate-based approaches. After isolating single cells into microtiter plates by FACS, a full-length transcript or 3`/5`-end protocol can be applied. **Smart-seq2**[47] is a widely-used method to reverse transcribe and amplify full-length transcripts. Following RT the enzyme adds cytosines to the cDNA, providing the basis for a template switching reaction. Here a template switching oligonucleotide (TSO) binds to the extra cytosine and provides the template for the addition of PCR adapter sequences for subsequent cDNA amplification. Compared to the original version[48], the updated protocol improves molecule capture efficiency and yield by using locked nucleic acids in the TSO and adding betaine to the RT reaction. Sequencing libraries are prepared using tagmentation, simultaneously fragmenting and indexing the cells. The Smart-seq2 protocol is highly efficient in capturing RNA molecules[37], although the late indexing step makes it more expensive than other methods. Furthermore, the absence of UMI makes downstream data analysis more challenging. Nevertheless, the protocol provides an adequate solution if deep single-cell phenotyping is required (e.g. for homogeneous samples or for analysing weakly expressed genes).

STRT-seq[49] uses a similar strategy for RT and template switching, but incorporates single-cell barcodes in the TSO. This allows early pooling of cells and cost-effective multiplex processing. STRT-seq enriches 5'-transcript ends using biotinylated purification and 5'-specific PCR primers. Analysing the 5'-transcript has the advantage of providing information about transcription start sites. Moreover, cell barcodes and transcripts are obtained in a single read, enabling cost-effective single-end sequencing. While the original STRT-seq protocol could not correct for amplification biases, later updates for the first time included UMI in a scRNAseq method[42]. The **SCRB-seq**[50] protocol incorporates single-cell barcodes and UMI in the polyT-primer, enabling 3' amplification of transcripts, and like STRT-seq, early indexing allows cell pooling to reduce costs. The RNA capture efficiency of the original protocol was improved by increasing the RT mix density: molecular crowding SCRB-seq (**mcSCRB-seq**[51]) includes polyethylene glycol to increase binding event probabilities. In addition, the PCR enzyme was switched from KAPA to the Terra polymerase to further improve library complexity. In **Quartz-seq**[52] the template-switching reaction is replaced by a polyA-tailing step. The additional adenosines provide a template for a polyT-primed second strand synthesis, followed by PCR amplification. The amplified transcriptome then undergoes ultrasound fragmentation and sequencing adapter ligation. A later version, **Quartz-seq2**[53], improved the molecule detection efficiency by using shorter RT primers and improving polyA-tagging efficiency.

Amplification biases during exponential PCR are addressed in **CEL-seq**[32], where transcripts are copied through IVT. The linear amplification of molecules enabled by inclusion of a T7 promoter in the polyT primer results in more evenly duplicated transcriptomes. Also, transcriptome amplification by IVT does not require template switching, which improves molecule capture efficiency. This workflow was further optimized in **MARS-seq**[54] by including UMI in the polyT primers and allowing upscaling of cell numbers through automation. Also, the original CEL-seq protocol was updated in **CEL-seq2**[55] for more efficient RNA capture and a simplified workflow. Briefly, the CEL-seq2 protocol uses UMI, a shorter RT primer, and more efficient RT and second strand synthesis enzymes. Furthermore, cDNA synthesis following IVT is initiated by random priming instead of adapter ligation.

scRNAseq methodologies: microfluidic systems-based approaches. Microfluidics allows higher throughput scRNAseq workflows, eliminating the technical constraints on scalability associated with using microtiter plates. Moreover, reducing reaction volumes from micro- to nanoliters reduces costs and technical variability[56], while improving cDNA yield[57]. There are three strategies for capturing cells; IFC, droplets, and nanowells, all of which increase the number of capture sites relative to microtiter plates. The first microfluidics system used for scRNAseq was designed as an automated array solution (**Fluidigm C1**), where single cells enter a fluidics circuit, are immobilized in hydrodynamic traps, lysed, and processed in consecutive nanoliter reaction chambers using a modified Smart-seq2 protocol. While early versions could only use commercial scRNAseq assays, a more recent open format accommodates custom scRNAseq protocols[42] and additional applications for genetics and epigenetics single-cell experiments[58]. Costs were further reduced by increasing throughput and cell capture from 96 to 800 sites (**C1 HT-IFC**), and including an early-indexing strategy that allows cell pooling. Notably, this high-throughput version switched from full-length to 3’ RNA sequencing. Also, the array formats restricted to specific cell sizes (small, medium and large arrays) affects the unbiased sampling from complex sample types. To further increase cell numbers, microfluidics progressed to open nanowell systems that allow better scalability. In **STRT-seq-2i**[59], the original protocol was applied in a nanowell platform with 9,600 sites, with cells loaded by limiting dilution or direct addressable FACS sorting. Positioning cells by FACS allows index sorting that assigns cell properties (e.g. fluorescence signal or size) to array coordinates and barcodes. Nanowells containing cells can be specifically utilised by targeted dispensing, significantly reducing reagent costs and contamination of ambient RNA. Moreover, the array format allows imaging to exclude doublets. To guarantee high cell viability during the time-consuming loading into nanowells, FCS can be added to the buffer and sample aliquots kept on ice. Alternatively, **Seq-Well**[60] provides a nanowell-based method that captures cells in 86,000 sub-nanoliter reactions. The underlying principle is the pre-loading of nanowells with barcoded beads before cells enter the capture sites through limited dilution. Subsequently, the arrays are sealed for cell lysis, RNA molecule capture on beads, before the immobilized molecules are pooled for 3`-end library production. The Seq-Well system is portable, and so allows sample processing at the sampling sites, as large equipment is not required. The fact that no major investments are required, makes the Seq-Well system a flexible and cost-effective alternative. However, while cells can be monitored by microscopy, the random distribution of barcoded beads does not allow the user to integrate imaging data. Also, the method requires experienced users to obtain reproducible high-quality results.

While scalable to higher throughputs, IFC and nanowell approaches are intrinsically constrained by the number of reaction sites. Droplet-based systems overcome this by encapsulating cells in nanoliter microreactor droplets. Here, cell numbers scale linearly with the emulsion volume, and large numbers of droplets are produced at high speed, facilitating large-scale scRNAseq experiments. Furthermore, droplet size can be adjusted to reduce potential biases during cell capture. Since barcodes are introduced into droplets randomly, this approach does not allow the assignment of cell barcodes to images and so precludes the visual detection of doublets and the integrative analysis of cell properties (e.g. fluorescence signal) with transcriptome profiles. Two droplet-based methods, inDrops[61] and Drop-seq[62], were developed in parallel, with related commercial systems allowing straightforward implementation. **inDrops**[61,63] encapsulates cells using hydrogel beads bearing polyT-primers with defined barcodes, after which the photo-releasable primers are detached from the beads to improve molecule capture efficiency and initiate in-drop RT reactions. The barcoded cDNAs are then pooled for linear amplification (IVT) and 3'-end sequencing library preparation. The technique has extremely high cell-capture efficiency (>75%) due to the synchronized delivery of deformable beads, allowing near-perfect loading of droplets. Therefore, the system is most suitable for experiments with limited total numbers of cells. The inDrops system is licensed to 1CellBio and a variant protocol is commercialized as **Chromium** Single Cell 3' Solution (10x Genomics)[64]. The Chromium system is straightforward to implement and standardize, although library preparation costs are significantly higher than in the original system. Unlike inDrops protocols, **Drop-seq**[62] uses beads with random barcodes. Following cell lysis and RNA capture, the drops are broken and pooled, covalent binding is performed through cDNA synthesis, the cDNA is amplified by PCR, and 3´-end sequencing libraries are produced by tagmentation. Drop-seq has lower cell capture efficiency than inDrops methods because beads and cells are delivered by double limiting dilution (double Poisson distribution), which results in 2-4% barcoded cells. The Drop-seq system is commercially available through Dolomite Bio and a similar system is provided by Illumina (**ddSEQ**).

scRNAseq methodologies: Split-pool barcoding-based approaches. Conceptually different from the above techniques are methods based on combinatorial barcoding. Here, cells are not processed as individual units but isolated in pools. These pools are split and mixed, with each round integrating pool-specific barcodes. The combination of such pool indices results in unique barcode combinations for each cell through their random assignment during consecutive pooling processes. Both split-pooling methods, SPLiT-seq[65] and sci-RNA-seq[12], were shown to reliably produce single-cell transcriptomes and to be scalable to hundred-thousands of cells per experiment. **SPLiT-seq** includes four rounds of indexing resulting in >20 million possible barcode combinations. Following initial indexing during reverse transcription, two rounds of index ligation and a final PCR indexing step create cell-specific barcoded 3`-transcript libraries. During the second ligation round UMIs are incorporated for the subsequent correction of amplification biases. Additional rounds of barcoding or switching from 96- to 384-well microtiter formats could further scale-up cell numbers. The original **sci-RNA-seq** protocol includes a two-step indexing workflow with the first index and UMI introduced during reverse transcription and a second index during PCR amplification (following tagmentation). The use of indexed tagmentation sequences could further scale-up possible barcode combinations and increase cell numbers per experiment. The formaldehyde- or methanol-based fixation of cells, used in SPLiT-seq and sci-RNA-seq respectively, allows sample storage, providing additional flexibility to the experimental designs. Both methods allow the processing of nuclei and consequently the analysis of more challenging cell types, such as neurons. The split-pool strategy employed in sci-RNA-seq was further shown to be applicable in different single-cell epigenomic analysis approaches, including open chromatin (sci-ATAC-seq[66]), chromatin conformation (sci-Hi-C[67]) or DNA methylation (sci-MET[68]).

Library preparation and sequencing. To prepare libraries for short-read sequencing applications the amplified cDNA (PCR) or RNA (IVT) is fragmented before sequencing adapters are added. Fragmentation can be achieved enzymatically (tagmentase, DNase), chemically (Zinc, KOAc and MgOAc) or through mechanic forces (ultrasound) (Table 3). 3'- or 5'-based libraries are subsequently amplified using primers specific for the transcript end or start, respectively. During this step of the protocol, a pool specific index can be introduced that allows the multiplexed sequencing of multiple experiments. Full-length methods introduce the cell specific barcodes only after fragmentation, thus impeding a pooled processing of cells at earlier stages of the protocol. Apart from STRT-seq, scRNAseq libraries require paired-end sequencing, where one read provides information about the transcripts while the other reads the single-cell barcodes and UMI sequences. STRT-seq incorporates the cell barcode and UMI at the 5'-transcript end, allowing the capturing of cell, molecule and transcript information in a single read, since no polyT stretch separates the respective sequences. High-throughput microfluidic-based experiments generally involve sequencing to lower depth (<100,000 reads/cell), while higher read numbers (~500,000 reads/cell) are optimal for many microtiter-plate formats[38]. Nevertheless, single-cell libraries are usually not sequenced to saturation and the phenotyping resolution (detection of more genes, and those expressed at lower levels) can benefit from further increasing the sequencing depth. Annotating splice variants from full-length transcriptomes requires deeper sequencing to better resolve the expression levels of transcript variants.

Further technical considerations

Cell doublets. An intrinsic problem for most microfluidics-based methods is that two cells may be captured per reaction site (nanowell or droplet), both receiving identical barcodes. Doublet rates can be experimentally determined in species mixture experiments, but otherwise can only be estimated. They occur when cells are positioned randomly in reaction sites by limiting dilution and can be controlled by the cell suspension concentration. The relationship between cell loading and doublet rate was systematically quantified for the Chromium system[64]. Up to the maximal recommended loading of 10,000 cells per droplet-lane, the doublet rates showed a linear relationship (in line with the Poisson loading of cells into droplets), with inferred rates ranging from 2% (2500 cells) to 8% (10,000 cells). Other microfluidics approaches report similar numbers; Drop-seq 0.36% to 11.3% (12.5 cells/µl to 100 cells/µl)[62], InDrops 4%[61], Seq-Well. 1.6%[60]. The doublet rate decreases at higher dilutions, with a resulting increase in reagent costs per cell, as fewer total cells are captured per experiment. This handicap can be partially overcome by jointly capturing samples from different individuals, where genotype differences allow the user to distinguish between donors and thereby reliably identify doublets[69]. Specifically, single nucleotide polymorphisms identified from the RNA sequencing reads are utilized to determine the donor origin of the cells and to discriminate samples that were processed in a single batch. However, such workflow is only practicable when the experimental design includes different human individuals or model organisms with distinct genetic backgrounds. Currently, there is no computational method for credibly identifying doublets, so doublet rates must be minimized by experimental design. Doublets can have dramatic consequences for data interpretation, as artefactual mixed transcriptomes can easily be mistaken for intermediate cell states in dynamic systems.

Cell capture efficiency. Cell capture efficiency is an important consideration, especially when working with primary or rare samples. The number of cells that receive barcodes is directly related to the proportion of sample that enters downstream analysis. The capture efficiency of FACS-based methods is constrained by the time the device requires to move between wells. To maximize capture rates of FACS based methods, cell suspensions can be diluted and sorted at low speed (e.g. 100 cells/second). Microfluidics technologies differ markedly in capture efficiency, mainly due to cell and bead loading mechanics. The HT-IFC system captures a maximum of 800 out of 6,000 injected cells. In nanowell systems that use limiting dilution for cell loading (no sorting), cells enter reaction sites by gravity with generally high efficiency. For example, 10,000 cells are added to the surface of a Seq-Well array, and around 3,000 cells are captured. For droplet-based systems, the rate at which cells enter the analysis is directly related to the loading efficiency of the beads. Where most droplets contain barcoded beads, cell capture is optimal (inDrops). In contrast, if beads and cells are encapsulated by limiting dilution, most cells do not enter a bead-containing droplet, resulting in lower capture efficiency (Drop-seq; see further discussion above).

Costs. The total cost of scRNAseq experiments is determined by three main components: Equipment, reagents and sequencing. For most methods, the cost of scRNAseq library preparation scales linearly with cell numbers, except with custom droplet methods. The actual costs per cell vary widely across methods and institutes, with microfluidic systems being generally cheaper (<0.30 USD/cell) than early-indexing plate-based 3'-digital counting methods (~1-2 USD/cell). Late-indexing full-length transcriptome profiling is more costly even in small volumes (~8-12 USD/cell). However, costs can be reduced using non-commercial tagmentase[70] or by minimizing reaction volumes and using automated workflows for plate-based formats[71]. Importantly, microtiter plates can be shipped and stored, thereby disconnecting sampling sites from scRNAseq processes such that expensive devices can be centralized in core units, optimizing resource management. Custom microfluidics methods further decrease costs per cell. Commercialized microfluidic methods are more expensive (0.5-2.0 USD/cell) than custom systems (<0.30 USD/cell), although their automated design reduces hands-on time and personnel costs. While the cost of library preparation is decreasing rapidly, sequencing costs become a major factor. Methods with higher molecule capture efficiency produce more complex sequencing libraries, making them informative at lower sequencing depth. Consequently, more efficient scRNAseq methods can compensate for higher library preparation costs, by decreasing overall sequencing costs.

3. Data processing

Data processing includes all the steps necessary to convert raw sequencing reads into gene expression matrices, following similar workflows to those used for bulk-RNAseq. After generating fastq reads and checking their quality (with tools such as FastQC; www.bioinformatics.babraham.ac.uk/projects/fastqc/), the next important step is to de-multiplex reads using cell barcodes. While only Smart-seq libraries can be directly de-multiplexed using the index reads, the 3`-end based methods require a dedicated processing step to identify the single-cell indexes in the sequencing reads. De-multiplexed reads are then mapped to reference genomes using alignment tools such as TopHat[72] or STAR[73], the latter showing proven accuracy and splice variant awareness. Recent alignment tools were optimized for fast handling of large-scale datasets without losing accuracy. For example, Kallisto[74] reduces the alignment time by two orders of magnitude through pseudo-alignment, compared to aligning individual bases. In a final processing step, mapped reads are quantified to create a transcript expression matrix. RSEM[75], Cufflinks[76] and HTSeq[77] can be used for full-length transcript datasets, while special tools are available for counting UMI-tagged data types, such as UMI-tools[78], which accounts for sequencing errors in UMI sequences.

In addition to the specific tools available for individual processing steps, single cell data processing pipelines have been developed that combine mapping and quantification steps, and include quality control measures for reads and cells. A pipeline developed by Ilicic *et al*. supports various mapping and quantification tools, and includes modules for filtering low quality cells[79]. Scater provides an organized workflow for converting raw sequencing reads into a 'single-cell expression set' (SCESet) class, a data structure that facilitates data handling and analysis[80]. Other available pipelines give either protocol-specific solutions (e.g. zUMI[81], scPIPE[82] and SEQC[83] for UMI data) or are technology-specific (e.g. Cell Ranger for Chromium systems). The scRNA-tools database (www.scRNA-tools.org) provides a comprehensive list of available computational tools for data processing and analysis[84]. Methods are categorized by analysis tasks and researchers can select tools according to the required analysis type.

Normalization. Single-cell RNAseq datasets show high levels of noise and variability related to non-biological technical effects, including dropout events due to stochastic RNA loss during sample preparation, biased amplification, or incomplete library sequencing. Technical variation also results from batch effects on processing units (e.g. plates or arrays), time points, facilities, and other sources. Moreover, natural variability complicates analysis because of, for example, variable cell size and RNA content, different cell cycle stages, and gender differences. Therefore, dataset normalization becomes an important step for meaningful data analysis. This can be guided by adding artificial spike-in RNA, which is used to model technical noise, as implemented in BASiCS[85]. However, it is not clear whether artificial RNA sufficiently reflects the behavior of endogenous RNA, and whether cellular RNA influences spike-in detection. Recent high-throughput methods distribute cells by limiting dilution, making the use of spike-in RNA impracticable due to the high number of otherwise empty reaction volumes. Alternative normalization methods originally developed for bulk-RNA sequencing, such as log-expression[86], trimmed mean M-values[87] or upper-quartiles[88] can also be used in scRNA-Seq, although more specialized normalization methods are being developed that can better handle many aspects of this specific type of data. Recent single-cell approaches perform between-sample normalization (SCnorm[89]) or normalize on cell-based factors following pool-based size factor deconvolution (SCRAN[90]). However, to correct for large-scale sources of variation, a recommended and standard procedure is to model the data with the correct distribution. Here, confounding factors can be incorporated as covariates into the model, and regressed out. While batch effects are usually detected by visual inspection of reduced-space representations (e.g. principal components), kBET[91] is a batch effect test based on k-nearest neighbors. It quantitatively measures batch effects within and between datasets without directly correcting the data. This approach concludes that a combination of log normalization or SCRAN pooling with ComBat[92] or limma[93] regression provides the best batch-corrected dataset while preserving the biological structure. The batch effect problem becomes magnified when integrating datasets from different time points, individuals or scRNAseq methods. In this regard, Haghverdi *et al*. propose an approach based on mutual nearest neighbors (MNNs), where a shared subset of populations is sufficient to correct for batch effects across experiments, although predefined or equal population compositions are required[94]. Alternatively, by inferring cell clusters from gene expression similarities and co-expression patterns, Biscuit (Bayesian Inference for Single-cell ClUstering and ImpuTing)[83] identifies and corrects for technical variation per cell. Also, the commonly used scRNAseq package Seurat provides a solution for integrating datasets based on common sources of variation[95], with a new feature enabling the identification of shared populations, and facilitating comparative analysis across datasets.

Imputation and gene selection. In addition to having a high noise level, scRNAseq datasets are also very sparse, further challenging cellular phenotyping and data interpretation. Non-expressed genes and technical shortcomings, such as dropout events (not sequenced transcripts), result in many zeros in the expression matrix and, thus, an incomplete description of a single cell's transcriptome. To reduce sparsity missing transcript values can be computationally inferred using imputation, for example using MAGIC[96], which uses diffusion maps to find data structures and restore missing information. Alternatively, scImpute[97] learns a gene's dropout probability by fitting a mixture model, and then imputes probable dropout events by borrowing information from similar cells (selected based on genes that are not severely affected).

A common strategy for determining heterogeneity in a sample is to analyse highly variable genes across datasets. A thorough feature selection step to remove uninformative or noisy genes increases the signal-to-noise ratio but also reduces the computational complexity. Commonly used strategies for extracting variable genes in scRNAseq tools exploit the relationship between the mean transcript abundance and a measure of dispersion such as the coefficient of variation[98], the dispersion parameter of the negative binomial distribution[99], or the proportion of total variability[85].

4. Data analysis

Some of the major applications of scRNAseq experiments include assessing sample heterogeneity and identifying novel cell types and states. This is achieved by determining co-expression patterns and clustering cells by similarity. Cells clusters can subsequently be interpreted by annotating gene sets that drive clusters (marker genes). A common way to visually inspect cellular subpopulation structures is to perform dimensionality reduction (DR) and to project cells into a two or three dimensional space. Principal component analysis (PCA) and t-distributed stochastic neighbour embedding (t-SNE) are commonly used approaches for data representation[100,101]. Diffusion components[102] and uniform manifold approximation and projection (UMAP)[103] are viable alternatives that overcome some limitations of PCA and t-SNE by preserving the global structures and pseudo-temporal ordering of cells as well as being faster[104]. Even though DR techniques can guide the initial data inspection; more robust clustering algorithms are needed to define subpopulations among cells.

While prior assumptions and canonical population markers allow supervised clustering (e.g. with Monocle2[105]), hypothesis-free unsupervised clustering is preferred in most cases. A commonly used unsupervised algorithm is hierarchical clustering, which provides consistent results without a pre-defined number of clusters. Hierarchical clustering can be performed in an agglomerative (bottom-top) or divisive (top-bottom) manner, with consecutive merging or splitting of clusters, respectively. Tools such as PAGODA[106], SINCERA[107] or bigSCale[7] implement hierarchical clustering. Another suitable unsupervised clustering algorithm is K-means, which estimates K centroids (centre of the clusters) and assigns cells to the nearest centroid, re-computes centroids based on the mean of cells in the centroid clusters, and then reiterates both steps. SC3, for example, integrates both K-means and hierarchical clustering to provide accurate and robust clustering of cells[108]. Other unsupervised approaches, like SNN-Cliq[109] and Seurat[95], use graph-based clustering, which builds graphs with nodes representing cells and edges indicating similar expression, and then partitions the graphs into interconnected 'quasi-cliques' or 'communities´. Clustering can be performed directly on expression values or further processed data types, such as principal components or similarity matrices, the latter showing improved yield in cluster separation. Cluster stability is measured using resampling methods (e.g. bootstrapping) or by measuring cell similarities within assigned clusters (e.g. silhouette index). To support cluster reproducibility, different algorithms can be compared using adjusted rand indexes[108]. Clusters can be represented by color coding cells in a low dimensional space produced by the DR algorithms discussed above (e.g. PCA or tSNE).

Marker genes that discriminate subpopulations can be identified by performing differential gene expression analysis between clusters using, for example, model-based approaches such as SCDE[110], MAST[111] or scDD[112], which account for data bimodality using a mixture model. Individual genes can be evaluated to serve as a binary classifier for cell identity using, for example, ROC or LRT tests based on the zero-inflated data[95,108]. A recent publication comprehensively compared differential expression analysis methods for scRNAseq and can guide the selection of appropriate differential expression tools[113].

Another important application of scRNA-seq is trajectory inference which estimates dynamic processes by ordering cells along a predicted differentiation path (pseudotime) using algorithms such as reversed graph embedding (Monocle2[114]) or minimum spanning tree (TSCAN[115]). Also, trajectory inference methods have been comprehensively benchmarked, testing their accuracy and overall performance[116]. To further facilitate the interpretation of the results, tools such as SCENIC[117] provide the opportunity to investigate active regulatory networks in subpopulations of cells. The analysis guides the identification of active transcription factors, eventually providing insights into the cellular mechanisms that drive heterogeneity. For cluster annotation, scmap facilitates comparison of data across experiments by projecting cells from one dataset onto cell types or individual cells from another scRNA-seq experiment[118]. With cell convolution tools, such as bigSCale[7], scRNAseq analysis can be expanded to millions of cells. Eventually, single cells can be mapped back to the spatial tissue context using experimental approaches[119,120] or pseudo-spatial ordering of cells[9,95,121].

To make scRNAseq data publically available, data storage and sharing repositories can be utilized. The Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/gds) is commonly used to provide access to raw data and further processed formats, such as gene expression quantification matrices. Large-scale projects, such as the Human Cell Atlas, set up specific data coordination platforms to further ease data query and accessibility. For data analysis, many researcher provide free open access to their computational pipelines through public databases, such as GitHub (https://github.com/) or as ready-to-use packages available through, for example, Bioconductor (https://www.bioconductor.org/).

Summary

Although it is challenging to define broadly applicable designs for single-cell RNA sequencing experiments, we here provide general guidelines to support the production of high quality datasets and their meaningful interpretation. A thoroughly planned and conducted sample preparation is crucial to preserve cellular and RNA integrity and the unbiased representation of the sample composition. The selection of downstream scRNAseq techniques is driven by the complexity of the underlying sample and the desired resolution per cell. While large cell numbers, processed in microfluidic systems, might represent better the composition of heterogeneous samples, an in-depth analysis of smaller sample sizes could be more appropriate to resolve subtle differences in homogenous mixtures. Budget restraints and reduced library complexity generally lead to the shallow sequencing of high cell numbers, whereas cell type focused experiments with sensitive methods can benefit from deeper sequencing. Eventually, the analysis and interpretation of single-cell transcriptomes is enabled by a wealth of computational methods, specifically tailored to answer biological questions hypothesis-free or guided by previous knowledge. Despite technical challenges scRNAseq experiments are a powerful tool to fully resolve sample heterogeneity and dynamic cellular systems or to identify perturbation effects at high resolution.

Future directions of the single-cell field.

Single-cell transcriptomics technologies are advancing rapidly. Cell numbers that can be analysed are increasing to hundreds of thousands of cells per experiment, markedly improving statistical power and resolution for detecting rare and transient cell types. However, high-throughput techniques come with the expense of decreased molecule capture rates, and future methods need to better balance cell numbers with cell resolution. This will be accompanied by decreasing sequencing costs, eventually enabling comprehensive, high-resolution snapshots of complex tissues to be achieved. Today, tissue and organism cell atlas projects perform sky-dive experimental strategies, initially creating a low-resolution atlas with thousands of cells to estimate sample heterogeneity, and then zoom in on target cell types using efficient scRNAseq methods to achieve higher per-cell resolution. In future, high-resolution maps will allow users to zoom in on the existing data, circumventing costly and time-consuming sample reprocessing. Microfluidics methods have already driven a paradigm shift in experimental designs, and conceptually different alternative methods such as combinatorial barcoding[12,65] might push the barrier back even farther. By not needing to physically separate individual cells, this approach enables cost-effective parallel processing of cells, which will allow cell numbers to be scaled up even further.

An additional future avenue of intense investigation will be based on advances in monitoring transcriptional profiles in spatial contexts. scRNAseq relies on disconnecting cells from their natural environment, but spatial methods, including *in situ* sequencing[122] and single-molecule (smFISH[119]) or multiplexed error-robust (MERFISH[120]) fluorescence *in situ* hybridization, profile gene expression within the tissue context. Although current methods have low transcriptome resolution or require prior marker selection, they were shown to be extremely powerful in resolving tissue complexity[9,123]. Future spatial methods should allow the field to advance from the current combinatory experimental designs[124], or pseudo-space analysis[95,125], to a full tissue expression profile in three dimensions. Eventually, phenotype heterogeneity and dynamics in living multicellular systems will be resolved by the fusion of unbiased transcriptome profiling in spatial and temporal dimensions and the combined profiling of additional layers of molecular information, such as genetic variation[126] or gene regulatory marks (e.g. DNA methylation[127] or open chromatin[128]), from the very same cell.

Acknowledgements

HH is a Miguel Servet (CP14/00229) researcher funded by the Spanish Institute of Health Carlos III (ISCIII). This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. H2020-MSCA-ITN-2015-675752 (SINGEK) and the Ministerio de Ciencia, Innovación y Universidades (SAF2017-89109-P; AEI/FEDER, UE). This project has been made possible in part by grant number 2018-182827 from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation. We thank ThePaperMill for critical reading and scientific editing services. Core funding is from the ISCIII and the Generalitat de Catalunya.

Author Contributions

The authors contributed to the sections of this Tutorial as follows: AL Data processing and Data analysis; CM Sample preparation; SP Optimization (Box 1); HH design, Sample preparation, Single-cell RNA sequencing, Further technical considerations and Future directions. All authors read and approved the final manuscript.

Competing financial and non-financial interests

The authors declare that they have no competing financial and non-financial interests.

Box 1. Optimizing reverse transcription for single-cell transcriptome sequencing.

Enzymes. Reverse transcription (RT) is one of the most crucial steps in the library preparation workflow. Despite its importance, however, relatively little has been done to improve the efficiency of the underlying enzymes. Reverse transcriptases are based on Moloney murine leukemia virus (MMLV)-derived enzymes, which originally had low processivity and high error rates due to their retroviral origins. Different point mutations have been introduced to improve processivity, resulting in enzymes that can reverse transcribe even very long RNAs (up to 12-14 kb). SuperScript II is a commonly used enzyme that became popular in the single-cell field due to its template-switching properties, exploited by methods such as Smart-seq2[129] and STRT-seq[49,59]. Most importantly, SuperScript II carries point mutations that inactivate its RNAse H domain, impairing competitive RNA degradation during cDNA synthesis. Alternative RT enzymes have been reported, with similar or superior performances, such as Maxima H (used in SCRB-seq[50,51]) or SMARTscribe in the SMARter v4 kit (Takara Bio). Protocols that do not require template switching and generate second strands by other means, such as polyA-tailing or random priming[130,131], can use SuperScript III, which carries different point mutations in the RNA polymerase, and displays increased thermal stability.

Additives. In an attempt to overcome the limitations of MMLV-based RT enzymes, several additives have been tested over the years. The challenge of generating full-length cDNA libraries has been a constant issue in molecular biology, pre-dating the advent of single-cell RNA-sequencing. Carninci and collaborators showed that the sugar trehalose had a thermo-stabilizing and thermo-protective effect on RT enzymes[132]. Conducting the RT reaction at a higher temperature enhances the unfolding of secondary RNA structures that could hinder enzyme processivity. This finding was later confirmed, and extended to the addition of betaine, alone or in combination with trehalose, to improve thermo-protection and related cDNA yield[133,134]. Smart-seq2[129] and STRT-seq-2i[59] use betaine in combination with magnesium chloride, where the latter, at concentrations higher than 1 mM, has been suggested to play a synergic destabilizing effect in the presence of betaine[135]. However, the extra magnesium chloride could also reduce the chelating function of 1,4-dithiothreitol (DTT), which is commonly used in RT reactions to guarantee higher cDNA yields and longer transcripts. In the very first published single-cell sequencing method[136], Tang and collaborators used the T4 gene 32 protein (T4g32p), a single-stranded binding protein that increases the yield and processivity during RT..

Template-switching oligonucleotides. The template-switching reaction relies on 2–5 untemplated cytosine nucleotides, which are added to newly synthesized cDNA (but not to fragmented or un-capped RNAs) when the enzyme reaches the 5′-end of the RNA. The presence of a template-switching oligonucleotide (TSO), carrying three complementary guanosines at its 3´-end, enables the enzyme to switch template and to add the complementary sequence of the TSO to the cDNA (including PCR adapter for subsequent amplification)). It has been suggested that the reduced RNA capture efficiency of single-cell RNAseq protocols might be due to the unstable binding of TSO to the untemplated nucleotides. The Smart-seq2 protocol addressed this issue by modifying the last nucleotide of the TSO with a locked nucleic acid. Furthermore, the importance of each nucleotide in the TSO has been extensively evaluated to define its optimal composition[137].

References

1. Regev, A. *et al.* Science Forum: The Human Cell Atlas. *eLife* **6,** e27041 (2017).
2. Ibarra-Soria, X. *et al.* Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat. Cell Biol.* **20,** 127–134 (2018).
3. Grün, D. *et al.* De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell* **19,** 266–277 (2016).
4. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The Technology and Biology of Single-Cell RNA Sequencing. *Mol. Cell* **58,** 610–620 (2015).
5. Bendall, S. C. *et al.* Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. *Cell* **157,** 714–725 (2014).
6. Kelsey, G., Stegle, O. & Reik, W. Single-cell epigenomics: Recording the past and predicting the future. *Science* **358,** 69–75 (2017).
7. Iacono, G. *et al.* bigSCale: An Analytical Framework for Big-Scale Single-Cell Data. *bioRxiv* 197244 (2017). doi:10.1101/197244
8. Consortium, T. T. M., Quake, S. R., Wyss-Coray, T. & Darmanis, S. Transcriptomic characterization of 20 organs and tissues from mouse at single cell resolution creates a Tabula Muris. *bioRxiv* 237446 (2017). doi:10.1101/237446
9. Halpern, K. B. *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542,** 352–356 (2017).
10. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551,** 333–339 (2017).
11. Karaiskos, N. *et al.* The Drosophila embryo at single-cell transcriptome resolution. *Science* **358,** 194–199 (2017).
12. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357,** 661–667 (2017).
13. Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M. & Reddien, P. W. Cell type transcriptome atlas for the planarian Schmidtea mediterranea. *Science* **360,** (2018).
14. Davie, K. *et al.* A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **174,** 982-998.e20 (2018).
15. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **173,** 1307 (2018).
16. Shahbazi, M. N. *et al.* Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature* **552,** 239–243 (2017).
17. Enge, M. *et al.* Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell* **171,** 321-330.e14 (2017).
18. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47,** 320–329 (2015).
19. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352,** 189–196 (2016).
20. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539,** 309–313 (2016).
21. Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171,** 1611-1624.e24 (2017).
22. Guillaumet-Adkins, A. *et al.* Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol.* **18,** 45 (2017).
23. Alles, J. *et al.* Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* **15,** 44 (2017).
24. Wang, W., Penland, L., Gokce, O., Croote, D. & Quake, S. R. High fidelity hypothermic preservation of primary tissues in organ transplant preservative for single cell transcriptome analysis. *BMC Genomics* **19,** 140 (2018).
25. Lacar, B. *et al.* Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat. Commun.* **7,** 11022 (2016).
26. Krishnaswami, S. R. *et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* **11,** 499–524 (2016).
27. Habib, N. *et al.* Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353,** 925–928 (2016).

28. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14,** 955–958 (2017).
29. Bakken, T. E. *et al.* Equivalent high-resolution identification of neuronal cell types with single-nucleus and single-cell RNA-sequencing. *bioRxiv* 239749 (2017). doi:10.1101/239749
30. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356,** (2017).
31. van den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14,** 935–936 (2017).
32. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Rep.* **2,** 666–673 (2012).
33. Paul, F. *et al.* Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163,** 1663–1677 (2015).
34. Prakadan, S. M., Shalek, A. K. & Weitz, D. A. Scaling by shrinking: empowering single-cell ‘omics’ with microfluidic devices. *Nat. Rev. Genet.* **18,** 345–361 (2017).
35. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14,** 865–868 (2017).
36. Barriga, F. M. *et al.* Mex3a Marks a Slowly Dividing Subpopulation of Lgr5+ Intestinal Stem Cells. *Cell Stem Cell* **20,** 801-816.e7 (2017).
37. Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **65,** 631-643.e4 (2017).
38. Svensson, V. *et al.* Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14,** 381–387 (2017).
39. Avital, G. *et al.* scDual-Seq: mapping the gene regulatory program of Salmonella infection by host and pathogen single-cell RNA-sequencing. *Genome Biol.* **18,** 200 (2017).
40. Hayashi, T. *et al.* Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.* **9,** 619 (2018).
41. Faridani, O. R. *et al.* Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.* **34,** 1264–1266 (2016).
42. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11,** 163–166 (2014).
43. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539,** 309–313 (2016).
44. Giustacchini, A. *et al.* Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat. Med.* **23,** 692–702 (2017).
45. Stubbington, M. J. T. *et al.* T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13,** 329–332 (2016).
46. Karlsson, K. & Linnarsson, S. Single-cell mRNA isoform diversity in the mouse brain. *BMC Genomics* **18,** 126 (2017).
47. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9,** 171–181 (2014).
48. Ramsköld, D. *et al.* Full-Length mRNA-Seq from single cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30,** 777–782 (2012).
49. Islam, S. *et al.* Highly multiplexed and strand-specific single-cell RNA 5’ end sequencing. *Nat. Protoc.* **7,** 813–828 (2012).
50. Soumillon, M., Cacchiarelli, D., Semrau, S., Oudenaarden, A. van & Mikkelsen, T. S. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv* 003236 (2014). doi:10.1101/003236
51. Bagnoli, J. W. *et al.* mcSCRB-seq: sensitive and powerful single-cell RNA sequencing. *bioRxiv* 188367 (2017). doi:10.1101/188367
52. Sasagawa, Y. *et al.* Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* **14,** 3097 (2013).
53. Sasagawa, Y. *et al.* Quartz-Seq2: a high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads. *Genome Biol.* **19,** (2018).
54. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343,** 776–779 (2014).
55. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17,** 77 (2016).
56. Wu, A. R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11,** 41–46 (2014).

57. Streets, A. M. *et al.* Microfluidic single-cell whole-transcriptome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 7048–7053 (2014).
58. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523,** 486–490 (2015).
59. Hochgerner, H. *et al.* STRT-seq-2i: dual-index 5' single cell and nucleus RNA-seq on an addressable microwell array. *Sci. Rep.* **7,** 16327 (2017).
60. Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14,** 395–398 (2017).
61. Klein, A. M. *et al.* Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **161,** 1187–1201 (2015).
62. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161,** 1202–1214 (2015).
63. Zilionis, R. *et al.* Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12,** 44–73 (2017).
64. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8,** 14049 (2017).
65. Rosenberg, A. B. *et al.* Scaling single cell transcriptomics through split pool barcoding. *bioRxiv* 105163 (2017). doi:10.1101/105163
66. Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348,** 910–914 (2015).
67. Ramani, V. *et al.* Massively multiplex single-cell Hi-C. *Nat. Methods* **14,** 263–266 (2017).
68. Mulqueen, R. M. *et al.* Scalable and efficient single-cell DNA methylation sequencing by combinatorial indexing. *bioRxiv* 157230 (2017). doi:10.1101/157230
69. Kang, H. M. *et al.* Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36,** 89–94 (2018).
70. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24,** 2033–2040 (2014).
71. Mora-Castilla, S. *et al.* Miniaturization Technologies for Efficient Single-Cell Library Preparation for Next-Generation Sequencing. *Slas Technol.* **21,** 557–567 (2016).
72. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinforma. Oxf. Engl.* **25,** 1105–1111 (2009).
73. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29,** 15–21 (2013).
74. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34,** 525–527 (2016).
75. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
76. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–515 (2010).
77. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **31,** 166–169 (2015).
78. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27,** 491–499 (2017).
79. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17,** 29 (2016).
80. McCarthy, D. J., Campbell, K. R., Lun, A. T. L. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33,** 1179–1186 (2017).
81. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. zUMIs: A fast and flexible pipeline to process RNA sequencing data with UMIs. *bioRxiv* 153940 (2018). doi:10.1101/153940
82. Tian, L. *et al.* scPipe: a flexible R/Bioconductor preprocessing pipeline for single-cell RNA-sequencing data. *bioRxiv* 175927 (2018). doi:10.1101/175927
83. Azizi, E. *et al.* Single-cell Map of Diverse Immune Phenotypes Driven by the Tumor Microenvironment. *bioRxiv* 221994 (2018). doi:10.1101/221994
84. Zappia, L., Phipson, B. & Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *bioRxiv* 206573 (2018). doi:10.1101/206573
85. Vallejos, C. A., Marioni, J. C. & Richardson, S. BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. *PLOS Comput. Biol.* **11,** e1004333 (2015).
86. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).

87. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11,** R25 (2010).
88. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11,** 94 (2010).
89. Bacher, R. *et al.* SCnorm: robust normalization of single-cell RNA-seq data. *Nat. Methods* **14,** 584–586 (2017).
90. L. Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17,** 75 (2016).
91. Buttner, M., Miao, Z., Wolf, A., Teichmann, S. A. & Theis, F. J. Assessment of batch-correction methods for scRNA-seq data with a new test metric. *bioRxiv* 200345 (2017). doi:10.1101/200345
92. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostat. Oxf. Engl.* **8,** 118–127 (2007).
93. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).
94. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36,** 421–427 (2018).
95. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33,** 495–502 (2015).
96. van Dijk, D. *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174,** 716-729.e27 (2018).
97. Li, W. V. & Li, J. J. scImpute: Accurate And Robust Imputation For Single Cell RNA-Seq Data. *bioRxiv* 141598 (2017). doi:10.1101/141598
98. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10,** 1093 (2013).
99. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40,** 4288–4297 (2012).
100. Abdi, H. & Williams, L. J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2,** 433–459 (2010).
101. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9,** 2579–2605 (2008).
102. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinforma. Oxf. Engl.* **31,** 2989–2998 (2015).
103. McInnes, L. & Healy, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv180203426 Cs Stat* (2018).
104. Becht, E. *et al.* Evaluation of UMAP as an alternative to t-SNE for single-cell data. *bioRxiv* 298430 (2018). doi:10.1101/298430
105. Qiu, X. *et al.* Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14,** 309–315 (2017).
106. Fan, J. *et al.* Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* **13,** 241–244 (2016).
107. Guo, M., Wang, H., Potter, S. S., Whitsett, J. A. & Xu, Y. SINCERA: A Pipeline for Single-Cell RNA-Seq Profiling Analysis. *PLOS Comput. Biol.* **11,** e1004575 (2015).
108. Kiselev, V. Y. *et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14,** 483 (2017).
109. Xu, C. & Su, Z. Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinforma. Oxf. Engl.* **31,** 1974–1980 (2015).
110. Kharchenko, P. V., Silberstein, L. & Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11,** 740–742 (2014).
111. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16,** (2015).
112. Korthauer, K. D. *et al.* A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biol.* **17,** 222 (2016).
113. Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15,** 255–261 (2018).
114. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14,** 979–982 (2017).
115. Ji, Z. & Ji, H. TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* **44,** e117 (2016).

116. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods: towards more accurate and robust tools. *bioRxiv* 276907 (2018). doi:10.1101/276907
117. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14,** 1083–1086 (2017).
118. Kiselev, V. Y., Yiu, A. & Hemberg, M. scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* **15,** 359–362 (2018).
119. Ji, N. & van Oudenaarden, A. Single molecule fluorescent in situ hybridization (smFISH) of C. elegans worms and embryos. *WormBook Online Rev. C Elegans Biol.* 1–16 (2012). doi:10.1895/wormbook.1.153.1
120. Moffitt, J. R. *et al.* High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 11046–11051 (2016).
121. Ibarra-Soria, X. *et al.* Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat. Cell Biol.* **20,** 127–134 (2018).
122. Ke, R., Mignardi, M., Hauling, T. & Nilsson, M. Fourth Generation of Next-Generation Sequencing Technologies: Promise and Consequences. *Hum. Mutat.* **37,** 1363–1367 (2016).
123. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10,** 857–860 (2013).
124. Lein, E., Borm, L. E. & Linnarsson, S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358,** 64–69 (2017).
125. Ibarra-Soria, X. *et al.* Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat. Cell Biol.* (2018). doi:10.1038/s41556-017-0013-z
126. Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12,** 519–522 (2015).
127. Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13,** 229–232 (2016).
128. Clark, S. J. *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9,** 781 (2018).
129. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10,** 1096–1098 (2013).
130. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17,** 77 (2016).
131. Sasagawa, Y. *et al.* Quartz-Seq2: a high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads. *Genome Biol.* **19,** (2018).
132. Carninci, P. *et al.* Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 520–524 (1998).
133. Spiess, A.-N. & Ivell, R. A highly efficient method for long-chain cDNA synthesis using trehalose and betaine. *Anal. Biochem.* **301,** 168–174 (2002).
134. Pinto, F. L. & Lindblad, P. A guide for in-house design of template-switch-based 5’ rapid amplification of cDNA ends systems. *Anal. Biochem.* **397,** 227–232 (2010).
135. Lambert, D. & Draper, D. E. Effects of osmolytes on RNA secondary and tertiary structure stabilities and RNA-Mg2+ interactions. *J. Mol. Biol.* **370,** 993–1005 (2007).
136. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6,** 377–382 (2009).
137. Zajac, P., Islam, S., Hochgerner, H., Lönnerberg, P. & Linnarsson, S. Base Preferences in Non-Templated Nucleotide Incorporation by MMLV-Derived Reverse Transcriptases. *PLOS ONE* **8,** e85270 (2013).
138. Liu, W. *et al.* Sample preparation method for isolation of single-cell types from mouse liver for proteomic studies. *Proteomics* **11,** 3556–3564 (2011).
139. Dorrell, C. *et al.* Surface markers for the murine oval cell response. *Hepatol. Baltim. Md* **48,** 1282–1291 (2008).
140. Su, X. *et al.* Single-cell RNA-Seq analysis reveals dynamic trajectories during mouse liver development. *BMC Genomics* **18,** 946 (2017).
141. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509,** 371–375 (2014).

142. Chapman, H. A. *et al.* Integrin α6β4 identifies an adult distal lung epithelial population with regenerative potential in mice. *J. Clin. Invest.* **121,** 2855–2862 (2011).
143. Xu, Y. *et al.* Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* **1,** e90558 (2016).
144. Joost, S. *et al.* Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity. *Cell Syst.* **3,** 221-237.e9 (2016).
145. Der, E. *et al.* Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis. *JCI Insight* **2,** (2017).
146. Autengruber, A., Gereke, M., Hansen, G., Hennig, C. & Bruder, D. Impact of enzymatic tissue disintegration on the level of surface molecule expression and immune cell function. *Eur. J. Microbiol. Immunol.* **2,** 112–120 (2012).
147. Barriga, F. M. *et al.* Mex3a Marks a Slowly Dividing Subpopulation of Lgr5+ Intestinal Stem Cells. *Cell Stem Cell* (2017). doi:10.1016/j.stem.2017.02.007
148. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525,** 251–255 (2015).
149. Glass, L. L. *et al.* Single-cell RNA-sequencing reveals a distinct population of proglucagon-expressing cells specific to the mouse upper small intestine. *Mol. Metab.* **6,** 1296–1303 (2017).
150. Herring, C. A. *et al.* Unsupervised Trajectory Analysis of Single-Cell RNA-Seq and Imaging Data Reveals Alternative Tuft Cell Origins in the Gut. *Cell Syst.* **6,** 37-51.e9 (2018).
151. Merlos-Suárez, A. *et al.* The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell* **8,** 511–524 (2011).
152. Wollny, D. *et al.* Single-Cell Analysis Uncovers Clonal Acinar Cell Heterogeneity in the Adult Pancreas. *Dev. Cell* **39,** 289–301 (2016).
153. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* **3,** 346-360.e4 (2016).
154. Segerstolpe, Å. *et al.* Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab.* **24,** 593–607 (2016).
155. Petersen, M. B. K. *et al.* Single-Cell Gene Expression Analysis of a Human ESC Model of Pancreatic Endocrine Development Reveals Different Paths to β-Cell Differentiation. *Stem Cell Rep.* **9,** 1246–1261 (2017).
156. Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst.* **3,** 385-394.e3 (2016).
157. Li, D. *et al.* Complete disassociation of adult pancreas into viable single cells through cold trypsin-EDTA digestion. *J. Zhejiang Univ. Sci. B* **14,** 596–603 (2013).
158. Shekhar, K. *et al.* Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* **166,** 1308-1323.e30 (2016).
159. Daniszewski, M. *et al.* Single cell RNA sequencing of stem cell-derived retinal ganglion cells. *Sci. Data* **5,** 180013 (2018).

Table 1: Glossary

TERM	DEFINITION
Algorithm	A process or set of rules to be followed in computational calculations or other problem-solving operations.
Barcode	A stretch of sequence used to uniquely label DNA/RNA molecules, cells or sequencing libraries (to allow multiplexing).
Batch effect	A technical sources of variation added during sample handling.
Benchmark	Systematically comparison of different techniques (experimental or computational) for their performance in a given scenario.
Binary classifier	A classification function that predicts the assignment of an elements to a set of groups.
Bulk RNA sequencing	The sequencing of RNA isolated from pools of cells.
Cell barcode	A cell specific unique sequence tag that is added to RNA transcripts during library preparation.
Cell capture	Positioning of single cells into reaction volumes (e.g. droplets or wells) for downstream processing.
Cluster annotation	Assigning a functions or identity to a group of cells based on the expression of marker genes.
Clustering	Clustering is the task of grouping cells in a way that cells in the same group (cluster) are more similar to each other than to cells of another group.
Combinatorial barcoding	The use of combinations of cell barcodes with repeated assignment of barcodes to cells during multiple indexing rounds.
Deconvolution	A process of resolving a complex mixtures (e.g. tissue) into its constituent elements (e.g. underlying cell type composition).
De-multiplexing	The process of separating the elements of interest in a mixed or multiplexed sample.
Digital counting	The counting of RNA molecules using unique molecular identifier (UMI) sequences.
Doublets	Two cells that are processed together in a reaction volume (e.g. well or droplet) and receive the same single-cell barcode.
Dropout events	Non-detected transcripts in the final dataset although the gene is expressed in the cell, leading to a false zero values in the expression matrix.
Fastq reads	A sequence of the four nucleotides ACGT obtained after sequencing in a specific format that represents the chain of nucleotides.
Gene expression matrix	A data matrix containing information about the level of gene expression per cell.
Imputation	The process of replacing missing data with inferred values.
Index-sorting	The isolation of single cells by FACS and the retrospective assignment of fluorescence signals during scRNAseq data analysis.
Library	DNA molecules that contain specifc sequences (primers) that enable the initiation high-throughput sequencing reactions.
Locked nuclic acids (LNA)	Modified RNA nucleotides with a bridge connecting the 2' oxygen and 4' carbon to increase the hybridization properties of oligonucleotides.
Microtiter plates	Also microplates or microwell plates, a flat plate with multiple wells used as individual reaction sites.
Pipeline	An analysis procedure where inputs go through a number of processing steps chained together to produce an output.
Poisson distribution	A discrete probability distribution that expresses the probability for the number of events in specified intervals such as distance, area or volume.
Pooling	Combining molecules or cells for their joint processing.
Promoter	A DNA sequence that initiates transcription of the downstream sequence.

Pseudotime	An inferred timeline of the progress cells make through a dynamic process such as cell differentiation.
Spike-in RNA	A pool of RNA transcripts of known sequence composition and quantity used to calibrate experiments.
Tagmentation	Reaction that involves the transposase-based cleaving of DNA and the tagging of the double-stranded DNA with universal overhangs.
Template swithching oligo (TSO)	DNA oligo sequence that carries three riboguanosines (rGrGrG) at its 3' end and binds to the cytosine extension of the cDNA molecules after reverse transcription.
Trajectory inference	Computational reconstruction of an underlying cellular developmental/differentiation path.
Unique molecular identifiers (UMI)	Random sequences attached to transcripts used as molecular tags to detect and quantify unique RNA molecules.
Zero-inflated data	Data with an excess of zero counts. To model zero-inflated data Poisson distribution is used.

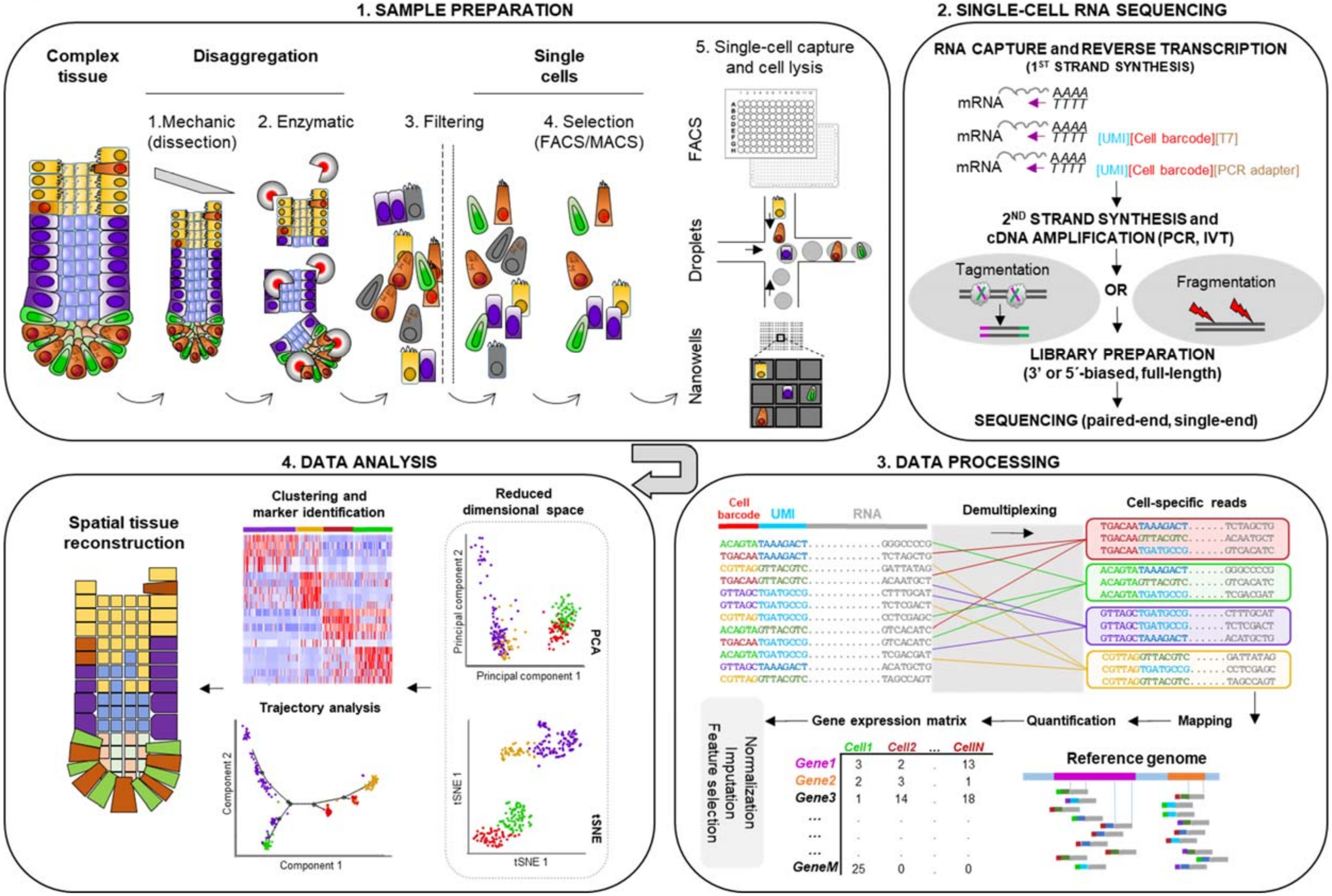
Table 2: Tissue-specific enzymatic treatments to prepare single-cell suspensions (from human and mouse samples).

Tissue	Digestion enzyme	Time (min)	Temperature (C°)	Final concentration	Ref.
Liver	Collagenase IV	10	37	0.16mg/ml	138
	Liberase Blendzyme 3	5-8	37	40ug/ml	9
	Collagenase, Collagenase D and Pronase, Trypsin	20, 20,10	37	2.5mg/ml, 10mg/ml and 10mg/ml, 0.05%	139
	Collagenase IV	30	37	0.05%	140
Lung	Dispase and Elastase	45	37	0.33U/ml and 3U/ml	141
	Collagenase and Dispase	45	37	0,2% solution	142
	Dispase, Elastase and Trypsin	60, 30, 15	4C and 37C	2mg/ml, 5U/ml plus 0.125%,	143
Skin	Trypsin	120	32	1X	144
	Liberase TL	15	37	2mg/ml	145
Spleen	Collagenase D	45	37	2mg/ml	146
GI tract	Dispase	20	37	0.4mg/ml	147
	Trypsin	30	37	2mg/ml	148
	TrypLE Express	1	37	1X	10
	Collagenase	40	37	1mg/ml	149
	Collagenase I	60	37	2.5 mg/ml	150
	Collagenase IV	30	37	2mg/ml	151
Pancreas	Collagenase type CLS IV	30	37	1mg/ml	152
	Collagenase P	30	37	0.8 mM	153
	TrypLE Express	1	37	1X	154
	Accutase and TrypLE Express	10 and 5-20	37	1X	155
	Accutase	8-10	37	1X	156
	Trypsin	30	37	1X	157
Kidney	Liberase TL	15	37	2mg/ml	145
Retina	Papain	45	37	4U/ml	62, 158
	Accutase	5	37	1X	159

Table 3: Key features of microtiter-plate- and microfluidics-based single-cell RNA sequencing methods.

Method	Capture format	Cell loading	Single-cell indexing	Molecule identifier	Additives in RT	cDNA amplification	Fragmentation	Transcript coverage	Sequencing	Ref.
Smart-seq	Plate	FACS	Tagmentation	N/A	N/A	PCR	Tagmentation	Full-length	Paired-end	48
Smart-seq2	Plate	FACS	Tagmentation	N/A	Betaine	PCR	Tagmentation	Full-length	Paired-end	129
STRT-seq	Plate	FACS	TSO	UMI	N/A	PCR	DNaseI	5`-end	Single-end	49
STRT-seq-2i	Nanowell	FACS/Poisson	TSO	UMI	Betaine	PCR	Tagmentation	5`-end	Single-end	59
SCRB-seq	Plate	FACS	OligoT primer	UMI	N/A	PCR	Tagmentation	3`-end	Paired-end	50
mcSCRB-seq	Plate	FACS	OligoT primer	UMI	PEG	PCR	Tagmentation	3´-end	Paired-end	51
Quartz-seq	Plate	FACS	OligoT primer	N/A	N/A	PCR	Ultrasound	Full-length	Paired-end	52
Quartz-seq2	Plate	FACS	OligoT primer	UMI	N/A	PCR	Ultrasound	3´-end	Paired-end	53
CEL-seq	Plate	FACS	OligoT primer	N/A	N/A	IVT	KOAc, MgOAc	3´-end	Paired-end	32
CEL-seq2	Plate	FACS	OligoT primer	UMI	N/A	IVT	Random priming	3´-end	Paired-end	55
MARS-seq	Plate	FACS	OligoT primer	UMI	N/A	IVT	Zinc	3´-end	Paired-end	54
Seq-Well	Nanowell	Poisson	OligoT beads	UMI	Ficoll	PCR	Tagmentation	3´-end	Paired-end	60
inDrops	Droplets	Poisson	OligoT beads	UMI	IGEPAL	IVT	KOAc, MgOAc	3´-end	Paired-end	61
Drop-seq	Droplets	Double Poisson	OligoT beads	UMI	Ficoll	PCR	Tagmentation	3´-end	Paired-end	62

Figure 1


1. SAMPLE PREPARATION
Complex tissue
Disaggregation
Single cells
5. Single-cell capture and cell lysis
1.Mechanic (dissection)
2. Enzymatic
3. Filtering
4. Selection (FACS/MACS)
FACS
Droplets
Nanowells
2. SINGLE-CELL RNA SEQUENCING
RNA CAPTURE and REVERSE TRANSCRIPTION
(1ST STRAND SYNTHESIS)
mRNA
AAAA
TTTT
[UMI][Cell barcode][T7]
[UMI][Cell barcode][PCR adapter]
2ND STRAND SYNTHESIS and cDNA AMPLIFICATION (PCR, IVT)
Tagmentation
OR
Fragmentation
LIBRARY PREPARATION
(3' or 5´-biased, full-length)
SEQUENCING (paired-end, single-end)
3. DATA PROCESSING
Cell barcode
UMI
RNA
Demultiplexing
Cell-specific reads
ACAGTATAAAGACT............GGGCCCCG
TGACAATAAAGACT............TCTAGCTG
CGTTAGGTTACGTC............GATTATAG
TGACAAGTTACGTC............ACAATGCT
GTTAGCTGATGCCG............CTTTGCAT
GTTAGCTGATGCCG............TCTCGACT
CGTTAGTGATGCCG............CCTCGAGC
ACAGTAGTTACGTC............GTCACATC
TGACAATGATGCCG............GTCACATC
ACAGTATGATGCCG............TCGACGAT
GTTAGCTAAAGACT............ACATGCTG
CGTTAGGTTACGTC............TAGCCAGT
TGACAATAAAGACT......TCTAGCTG
TGACAAGTTACGTC......ACAATGCT
TGACAATGATGCCG......GTCACATC
ACAGTATAAAGACT......GGGCCCCG
ACAGTAGTTACGTC......GTCACATC
ACAGTATGATGCCG......TCGACGAT
GTTAGCTGATGCCG......CTTTGCAT
GTTAGCTGATGCCG......TCTCGACT
GTTAGCTAAAGACT......ACATGCTG
CGTTAGGTTACGTC .....GATTATAG
CGTTAGTGATGCCG......CCTCGAGC
CGTTAGGTTACGTC......TAGCCAGT
Gene expression matrix
Quantification
Mapping
Normalization
Imputation
Feature selection
Reference genome
4. DATA ANALYSIS
Spatial tissue reconstruction
Clustering and marker identification
Trajectory analysis
Component 2
Component 1
Reduced dimensional space
Principal component 2
Principal component 1
PCA
tSNE 1
tSNE 1
tSNE

	Cell1	Cell2	...	CellN
Gene1	3	2	.	13
Gene2	2	3	.	1
Gene3	1	14	.	18
...	.	.	.	.
...	.	.	.	.
...	.	.	.	.
GeneM	25	0	.	0

Figure 1. The single-cell RNA sequencing process. The successful design of single-cell transcriptomics experiments includes four major phases: 1) During sample preparation cells are physically separated into a single-cell solution from which specific cell types can be enriched or excluded (optional). Following their capture in wells or droplets single cells are lysed and the RNA is released for subsequent processing. 2) To convert RNA into sequencing ready libraries, polyA-tailed RNA molecules are captured on polyT oligonucleotides that can contain unique molecule identifier (UMI) sequences and single-cell specific barcodes (5` and 3'-biased methods). To enable the subsequent amplification of the RNA by PCR or IVT, adapters or T7 polymerase promoter sequences are included in the oligonucleotides, respectively. Following reverse transcription into cDNA and second strand synthesis (optional), the transcriptome is amplified (PCR or IVT). In order to be converted into sequencing libraries, the amplicons are fragmented by enzymatic (e.g. tagmentation) or mechanic (e.g. ultrasound) forces. Sequencing adapters are attached during a final amplification step. Full-length sequencing can be carried out, or 5' or 3' transcript ends can be selected for sequencing using specific amplification primers (optional). For most applications, paired-end sequencing is required. 3) The sequencing reads are de-multiplexed based on cell-specific barcodes and mapped to the respective reference genome. UMI sequences are used for the digital counting of RNA molecules and for correction of amplification biases. The resulting gene expression quantification matrix can subsequently be normalized, missing values imputed, before extracting informative genes for the analysis. 4) Dimensional reduction representations guide the estimation of sample heterogeneity and the data interpretation. Data analysis can then be tailored to the underlying dataset, enabling cells to be clustered into potential cell types and states, or ordered along a predicted trajectory in pseudotime. Eventually, the spatial cellular organization can be reconstructed through the interrogation of marker genes (experimentally) or through marker-guided computational reconstruction (inference).