

Video Article

Droplet Barcoding-Based Single Cell Transcriptomics of Adult Mammalian Tissues

Jo Anne Stratton^{1,2,3}, Sarthak Sinha², Wisoo Shin², Elodie Labit², Tak-Ho Chu^{1,4}, Prajay T. Shah², Rajiv Midha^{1,4}, Jeff Biernaskie^{1,2,3}¹Hotchkiss Brain Institute, University of Calgary²Department of Comparative Biology and Experimental Medicine, Faculty of Veterinary Medicine, University of Calgary³Alberta Children's Hospital Research Institute, University of Calgary⁴Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary

*These authors contributed equally

Correspondence to: Jeff Biernaskie at jabierna@ucalgary.caURL: <https://www.jove.com/video/58709>DOI: [doi:10.3791/58709](https://doi.org/10.3791/58709)

Keywords: Biology, Issue 143, Single cell, RNA-Seq, Adult, Transcriptome, RNA, FACS

Date Published: 1/10/2019

Citation: Stratton, J.A., Sinha, S., Shin, W., Labit, E., Chu, T.H., Shah, P.T., Midha, R., Biernaskie, J. Droplet Barcoding-Based Single Cell Transcriptomics of Adult Mammalian Tissues. *J. Vis. Exp.* (143), e58709, doi:10.3791/58709 (2019).

Abstract

The analysis of single cell gene expression across thousands of individual cells within a tissue or microenvironment is a valuable tool for identifying cell composition, discrimination of functional states, and molecular pathways underlying observed tissue functions and animal behaviors. However, the isolation of intact, healthy single cells from adult mammalian tissues for subsequent downstream single cell molecular analysis can be challenging. This protocol describes the general processes and quality control checks necessary to obtain high-quality adult single cell preparations from the nervous system or skin that enabled subsequent unbiased single cell RNA sequencing and analysis. Guidelines for downstream bioinformatic analysis are also provided.

Video Link

The video component of this article can be found at <https://www.jove.com/video/58709/>

Introduction

With the development of high throughput single cell technology^{1,2} and advancements in user-friendly bioinformatics tools over the last decade³, a new field of high-resolution gene expression analysis has emerged – single-cell RNA sequencing (scRNA-Seq). The study of single cell gene expression was first developed to identify heterogeneity within defined cell populations, such as in stem cells or cancer cells, or to identify rare populations of cells^{4,5}, which were unattainable using traditional bulk RNA sequencing techniques. Bioinformatic tools have enabled the identification of novel sub-populations (Seurat)², visualization of the order of cells along a pseudotime space (Monocle)⁶, definition of active signaling networks within or between populations (SCENIC)⁷, prediction of the assembly of single-cells in an artificial 3D space (Seurat, and more)⁸. With these new and exciting analyses available to the scientific community, scRNA-Seq is fast-becoming the new standard approach for gene expression analysis.

Despite the vast potential of scRNA-Seq, the technical skillsets required to produce a clean dataset and to accurately interpret results can be challenging to newcomers. Here, a basic, but comprehensive protocol, starting from the isolation of single cells from whole primary tissues to visualization and presentation of data for publication is presented (**Figure 1**). First, the isolation of healthy single cells can be deemed challenging, as different tissues vary in their degree of sensitivity to enzymatic digestion and subsequent mechanical dissociation. This protocol provides guidance in these isolation steps and identifies important quality control checkpoints throughout the process. Second, understanding the compatibility and requirements between single cell technology and next-generation sequencing can be confusing. This protocol provides guidelines to implement a user-friendly, droplet-based single-cell barcoding platform and perform sequencing. Finally, computer programming is an important prerequisite for analyzing single-cell transcriptomic datasets. This protocol provides resources for getting started with the R programming language and provides guidance on implementing two popular scRNA-Seq-specific R packages. Together, this protocol can guide newcomers in performing scRNA-Seq analysis for obtaining clear, interpretable results. This protocol can be adjusted to most tissues in the mouse, and importantly could be modified for use with other organisms, including human tissue. Adjustments depending on the tissue and user will be required.

There are several considerations to keep in mind while following this protocol; including, 1) Following all quality control guidelines in Steps 1 and 2 of this protocol is recommended to ensure a viable single cell suspension of all cells within the sample of interest while ensuring accurate total cell number counts (summarized in **Figure 2**). Once this is achieved, and if all the optimized conditions are followed, the quality control steps can be dropped (to save time - preserving RNA quality and reducing cell loss). Confirming successful isolation of high viability single cells from the tissue of interest is highly recommended before any downstream processing. 2) Since some cell types are more sensitive than others to stress, excessive dissociation techniques can inadvertently bias the population, therefore confounding downstream analysis. Gentle dissociation without

unnecessary cellular shearing and digestion is critical for achieving high cellular yields and an accurate representation of tissue composition. Shear forces occur during the trituration, FACS and resuspension steps. 3) As with any RNA work, is it best to introduce as little additional RNase into the sample as possible during preparation. This will help maintain high-quality RNA. Use ribonuclease inhibitor solutions with rinsing to clean tools and any equipment that is not RNase-free but avoid DEPC-treated products. 4) Perform preparations as quickly as possible. This will help maintain high-quality RNA and reduce cell death. Depending on the tissue dissection length and animal number, consider starting multiple dissections/preparations at the same time. 5) Prepare cells on ice when possible to maintain high quality RNA, reduce cell death, and slow cell signaling and transcriptional activity. Albeit, ice-cold processing is ideal for most cell types, some cell types (e.g., neutrophils) perform better when processed at room temperature. 6) Avoid calcium, magnesium, EDTA, and DEPC-treated products during cell preparation.

Protocol

All protocols described here are in accordance with and approved by the University of Calgary's Animal Care Committee.

1. Dissociating Tissue (Day 1)

1. Euthanize mice with an overdose of sodium pentobarbital (i.p., 50 mg/kg) or as appropriate according to animal ethics protocol. Then remove unwanted hair from the back and legs of the mouse, and ethanol-sterilize the region of dissection.
2. Dissect the tissue or microenvironment of interest. For this protocol, we use skin and nerve tissues to demonstrate the generalizability of droplet barcoding-based single cell transcriptomics following adult tissue dissociation.
 1. For the sciatic nerve, use the detailed protocol found at Stratton *et al.*⁹. Briefly, cut the skin away from the hind region of the back/legs of mouse. Make an incision along the length of the thigh with a sterile scalpel blade. Use fine forceps and scissors to expose and remove the sciatic nerve.
 2. For the back skin, use the detailed protocol found at Biernaskie *et al.*¹⁰. Briefly, dissect dorsal back skin by making incisions from shoulder to shoulder, across the rump and down the back using fine forceps and scissors. Cut the skin into thin slices (0.5 cm thickness) using a sterile scalpel blade.
3. Wash tissue 2 times with ice-cold HBSS, and remove unwanted connective tissue, fat deposits or debris under a dissecting microscope.
 1. For the skin dermis only, float the slices in dispase (5 mg/mL, 5 U/mL) in HBSS for 30 - 40 min at 37 °C. Surgically separate the epidermis from the dermis. Discard the epidermis or further dissociate using trypsin if of interest.
4. Mince sample into 1-2 mm pieces using a sterile scalpel blades and put into freshly-thawed 2 mg/mL cold collagenase-IV enzyme (2 mg/mL, 125 CDU/mg, in F12 media).
 1. For the nerve, use ~500 µL per 2x sciatic nerves. For the skin, use ~8 mL per 1x mouse back skin.
NOTE: The tissues should be fully submerged in the collagenase-IV solution. It is critical that any digestion enzymes are handled, stored, and prepared appropriately. If enzymes are left at room temperature for long periods of time, single-cell isolation will require excessive mechanical trituration and reduce cell viability. Collagenase-IV can also be made up in cell culture media where cell viability is most optimal. However, this could alter enzyme activity or transcriptional signature so should be optimized by the user.
5. Incubate the sample in the enzyme in a 37 °C bath for 30 min with gentle shaking every 10 min. A shaker placed at 37 °C is also an appropriate alternative.
6. Triturate with a P1000 pipettor 20-30 times at 30 min post-enzyme addition.
7. Repeat trituration every 30 min until solution appears cloudy and chunks of tissue are largely dissociated.
NOTE: Ensure full release of cells (**Figure 2b, 2c**). To confirm full release, plate cells with Nuc Blue (2 drops per 1 mL) and after 20 min, check under the microscope to ensure all nuclei are associated with single cells rather than debris. It is critical to check the degree of cell release within a given experiment for each tissue type or condition. In fibrotic tissue (*i.e.*, chronic injury) or uninjured adult tissue, the release of cells will vary dramatically from acute injury or embryonic tissue. This is especially important because certain cell types are less likely to release from the tissue than others, thus preferentially excluding those cells from downstream analysis.
 1. For the nerve, dissociate tissue for 0.5-1.5 hours total. For the skin, dissociate tissue for 2 hours total (in the last hour of incubation, add DNase (1 mg/mL) to the skin sample).
8. Filter twice with a 40 µm filter. Rinse the filter with ice-cold 1% BSA/HBSS.
9. Centrifuge at 260 x g for 8 min. Then remove supernatant.
 1. Resuspend the cell pellet in HBSS containing 1% BSA using a wide-bore tip, and place on ice. The resuspension volume is based on tissue volume (800 mg wet weight for skin = 800 µL volume; 10 mg wet weight for nerves = 100 µL volume).
 2. Optionally, start with a low resuspension volume and then adjust as necessary based on flow rate (events per second) on FACS sorter. The most efficient sort density (maximize number of cells collected while minimalizing time) for collections is 3,000-7,000 events/second.
10. If using viability dye, take out a subaliquot for an unstained control. Then add 1:15,000 viability dye (stock: 20,000 nM/µL) to sample (1.3 nM/µL final concentration) using a wide-bore tip to reduce shearing.
NOTE: It is critical to check for the degree of cell death within a given experiment for each tissue type or condition. Some cell types within a sample are more likely to die than others, thus being preferentially excluded from downstream analysis.
 1. Incubate sample with viability dye for 5-10 min on ice in the dark. Then add 4 mL of ice-cold 1% BSA/HBSS to sample. Centrifuge at 260 x g for 8 min to remove excess viability dye. Treat the subaliquot with no viability dye unstained control in the same manner.

2. Isolating Viable and Healthy Cells (Day 1)

1. Ensure that the FACS facility follows appropriate fluorescence activated cell sorting (FACS) parameters.

1. Prepare the FACS machine in advance to ensure that it is ready once the final centrifuge in Step 1 is complete, and ensure that the collection compartment is kept cold using ice blocks.
2. Use the following parameters: Flow rate: 1.0 (corresponds roughly to 10 $\mu\text{L}/\text{min}$); Filter: 1.5 ND; Nozzle size: 100 μm ; Forward scatter: 80 - 180 V (change as necessary in order to distinguish size of events); Side scatter: 150 - 220 V (change as necessary in order to distinguish granularity/shape of events); Laser: 100 - 400 V (change as necessary in order to distinguish viability dye positive vs negative events & check this against no viability dye control); Gates: Change as necessary to ensure all cells are collected. See **Figure 2d-2g**.
NOTE: FACS parameters are highly dependent on the cell types and the sorter employed and therefore needs to be optimized by the user.
2. Prepare 15 mL narrow-bottom tubes with 8 mL of ice-cold 1% BSA/HBSS for sample collections. Static inside tube and surface tension can affect collection efficiency. Invert the tubes before collections to ensure interface between surface of liquid and the inside of tube is moist.
NOTE: If working with very low cell numbers, adjust to small collection vessel as appropriate.
3. Once all cells are collected, centrifuge sample at 260 x g for 8 min.
NOTE: Before centrifugation, add 1% BSA/HBSS to wash/push cells down from the side surface and invert/mix the tube immediately after FACS.
4. Resuspend cell pellet in 1% BSA/HBSS and keep on ice. The maximum volume per sample that is compatible with Step 3 processing is 33.8 μL , so ensure that final cell dilution/resuspension volume is appropriate to obtain ideal cell number in 33.8 μL . Other dilution media options for this step (and all previous dilutions in 1% BSA/HBSS) include DMEM, and up to 40 % serum, but avoid calcium, magnesium, or EDTA containing reagents.
 1. Leave cells on ice for a minimal amount of time. Ideally co-worker should prepare all equipment and reagents for the following step (Step 3) during final steps of Step 2.
5. Cell preparation critical checks
 1. Confirm estimates of cell numbers obtained from FACS. Depending on tissue type and dissociation lengths, debris and cells can be very similar in size and shape. Thus, unless a fluorescent reporter is used, FACS cannot exclude all debris. It is recommended that a final cell count after FACS collection is performed to understand what percentage of events (according to FACS) are in fact cells for a given preparation (**Figure 2g**). Perform cell count using a hemocytometer or automated cell counter (repeat twice) and calculate the percentage of viable cells that is represented by the total events collected according to FACS machine.
 2. Validate cell preparation. Validate that no large particles (>100 μm) are present as they may clog equipment in downstream steps. Inadequate removal of debris may risk clogging the single-cell microfluidic chip. Plate remaining cells with Nuc Blue (as above) to ensure no large debris fragments are present. This will also allow for the confirmation that cells are singular (*i.e.*, not sticking together) giving confidence that downstream single cell genetic analysis represents single cells rather than multiple cells.
 3. Decide on cell numbers to sequence: There is a large range of adult tissue-derived cell numbers per sample that can be loaded into the system with up to 8 samples that can be run at one time. Authors have loaded anywhere from 500 – 50,000 cells per sample and obtained good quality scRNA-Seq datasets. More discussion regarding the most appropriate cell numbers to load can be found in the Discussion section. The final output of sequenced cell numbers depends heavily on the quality of single-cells isolated. Loading 10,000 adult tissue-derived cells can return anywhere from 1,000 to 4,000 sequenced cells (10 - 40% return). If interested in sequencing high cell numbers (~10,000 cells, the maximum number recommended for this system), then loading 25,000 - 100,000 cells will be required.

3. GEM (Gel Bead in Emulsion) Generation and Barcoding (Day 1)

NOTE: Steps 3-6 of this protocol are designed to be used in conjunction with the most common microdroplet-based single-cell platform, manufactured by 10X Genomics. Detailed guidelines for Steps 3 and 4 are outlined in the manufacturer's protocol (Refer to the Chromium Single Cell 3' protocol)^{11,12} and must be followed in conjunction with this protocol. For best results, Step 3 must be completed immediately after dissociation (Step 1) and cell isolation (Step 2) steps on day 1 of this protocol.

1. Prepare chip according to the manufacturer's protocol^{11,12}. This microdroplet-based single-cell platform uses technology that samples ~750,000 barcodes to separately index each cell's transcriptome. This is achieved by partitioning cells into Gel Bead in EMulsions (GEMs) where generated cDNA share a common barcode. During GEM generation, cells are delivered so that the majority (90 - 99%) of generated GEMs contain no cells, while the remainder, for the most part, contain a single cell.
 1. Place the chip in the chip holder.
 2. Prepare cell master mix on ice.
 3. Add 50% glycerol to unused wells and add 90 μL of the cell master mix to well 1, 90 μL of gel beads to well 2, and 270 μL of partitioning oil to well 3.
 4. Cover the chip with the gasket.
2. Load the chip and run in a single-cell controller.
 1. Eject the tray, place the chip in tray, retract the tray, and press **Play**. A single cell 3' gel bead in a GEM includes primers containing a partial Illumina R1 sequence (read 1 sequencing primer), a 16 nucleotides (nt) 10x Barcode, a 10 nt Unique Molecular Identifier (UMI), and a poly-dT primer sequence. During the run, gel beads in the controller are released and mixed with cell lysate and master mix.
 2. Collect 100 μL of the sample and place in a PCR tube.
 3. Place PCR tubes in pre-set PCR machine and run the PCR according to the kit. Following incubation, GEMs will include full-length, barcoded cDNA from poly-adenylated mRNA.
3. Following the run, place at -20 °C overnight for up to 1 week before preceding to next step.

4. Clean-Up, Amplification, Library Construction and Library Quantification (Day 2 Onwards)

NOTE: Detailed guidelines for Steps 4 are outlined in the manufacturer's protocol^{11,12}, and must be followed in conjunction with this protocol.

1. Use silane magnetic beads to remove leftover biochemical reagents/primers from GEM reaction mixture.
2. Amplify full-length, barcoded cDNA to generate sufficient mass for library construction.
3. Assess DNA yield. Prior to library construction, assess DNA yield of the sample. This will determine how many cycles to use in downstream PCR step (Sample Index PCR during library construction). Depending on the RNA content of a given sample, which can vary depending on activation states (e.g., control vs injured, etc.), cell type, and cell yield, the recommended cycle number may change.
 1. For sequencing ~3,000 tissue-derived cells (irrelevant to activation states), the authors have found that 14 cycles (samples: ~10 - 100 ng DNA) is standard.
 2. Use a Bioanalyzer for DNA analysis. Refer to the User Guide¹³.
4. Fragment sample and select the size of DNA. Prior to library construction, use enzymatic fragmentation and size selection protocols to obtain appropriate cDNA amplicon size.
5. Prepare sample for library construction. While R1 (read 1 primer sequence) is added to the molecules during GEM incubation; P5, P7 (a sample index), and R2 (read 2 primer sequence) are added during library construction.
6. Assess DNA yield. Most sequencing facilities require submission of final libraries that include DNA yield and quality information. Thus, run the bioanalyzer following the completion of the entire protocol and before transporting to the sequencing facility.
7. Store samples at -80 °C for up to 2 months.
8. Before sequencing, quantify samples using a DNA quantification kit. This can be done at the sequencing facility.

5. Library Sequencing (Day 3 Onwards)

NOTE: The single-cell transcriptome barcoding platform used in this protocol generates Illumina-compatible paired-end libraries beginning and ending with P5 and P7 sequences. Although minimum depth needed to resolve cell-type identity can be as few as 10,000–50,000 reads/cell^{15,16}, ~100,000 reads/cell is recommended as an optimal cost-coverage trade-off for adult in vivo cells (keeping in mind some cell types or minimally activated cell states will reach saturation at 30,000-50,000 reads/cell).

1. Transport cDNA libraries on dry ice to a sequencing facility equipped with an appropriate Illumina sequencer.
2. Provide the following information to the sequencing facility:
 1. Provide sample details: Sample index IDs corresponding to each library; species; genomic database for primary assembly (i.e., GRCm38 for mouse); electropherogram showing fragment sizes from the bioanalyzer (between 200 and 9,000 bp); cDNA concentration (ng/μL) and total library concentration (total yields range from 200 – 1400 ng); volume (μL) of sample.
 2. Provide sequencing requests: Quantify samples using a DNA quantification kit; adapter/index type (TruSeq DNA); plate type (Eppendorf twin.tec, Full Skirt - recommended for DNA); sequencing technology/library type (10x, full sequencing instructions and cycle recommendations)¹⁷.
3. Run shallow sequencing (optional): Studies analyzing multiple biological samples will benefit from pooling samples (aggregation) to generate a single gene-barcode matrix containing data from all samples. To minimize batch effects between samples when pooling, the read depth between different libraries should be standardized. In order to do this, an accurate approximation of single cell numbers is necessary. The MiSeq sequencer will allow shallow sequencing and is a cost-effective, practical way to obtain accurate cell estimations.

NOTE: One run using a MiSeq SR50 sequencer provides sufficient coverage to accurately estimate approximately 20,000 cells. This run will approximate the number of UMI recovered for each unique barcode. In **Figure 3a**, the header of an example (Sample 1.6) output (.csv) is shown, listing barcodes and its corresponding UMI counts as determined by confidently mapped reads.

 1. Consult with a bioinformatician to become familiar with the R programming language. Refer to DataCamp tutorials for more information¹⁸.
 2. Assess raw data obtained from the sequencer using the provided R script as a template¹⁹. Raw data refers to number of UMIs mapped to each unique cell barcode. The script reads a .csv file where the first column is a list of barcodes and the second column are its corresponding UMI counts. This script will provide a plot (**Figure 3b**) as well as the estimated number of barcoded cells in each sample. Adjust script to ensure that the inputted number of UMI counts for a given sample is at the one-third point of the first steep drop. In **Figure 3b**, this elbow falls around 225 UMIs corresponding to 3,480 barcoded cells.
 3. Comparable to full-depth sequencing using HiSeq (where 3,516 cells were successfully sequenced, **Figure 3c**), shallow sequencing estimates predicted 3,480 cells.
4. Either use cell recovery approximations (from step 5.3) or use the recovery chart found in the manufacturer's protocol²⁰ to plan lane distribution for deeper sequencing. Each sample should receive comparable coverage, so if shallow sequencing reveals that there are differing numbers of cells in each sample (which is often the case) then lane distribution should be calculated accordingly. One HiSeq flow cell (which comprises 8 lanes) can sequence up to 2.4 billion custom paired end reads. Example flow cell set-up is presented in **Figure 3d**.

6. Processing Read Files

NOTE: Sequencing a single cell 3' Library using this protocol generates raw data in binary base call (BCL) format. The Cell Ranger package is used to generate text-based FASTQ files from BCL files, perform genomic and transcriptomic alignments, gene counts, demultiplexing, and aggregation of samples. In this section, the key steps that enable users to download raw BCL data from a sequencing facility and generate filtered gene-barcode matrices ready for downstream bioinformatics is presented.

1. Use a centralized server for running the program. BCL files, FASTQ files and most of the downstream bioinformatics processing demands significant processing power.
2. Download all raw read files onto server (or FASTQ files if they are available).
 1. Consult with server administrator to set up an account on a centralized server or cluster, and to get familiar with Unix²¹.
 2. Use a fetch command appropriate for the server's operating system to download all files from the sequencing facility's server.
 1. Most sequencing facilities provide a command to download files from a secure path that can be run from the command line (see example below).
 2. Replace the "<username>" and "<password>" placeholders in command line with credentials provided.
`wget -O - "https://your_sequencing_facilitys_server.com/path_to_raw_read_files/ --no-cookies --no-check-certificate --post-data 'j_username=username&j_password=password' | wget --no-cookies --no-check-certificate --post-data 'j_username=username&j_password=password' -ci -`
 3. If only an absolute path to files is provided (*i.e.* https://your_sequencing_facilitys_server.com/path_to_raw_read_files/), insert this path into a fetch command.
3. Unzip files: If downloaded files end with a ".gz" extension, it has been compressed using the "gzip" command. To unzip, run unzip command in the command line (see example below).
`gunzip raw_read_files.gz`
4. Download the latest version of Cell Ranger onto the server as a self-contained .tar²².
 1. **Critical:** Prior to download, ensure the Linux system meets minimum requirements²³. Ensure a minimum of 8-core Intel processor with 64 GB RAM and 1 TB of free disk space.
 NOTE: Cell Ranger provides pre-built human and rodent reference transcriptomes. These can be modified using *cellranger mkref* command to detect genes like GFP²⁴.
5. Generate FASTQ files from sequencer's base call files (BCLs) using the *cellranger mkfastq* command.
 NOTE: The program will align raw reads (from FASTQ files) to a reference genome and generate gene-cell matrices for downstream analysis. It uses STAR aligner that performs splicing-aware alignment of reads to a reference genome. Only confidently mapped reads (*i.e.*, reads compatible with a single gene annotation) are used for UMI counting.
 1. For example, use the *cellranger mkfastq* command:
`cellranger mkfastq --id=sample_name \
 --run=/path/to/sample \
 --csv=csv_file_containing_lane_sample_index.csv`
6. Run *cellranger count* on FASTQ files generated using *mkfastq* to generate single-cell gene counts.
 1. For example, use the *cellranger count* command:
`cellranger count --id=sample_name \
 --transcriptome=refdata-cellranger-mm10-1.2.0 \
 --fastqs=/absolute/path/to/fastq/files \
 --sample=same_sample_name_supplied_to_cellranger_mkfastq \
 --localcores=30`
7. Multi-library aggregation (optional): To combine samples, pool *cellranger count* outputs using *cellranger aggr*. This results in a single gene-barcode matrix containing data pooled from multiple libraries. Example *cellranger aggr* command:
`cellranger aggr --id=sample_name \
 --csv=csv_with_libraryID_&_path_to_molecule_h5.csv \
 --normalize=mapped`
 NOTE: Libraries can be aggregated using three normalization modes (mapped, raw, none). Mapped is recommended as it subsamples higher-depth libraries until all libraries have equal sequencing depth²⁵.
8. For immediate visualization/analysis of data, import the .cloupe output file (generated using *cellranger count* or *cellranger aggr*) into 10x Loupe Cell Browser²⁶.

7. Advanced Analysis of scRNA-Seq Datasets

NOTE: A complete scRNA-Seq tools database can be found at scRNA-tools^{3,27}. Below is a framework for unsupervised cell clustering using Seurat² and pseudotemporal ordering using Monocle⁶. Although much of this work can be done on a local computer, the following steps assume that computation will be completed using an institutional server.

1. Download the most recent version of Miniconda onto server account using the Linux platform²⁸.
2. Install the latest version of R using conda²⁹.
3. Plot data using the provided Seurat R script as a template³⁰.
 NOTE: Seurat is an R-based toolkit that enables quality control checks, clustering, differential gene expression analysis, marker gene identification, dimensionality reduction, and visualization of scRNA-Seq data. A comprehensive description of Seurat coding and tutorials can be found on the Satija Lab website³¹.
4. Plot data using the provided Monocle R script as a template³².
 NOTE: Monocle is another R-based toolkit that enables visualization of expression changes over pseudotime and identifies genes underlying cell fate decisions. A comprehensive description of Monocle coding and tutorials can be found on the Monocle website³³.
5. R-packages such as kBET can be employed to test and correct batch effects as a result of pooling datasets³⁴.

8. NCBI's GEO and SRA Submissions

NOTE: Since easy access to raw sequencing files ensure reproducibility and reanalysis, accessioned submissions to online publicly available repositories are recommended or required prior to manuscript submission. National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) are publicly accessible data repositories for high-throughput sequencing data^{35,36}.

1. Register for NCBI's GEO Submitter account³⁷.
2. Complete GEO submission which includes three components compiled into a directory/folder (titled as the GEO submitter's username): 1) Metadata record (one spreadsheet per project submission); 2) Raw Data Files; 3) Processed Data Files.
 1. Download and complete the Metadata spreadsheet³⁸. The following public GEO submission can be used as a guide (GSE100320)³⁹. Place the spreadsheet in the directory.
 2. Place Raw Data Files generated from *cellranger count* script for all libraries into the directory.
 3. Place Processed Data Files (filtered barcodes.tsv, genes.tsv, and matrix.mtx files) generated from *cellranger count* script for all libraries into the directory.
3. Use GEO submitter's FTP server credentials to transfer directory containing all three components. For Linux/Unix users: ncftp, lftp, ftp, sftp, and ncftpput can be used.
4. Notify GEO for all transfers³⁸.

Representative Results

The repertoire of open source packages designed to analyze scRNA-Seq datasets has increased dramatically⁴⁰ with the majority of these packages use R-based languages³. Here, representative results using two of these packages are presented: assessing unsupervised grouping of single cells based on gene expression, and ordering single cells along a trajectory in order to resolve cell heterogeneity and deconstruct biological processes.

Figure 4 illustrates the use of Seurat for pre-processing quality checks and downstream bioinformatics analysis. First, filtration and removal of deviant cells from analysis is essential for quality checking. This was done using violin (**Figure 4a**) and scatter plots (**Figure 4b**) to visualize the percentage of mitochondrial genes, number of genes (nGene), and number of UMI (nUMI) to identify cell doublets and outliers. Any cell with a clear outlier number of genes, UMI, or percentage of mitochondrial genes was removed using Seurat's FilterCells function. Since Seurat uses principal component (PC) analysis scores to clusters cells, determining statistically significant PCs to include is a critical step. Elbow plots (**Figure 4c**) were used for PC selection, in which PCs beyond the plateau of the 'standard deviation of PC' axis were excluded. The resolution of clustering was also manipulated demonstrating that the number of clusters can be changed, ranging from 0.4 (low resolution leading to fewer cell clusters, **Figure 4d**) to 4 (high resolution leading to higher cell clusters, **Figure 4e**). At low resolution, it is likely that each cluster represents a defined cell type, whereas at high resolution this may also represent subtypes or transitional states of a cell population. In this instance, low-resolution cluster settings were used for further analyzing expression heatmaps (using Seurat's DoHeatmap function) to identify the most highly expressed genes in a given cluster (**Figure 4f**). In this instance, the most highly expressed genes were identified by assessing differential expression in a given cluster versus all other clusters combined, demonstrating that each cluster was uniquely represented by defined genes. Additionally, individual candidate genes can be visualized on tSNE plots using Seurat's FeaturePlot function (**Figure 4g**). This allowed for deciphering whether there were clusters that represented macrophages. Using FeaturePlot, we found that both cluster 2 and 4 were expressing Cd68 - a pan-macrophage marker.

The Monocle package was used for corroborating cell clusters identified in Seurat, and for building cell trajectories, or pseudotemporal ordering, to recapitulate biological processes (**Figure 5**). Pseudotemporal ordering can be used for samples where single-cell expression profiles are expected to follow a biological time course. Cells can be ordered along a pseudotemporal continuum to resolve intermediate states, bifurcation points of two alternative cell fates, and identify gene signatures underlying acquisition of each fate. Firstly, similar to Seurat's filtration, poor-quality cells were removed such that the distribution of mRNA across all cells was log normal and fell between upper and lower bounds as identified in **Figure 5a**. Then, using Monocle's newCellTypeHierarchy function, single cells were classified and counted using known lineage marker genes (**Figure 5b, 5c**). For example, cells expressing PDGF receptor alpha or Fibroblast Specific Protein 1 were assigned to Cell Type #1 to create a criterion for defining fibroblasts. Next, this population (Cell Type #1) was assessed to decipher fibroblast trajectories. To do this, Monocle's differential GeneTest function was utilized, which compared the cells representing the extreme states within the population and found differential genes for ordering the remaining cells in the population (**Figure 5d**). By applying manifold learning methods (a type of non-linear dimensionality reduction) across all cells, a coordinate along a pseudotemporal path was assigned. This trajectory was then visualized by cell state (**Figure 5e**) and pseudotime (**Figure 5f**).

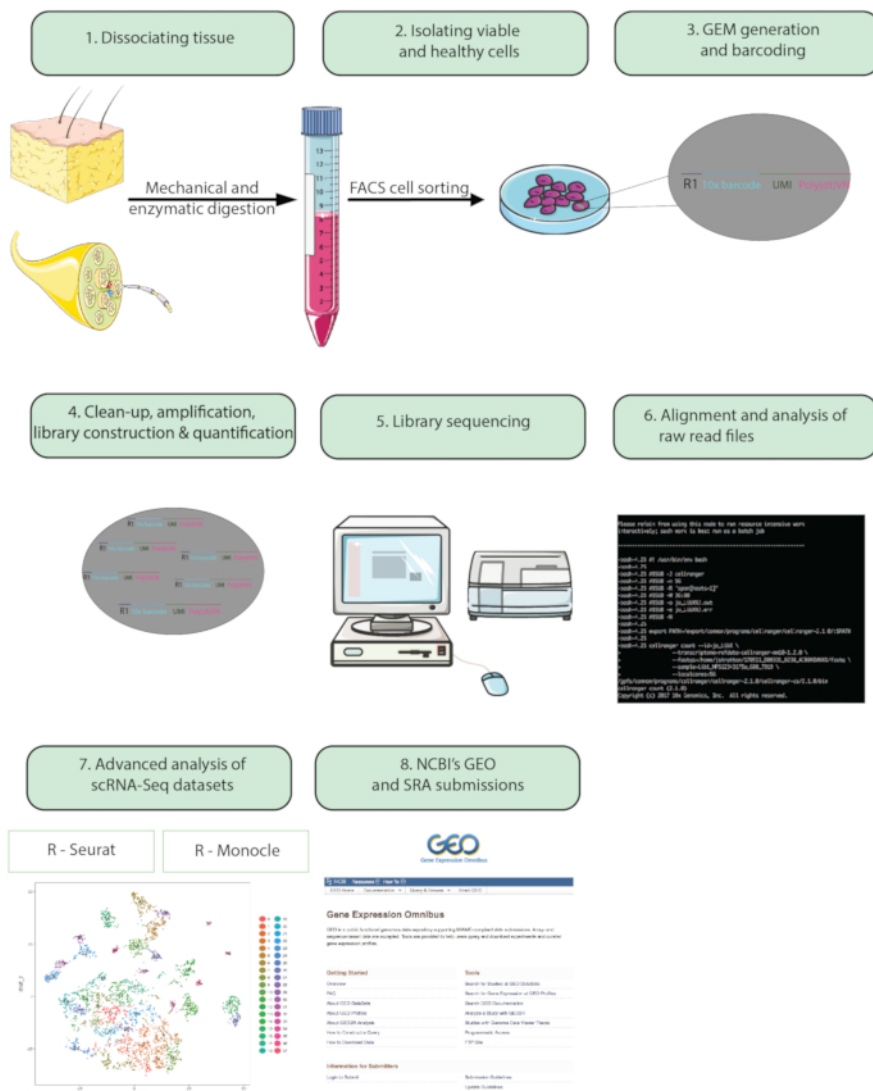


Figure 1: Flow chart. Steps from whole animal preparation to analyzing single cell RNA-Seq datasets to submitting final datasets to a publicly available repository. Gel beads in Emulsion (GEMs) refer to beads with barcoded oligonucleotides which encapsulate thousands of single cells. [Please click here to view a larger version of this figure.](#)

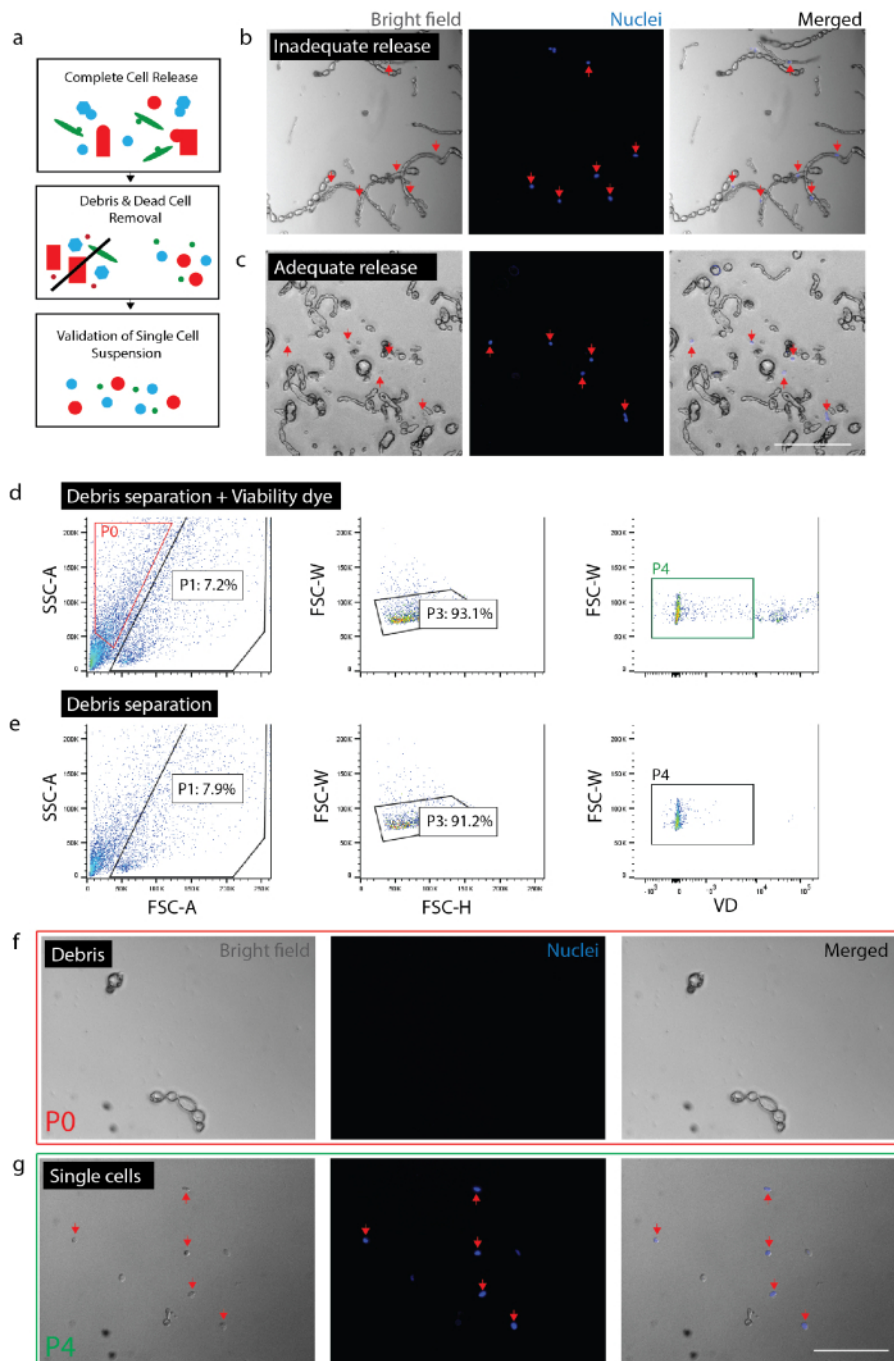


Figure 2: Creating viable single cell suspension from nerve tissue. (a) Cartoon overview of quality control checks. (b) Cells and debris with cells still incorporated in debris (red arrows). (c) Cells released from debris (red arrows). (d) Cell isolation by FACS. P0: debris fraction; P1: cell-like fraction; P3: exclusion of duplets; P4: viability dye (Sytox Orange) negative fraction. (e) No viability dye control. (f) Image of P0 fraction representing isolated debris. (g) Image of P4 fraction representing isolated viable cells (red arrows). (b)(c)(f) and (g) had nuclear dye added 20 minutes before imaging. Scale Bars: 80 μ m. [Please click here to view a larger version of this figure.](#)

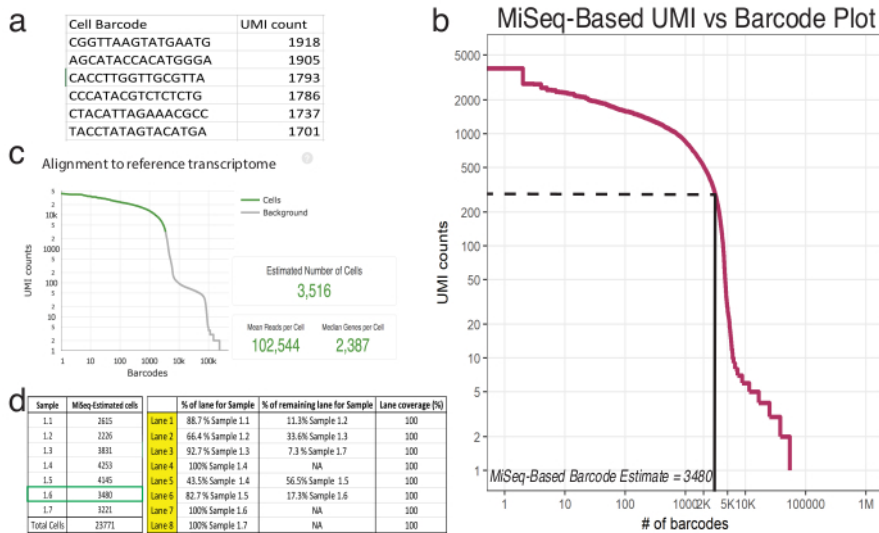


Figure 3: Shallow sequencing predicts the number of recovered cells in 10X processed samples. (a) An example (Sample 1.6) of MiSeq-generated csv listing cell barcodes and its corresponding UMI counts as determined by confidently mapped reads. (b) Barcode rank plot for Sample 1.6 shows one significant drop in UMI count as a function of cell barcodes. The dashed and solid lines represent the cutoff between cells and background as determined by visual inspection. (c) Cell barcodes observed using the Cell Ranger pipeline post-HiSeq reveals shallow sequencing accurately approximated the number of cells for Sample 1.6. (d) An example of a flow-cell set-up based on shallow sequencing derived cell estimates. For Sample 1.6, since shallow sequencing predicted 3480 cells, 1.17 lanes were assigned to ensure >100,000 reads per cell sequencing coverage in HiSeq. Note: All lanes must add to 100%. [Please click here to view a larger version of this figure.](#)

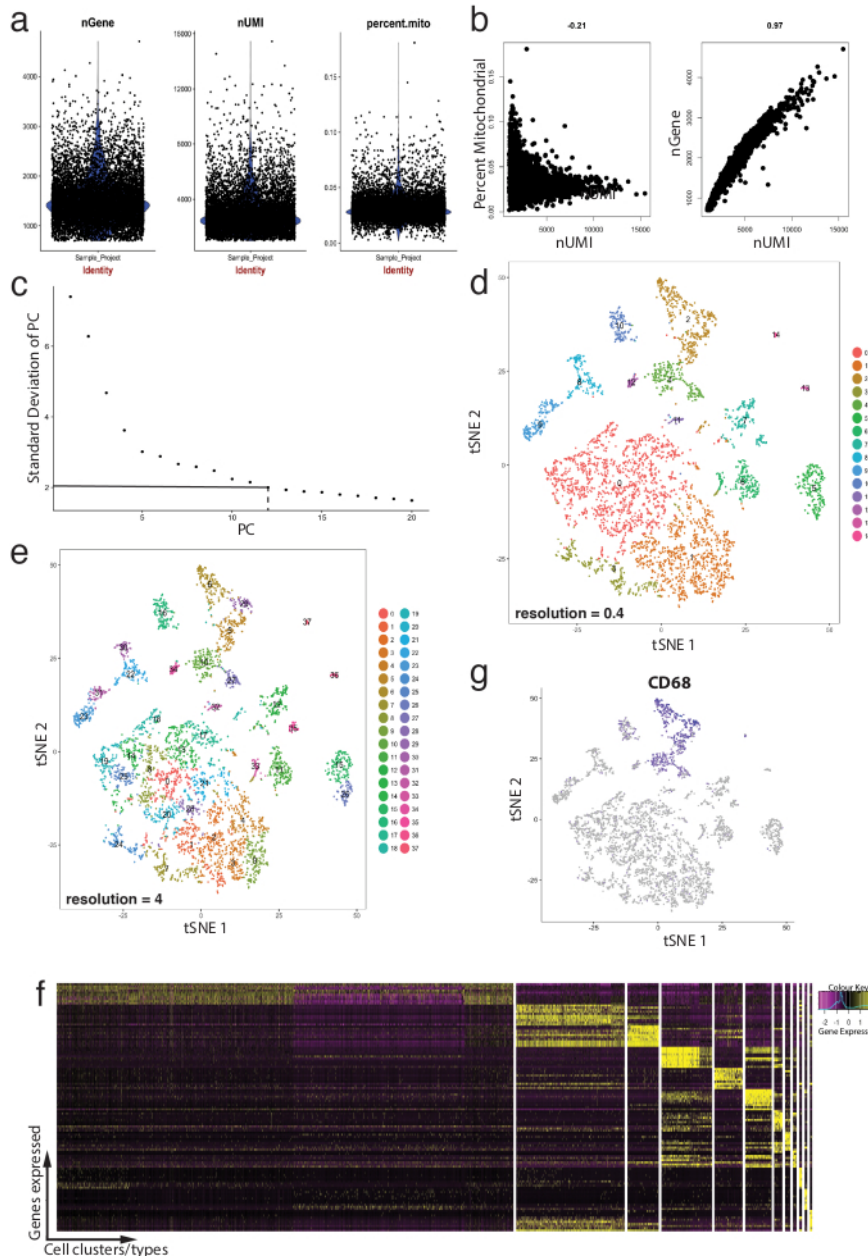


Figure 4: Quality control and bioinformatics of single-cell RNA-Seq dataset using Seurat R package. (a) Plots of quality control metrics which include number of genes, number of unique molecular identifiers (UMIs), and the percentage of transcripts mapping to the mitochondrial genome. (b) Sample gene plots detecting cells with deviant levels of mitochondrial transcripts and UMIs. (c) Sample elbow plot used for ad hoc determination of statistically significant PCs. The dashed and dot-dashed lines represent the cutoff where a clear "elbow" becomes apparent in the graph. PC dimensions before this elbow are included in downstream analysis. (d, e) Graph-based cell clusters visualized at two different resolutions in a low-dimensional space using a tSNE plot. (f) Top marker genes (yellow) for each cluster visualized on an expression heatmap using Seurat's DoHeatmap function. (g) Visualizing marker expression of, for example, Cd68 gene representing macrophages (purple) using Seurat's FeaturePlot function. This suggests that cluster 2 and 4 (in panel d) of this dataset represents macrophages. [Please click here to view a larger version of this figure.](#)

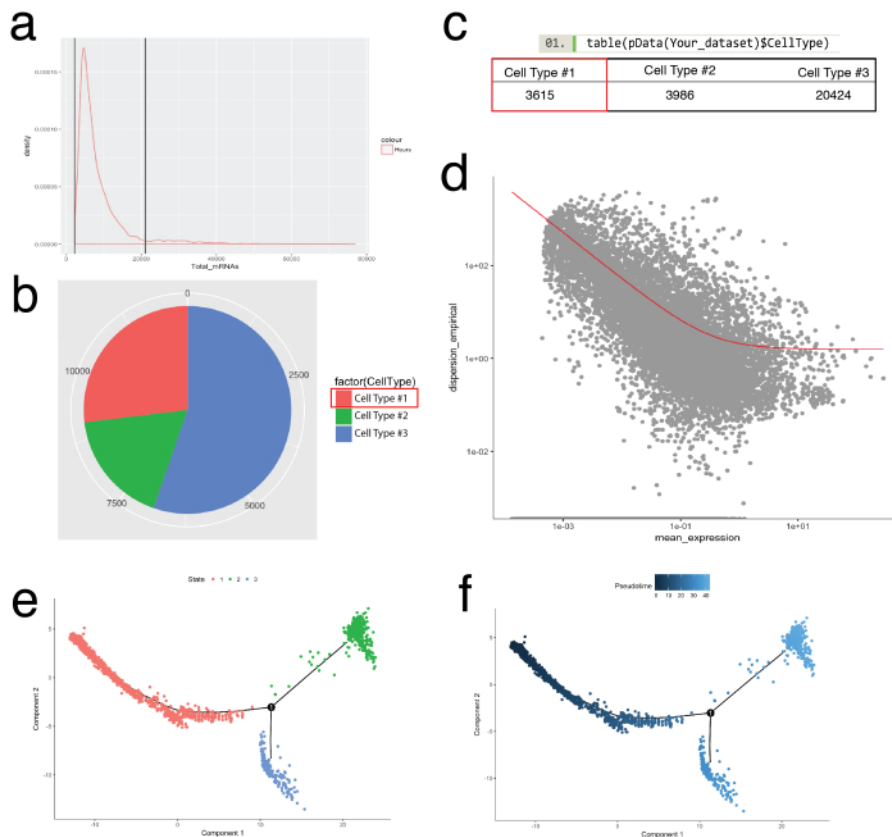


Figure 5: Cell categorization and ordering along pseudotemporal trajectory using Monocle toolkit. (a) Inspecting the distribution of mRNA (inferred from UMI counts) across all cells in a sample. Only cells with mRNA between 0 - ~20,000 were used for downstream analysis. (b, c) Assigning and counting cell types based on known lineage cell markers. For example, cells expressing PDGF receptor alpha or Fibroblast Specific Protein 1 were assigned to Cell Type #1 representing pan-fibroblasts using Monocle's newCellTypeHierarchy function. Number of different cell types can be visualized as a pie chart (b) and as a table (c). (d) Using Cell Type #1 (fibroblasts) as an example, the genes used for ordering cells can be visualized using a scatter plot that demonstrates gene dispersion vs. mean expression. The red curve shows the cutoff for genes used for ordering calculated by the mean-variance model using Monocle's estimateDispersions function. Genes that meet this cutoff were used for downstream pseudotime ordering. (e, f) Visualization of cell trajectories in a reduced two-dimensional space colored by cell's "State" (e), and by Monocle-assigned "Pseudotime" (f). [Please click here to view a larger version of this figure.](#)

Discussion

This protocol demonstrates how the appropriate preparation of single cells can uncover the transcriptional heterogeneity of thousands of single cells and discriminate functional states or unique cellular identities within a tissue. The protocol does not require fluorescent reporter proteins or transgenic tools and can be applied to the isolation of single cells from various tissues of interest including those from humans; keeping in mind each tissue is unique and this protocol will require some degree of adjustment/modification.

The diverse and highly dynamic transcriptional programs within cells have emphasized the value of single-cell genomics. Aside from isolating high-quality RNA, a critical sample preparation step necessary for high quality datasets is ensuring that cells are completely released from tissue and that cells are healthy and intact. This is relatively straight forward for collecting cells that are easily released, such as circulating cells or in tissues where cells are loosely retained, such as in lymphoid tissues. But this can be challenging for other adult tissues, due to the highly-developed cellular architecture spanning large distances, surrounding extracellular matrix and the often-rigid cytoskeletal proteins involved in maintaining cell structure. Even with appropriate dissociation techniques for the full release of cells, there is potential that the rigorous and often lengthy processing required would alter mRNA quality and cell integrity. In addition, the high temperatures used for enzyme-assisted dissociation also affect transcriptional signatures^{29,30}. The intent of the protocol is to present quality control checks, using tissues such as the myelinated adult nerve and the extracellular matrix-rich adult skin, to demonstrate how optimization can help to overcome these obstacles.

A major consideration when designing any scRNA-Seq experiment is the choice of sequencing depth. Sequencing can be highly multiplexed and read depth can vary from being very low using Drop-Seq² to up to 5 million reads/cell¹⁴ using a full-length RNA-Seq method such as Smart-Seq. Most scRNA-Seq experiments can detect moderate-to-high expression transcripts with sequencing as low as 10,000 reads/cell, which is usually sufficient for cell type classification^{41,42}. Shallow sequencing depth is of value to save on sequencing costs when trying to detect rare cell populations across complex tissues where thousands of cells may be needed to confidently ascribe rare populations. But shallow depth sequencing is not adequate when detailed information on gene expression and processes associated with subtle transcriptional signatures is necessary. Currently, it is estimated that the large majority of genes in a cell are detected with 500,000 reads/cell, but this can vary depending on the protocol and tissue type^{43,44}. While full-length transcript sequencing circumvents the need for assembly and can, therefore, detect novel or rare splice variants, sequencing costs often limit scaling such approaches to examine thousands of cells comprising a complex tissue system.

In contrast, 3' tagged single-cell libraries such as the ones described in this protocol typically have lower complexity and require shallower sequencing. It is important to note that libraries generated using the described protocol can be sequenced on one of five supported sequencers: 1) NovaSeq, 2) HiSeq 3000/4000, 3) HiSeq 2500 Rapid Run and High Output, 4) NextSeq 500/550, and 5) MiSeq.

An alternative approach to single cell RNA-Seq, that reduces the need for delicate tissue and cell handling yet maintains some of the benefits of single cell RNA-Seq, is the analysis of RNA from single nuclei⁴⁵. This approach allows more rapid processing reducing RNA degradation, and more extreme measures to ensure adequate release of nuclei, and thus likely allows for a more confident capture of the transcriptional profiles representing all cells within a given tissue. This would, of course, only provide a portion of the transcriptional activity present within a given cell, thus depending on what experimental objectives are of interest this approach may or may not be appropriate.

Besides the complete characterization of cellular identities within a given tissue, one of the most valuable analyses for scRNA-Seq datasets is the assessment of intermediate transcriptional states across 'defined' cell populations. These intermediary states can impart insights into the lineage relationships between cells within identified populations, which was not possible with traditional bulk RNA-Seq approaches. Several scRNA-Seq bioinformatic tools have now been developed to elucidate this. Such tools can assess the processes involved in, for example, cancer cells transitioning to an oncogenic/metastatic state, stem cells maturing into diverse terminal fates or immune cells shuttling between active and quiescent states. Subtle transcriptome differences in cells may also be indicative of lineage biases that, recently developed bioinformatic tools like FateID, can infer⁴⁷. Since the distinctions between transitioning cells can be difficult to ascertain given the transcriptional differences may be subtle, deeper sequencing may be necessary⁴⁶. Fortunately, coverage of a shallowly sequenced library can be increased if interested in probing the dataset further by re-running the library on another flow cell.

Taken together, this protocol provides an easy-to-adapt workflow that enables users to transcriptionally profile hundreds to thousands of single-cells within one experiment. The final quality of a scRNA-Seq dataset relies on optimized cell isolation, flow cytometry, cDNA library generation, and interpretation of raw gene-barcode matrices. To this end, this protocol provides a comprehensive overview of all key steps that can be easily modified to enable studies of diverse tissue types.

Disclosures

No disclosures

Acknowledgements

We acknowledge the support staff at the UC DNA Services Facility, as well as the Animal Care facility staff at the University of Calgary. We thank Matt Workentine for his bioinformatics support and Jens Durruthy for his technical support. This work was funded by a CIHR grant (R.M. and J.B.), a CIHR New Investigator Award to J.B., and an Alberta Children's Health Research Institute Fellowship (J.S.).

References

- Shalek, A.K., *et al.* Single-cell RNA-seq reveals dynamic paracrine control for cellular variation. *Nature*. **510**, 363-369 (2014).
- Macosko, E.Z., *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. **161**, 1202–1214 (2015).
- Zappia, L., Phipson, B., Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *bioRxiv:206573*. <https://www.biorxiv.org/content/early/2018/03/23/206573> (2018).
- Dulken, B.W., Leeman, D.S., Boutet, S.C., Hebestreit, K., Brunet, A. Single cell transcriptomic analysis defines heterogeneity and transcriptional dynamics in the adult neural stem cell lineage. *Cell Reports*. **18** (3), 777-790 (2017).
- Llorens-Bobadilla, E., *et al.* Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury. *Cell Stem Cell*. **17** (3), 329-340 (2015).
- Trapnell, C., *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*. **32**, 381–386 (2014).
- Aibar, S., *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nature Methods*. **14**, 1083-1086 (2017).
- Mayer, C., *et al.* Developmental diversification of cortical inhibitory interneurons. *Nature*. **555** (7697), 457-462 (2018).
- Stratton, J.A., *et al.* Purification and Characterization of Schwann Cells from Adult Human Skin and Nerve. *eNeuro*. **4** (3), ENEURO.0307-16.2017 (2017).
- Biernaskie, J.A., McKenzie, I.A., Toma, J.G., Miller, F.D. Isolation of skin-derived precursors (SKPs) and differentiation and enrichment of their Schwann cell progeny. *Nature Protocols*. **1** (6), 2803-2812 (2007).
- 10X Genomics. *User Guides*. <https://www.10xgenomics.com/resources/user-guides/> (2018).
- 10X Genomics. *Chromium Single Cell 3' Training Module*. <http://go.10xgenomics.com/training-modules/single-cell-gene-expression> (2018).
- Agilent*. <https://www.agilent.com/en-us/library/usermanuals?N=135> (2018).
- Kolodziejczyk, A.A., *et al.* Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell*. **17**, 471–485 (2015).
- 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).
- Pollen, A.A., *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nature Biotechnology*. **32**, 1053–1058 (2014).
- 10X Genomics. *Sequencing Requirements for Single Cell 3'*. <https://support.10xgenomics.com/single-cell-gene-expression/sequencing/doc/specifications-sequencing-requirements-for-single-cell-3> (2018).
- DataCamp. *Introduction to R*. <https://www.datacamp.com/courses/free-introduction-to-r> (2018).

19. Droplet-based, high-throughput single cell transcriptional analysis of adult mouse tissue using 10X Genomics' Chromium Single Cell 3' (v2) system: From tissue preparation to bioinformatic analysis. <https://figshare.com/s/97b83e649e5eefd01357> (2018).
20. 10X Genomics. *User Guides*. <https://www.10xgenomics.com/resources/user-guides/> (2018).
21. *UNIX Tutorial for Beginners*. <http://www.ee.surrey.ac.uk/Teaching/Unix/> (2018).
22. 10X Genomics. *Creating a Reference Package with cellranger mkref*. <https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/advanced/references> (2018).
23. 10X Genomics. *System Requirements*. <https://support.10xgenomics.com/single-cell-gene-expression/software/overview/system-requirements> (2018).
24. 10X Genomics. *Software Downloads*. <https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest> (2018).
25. 10X Genomics. *Aggregating Multiple Libraries with cellranger aggr*. https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/aggregate#depth_normalization (2018).
26. 10X Genomics. *Loupe Cell Browser Gene Expression Tutorial*. <https://support.10xgenomics.com/single-cell-gene-expression/software/visualization/latest/tutorial> (2018).
27. scRNA-tools. *A table of tools for the analysis of single-cell RNA-seq data*. <https://www.scrna-tools.org/> (2018).
28. Conda. *Downloading conda*. <https://conda.io/docs/user-guide/install/download.html>. (2018).
29. Anaconda. *r / packages / r 3.5.1*. <https://anaconda.org/r/r> (2018).
30. <https://figshare.com/s/97b83e649e5eefd01357>. (2018).
31. *Satija Lab*. Seurat - Guided Clustering Tutorial. https://satijalab.org/seurat/pbmc3k_tutorial.html (2018).
32. <https://figshare.com/s/97b83e649e5eefd01357>. (2018).
33. *Monocle*. <http://cole-trapnell-lab.github.io/monocle-release/docs/#constructing-single-cell-trajectories> (2018).
34. *Github*. An R package to test for batch effects in high-dimensional single-cell RNA sequencing data. (2018).
35. Edgar, R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*. **30**, 207–210 (2002).
36. Leinonen, R., Sugawara, H., Shumway, M. The sequence read archive. *Nucleic Acids Research*. **39**, D19-21 (2011).
37. NIH. *GenBank Submission Portal Wizards*. https://www.ncbi.nlm.nih.gov/account/register/?back_url=/geo/submitter/ (2018).
38. NIH. *Submitting data*. <https://submit.ncbi.nlm.nih.gov/geo/submission/> (2018).
39. Shah, P.T., *et al*. Single-Cell Transcriptomics and Fate Mapping of Ependymal Cells Reveals an Absence of Neural Stem Cell Function. *Cell*. **173**, 1045–1057.e9 (2018).
40. Anon. Method of the Year 2013. *Nature Methods*. **11**, 1 (2013).
41. Adam, M, Potter, A.S., Potter, S.S. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development*. **144** (19), 3625-3632 (2017).
42. Wu, Y.E., Pan, L., Zuo, Y., Li, X., Hong, W. Detecting activated cell populations using single-cell RNA-seq. *Neuron*. **96** (2), 313-329 (2017).
43. Zeigenhain, C., *et al*. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular Cell*. **65**, (4) 631-643 (2017).
44. Wu, A.R., *et al*. Quantitative assessment of single-cell RNA-sequencing methods. *Nature Methods*. **11**, (1) 41-46 (2014).
45. Habib, N., *et al*. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science*. **353**, 929-928 (2016).
46. Janes, K.A. Single-cell states versus single-cell atlases - two classes of heterogeneity that differ in meaning and method. *Current Opinions in Biotechnology*. **39**, 120-125 (2016).
47. Herman, J.S., Sagar, Grün, D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nature Methods*. **15** (5), 379-386 (2018).