Experimental Considerations for Single-Cell RNA Sequencing Approaches

Quy H. Nguyen[1], Nicholas Pervolarakis[2], Kevin Nee[1] and Kai Kessenbrock[1]*

[1] *Department of Biological Chemistry, University of California, Irvine, Irvine, CA, United States,* [2] *Center for Complex Biological Systems, University of California, Irvine, Irvine, CA, United States*

Single-cell transcriptomic technologies have emerged as powerful tools to explore cellular heterogeneity at the resolution of individual cells. Previous scientific knowledge in cell biology is largely limited to data generated by bulk profiling methods, which only provide averaged read-outs that generally mask cellular heterogeneity. This averaged approach is particularly problematic when the biological effect of interest is limited to only a subpopulation of cells such as stem/progenitor cells within a given tissue, or immune cell subsets infiltrating a tumor. Great advances in single-cell RNA sequencing (scRNAseq) enabled scientists to overcome this limitation and allow for in depth interrogation of previously unexplored rare cell types. Due to the high sensitivity of scRNAseq, adequate attention must be put into experimental setup and execution. Careful handling and processing of cells for scRNAseq is critical to preserve the native expression profile that will ensure meaningful analysis and conclusions. Here, we delineate the individual steps of a typical single-cell analysis workflow from tissue procurement, cell preparation, to platform selection and data analysis, and we discuss critical challenges in each of these steps, which will serve as a helpful guide to navigate the complex field of single-cell sequencing.

Keywords: single-cell genomics, single-cell analysis, cell isolation, computational biology, cellular heterogeneity

INTRODUCTION

Elucidating cellular heterogeneity represents a major scientific challenge in many areas of biology and biomedical research including developmental and stem cell biology, immunology, neurobiology, and cancer research (Wagner et al., 2016). Recent convergence of next generation sequencing (NGS) and bioengineering approaches to manipulate individual cells has led to unbiased single-cell DNA (Navin et al., 2011), RNA (Pollen et al., 2014; Treutlein et al., 2014; Tanay and Regev, 2017), and ATAC (Buenrostro et al., 2015) sequencing. These technological advances are redefining our understanding of how biological systems function and have formed the basis for large-scale, international collaborations such as the Human Cell Atlas project (Rozenblatt-Rosen et al., 2017). In this spirit, a recent endeavor using microwell-based single-cell RNAseq (scRNAseq) created the first cell atlas to map out most tissues of the mouse (Han et al., 2018). Moreover, scRNAseq has provided critical new insights into key developmental processes such as the earliest steps of cardiovascular lineage segregation in mice (Lescroart et al., 2018), and our recent work utilized scRNAseq to reveal the spectrum of cellular heterogeneity within the human

breast epithelium identifying three major cell types each harboring multiple distinct cell states (Nguyen et al., 2018).

Due to the high sensitivity of these methods, in particular scRNAseq, it can be difficult to choose an adequate approach to minimize batch effects and unwanted technical variation that may overshadow true biological insights. Here, we provide helpful insights and delineate a step-wise approach for designing single-cell analysis workflows (**Figure 1**).

CELL DISSOCIATION AND SINGLE-CELL PREPARATION

The process of single-cell preparation is arguably the greatest source of unwanted technical variation and batch effects in any single-cell study (Tung et al., 2017). Different tissues can vary significantly in extracellular matrix (ECM) composition, cellularity, and stiffness, and therefore dissociation protocols must be optimized for the specific tissue type of interest. Conventional protocols for single-cell preparation typically involve the following steps: (1) tissue dissection, (2) mechanical mincing, (3) enzymatic/proteolytic ECM breakdown (e.g., dispase, collagenase, trypsin) often accompanied by mechanical agitation, and (4) optional enrichment for cell types of interest by flow cytometry, bead-based immune-selection, differential centrifugation, or sedimentation. Each step can affect the cells' expression signatures, and should therefore be carefully optimized to introduce the least artifact. An optimal tissue dissociation protocol will yield as many viable cells as possible in the shortest possible duration without preferentially depleting or significantly altering the frequencies of certain cell types.

Recent advances in bioengineering of innovative microfluidic cell dissociation devices (Qiu et al., 2014) have the potential to radically change the way tissue samples are dissociated into single cells, while avoiding inter-assay variation due to human handling of the tissue. Several microfluidic devices have been optimized for streamlined tissue digestion, cell dissociation, filtering, and polishing. In brief, these devices were designed to work with tissue sequentially through progressively smaller size scales, starting from tissue specimen, through cellular aggregates and clusters, and finally eluting a solution containing close to 100% single cells, which will be ideal for scRNAseq applications. In addition, new semi-automated commercially available systems can help streamline tissue dissociation (e.g., Miltenyi gentleMACS). These devices offer tissue-type specific kits that may allow more reproducible, time-saving and efficient tissue dissociation and single-cell preparation (Meeson et al., 2013; Baldan et al., 2015). Ultimately, determining a "best practices" dissociation strategy through heuristic optimization will be critical for downstream single-cell library quality.

Cell Type Enrichment
There are various methods for isolating specific cell populations or removal of unwanted populations that should be optimized for any specific tissues type. Manual isolation utilizing magnetic beads or gradient purification are potential methods for removal of unwanted cells such as de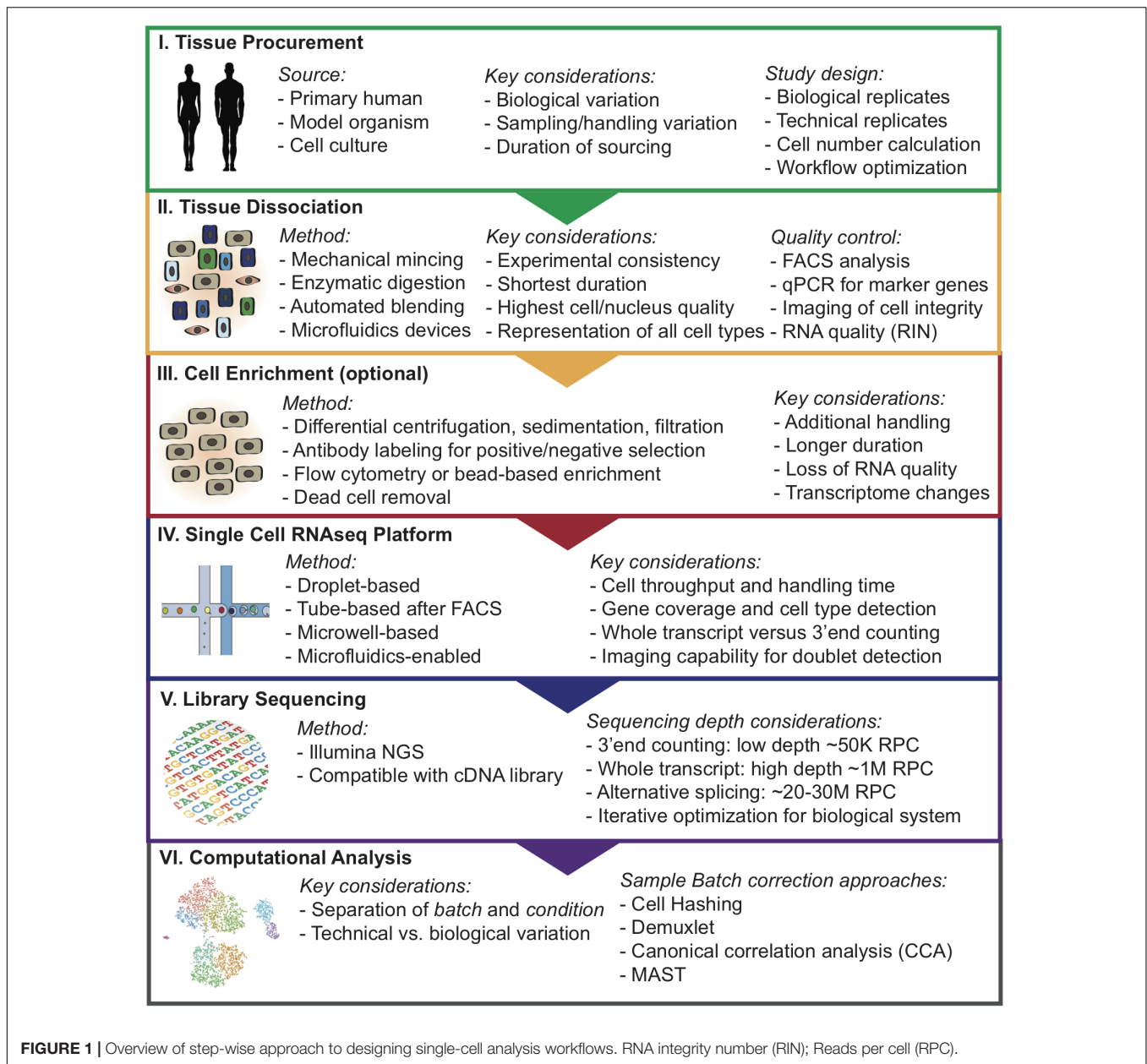ad cells. Flow cytometry is a widely used, high-throughput method to enrich for rare cells such as hematopoietic stem cells (Radbruch and Recktenwald, 1995; Will and Steidl, 2010). However, these methods are not without drawbacks, since they can introduce artificial stress on cells and change their expression profile (Van Den Brink et al., 2017). Methods that involve antibody binding for purification can also affect the cell expression profile if binding of the antibodies to cell surface molecules induce intracellular signaling (Kornbluth and Hoover, 1989; Christaki et al., 2011). Flow cytometry-isolated cells are exposed to high pressure during sorting and these osmotic and pressure changes introduced to cells during cell sorting and handling can induce change to the cell expression profile of multiple cell types (Xiong et al., 2002; Romero-santacreu et al., 2009; Van Den Brink et al., 2017).

Quality Control
Due to the high cost of single-cell sequencing experiments, careful quality control measurements should be executed. The performance of alternative protocols can be assessed using a number of readouts. A useful first metric can be acquired using imaging of viability such as using the Countess platform (Thermo Fisher Scientific). Flow cytometry is particularly valuable to measure several critical metrics simultaneously, such as cell viability, and contamination with doublets and small cell clusters which can confound single-cell sequencing results. Flow cytometry can also be used to evaluate whether cell populations of interest, such as immune cells, stromal fibroblasts, or stem cell populations, are maintained in the cell preparation and in the appropriate frequency. Finally, an additional metric on RNA quality can be acquired using the RNA integrity number (RIN) method (Schroeder et al., 2006).

SINGLE-CELL TRANSCRIPTOMIC PLATFORM

Protocols for transcriptome analysis have advanced rapidly, resulting in several robust methods which range in cell and mRNA capture strategy, barcoding, throughput, and level of automation (Fan et al., 2015; Macosko et al., 2015). Selection of the optimal approach depends largely on the research question. Recent high-throughput protocols for scRNAseq have dramatically increased scalability through automation, increasing the number of cells that can be processed simultaneously, and decreasing reagent cost through reaction miniaturization. Using microwell-based (Cytoseq, Wayfergen), microfluidics-based (Fluidigm C1 HT), or droplet-based (inDrop, Drop-seq, and 10× Chromium) approaches, hundreds to thousands of cells can be captured in a single experiment (Islam et al., 2014; Picelli et al., 2014; Klein et al., 2015; Heath et al., 2016; Zheng et al., 2017). The newest of these protocols utilize beads functionalized with oligonucleotide primers, which each contain a universal PCR priming site, a cell-specific barcode, an mRNA capture sequence, and Unique Molecular Identifiers (UMI). Individual cells are captured in wells or droplets with a single bead. Cell-specific barcode are similar within a droplet but unique UMI sequence on the primer allows for individual transcripts within

FIGURE 1 | Overview of step-wise approach to designing single-cell analysis workflows. RNA integrity number (RIN); Reads per cell (RPC).

a cell to be counted. This provides a quantitative readout of the number of transcripts of each gene detected in a cell, thereby reducing the effects of amplification duplicates that occur with earlier technologies (Ramsköld et al., 2012; Patel et al., 2014). High-throughput 3′-end counting approaches have several important limitations. Since only the 3′-end of each mRNA are sequenced, differential splicing analyses are not feasible (Macosko et al., 2015; Heath et al., 2016). High-throughput approaches typically only achieve ∼10% transcriptome coverage, relative to ∼40% for full-length scRNAseq protocols that use Switching Mechanism at 5′End of RNA Template (SMART) chemistry (Tirosh et al., 2016; Yuan et al., 2017). This is partly due to lower mRNA capture efficiency, but also due to lower sequencing depth. Single-cell qPCR platforms (e.g., Fluidigm C1 and Biomark)

remain superior in sensitivity for detecting low-expressed genes (Lawson et al., 2015).

Protocols for processing rare cells usually involve an upstream capture step by flow cytometry or micromanipulation, followed by dispensing single cells into microtubes or microwell plates. Studies investigating rare cell populations that require selection via specific markers (e.g., adult tissue stem cell populations), are best performed using these protocols. Single-cell libraries are prepared using SMART-based chemistry, which utilizes a template-switching oligonucleotide (TSO) (Tirosh et al., 2016). This TSO can be used to prime off of the untemplated nucleotides added by the reverse transcriptase, enabling subsequent PCR using a single primer and capture of full length transcripts (Tirosh et al., 2016; Yuan et al., 2017). cDNAs are then

amplified by PCR and libraries are prepared for sequencing using standard protocols. Although there have been several large scale projects utilizing these protocols, because they are manual in nature and utilize larger microliter reaction volumes, they limit the number of cells that can be processed at reasonable cost.

Another area of ongoing debate is how to determine how many cells one should be analyzed to reach sufficient statistical power. Several methods have been developed using power analysis statistics, such as Scotty[1] or web-based tools[2], but one must estimate the number and expected frequencies of cell populations present in the sample, and such information is often not available. Therefore, these decisions are usually made based on logistical restraints (i.e., the number of cells available), financial considerations, or re-iterative experiments where an initial sample of cells is sequenced to get a sense for overall population structure, and then increasing numbers of cells are sequenced until one is satisfied that all the main populations have been identified.

SINGLE NUCLEI ISOLATION AND SEQUENCING

Single-cell RNA sequencing methods are optimal when cells can be harvested intact and viable (Grindberg et al., 2013). However, certain cell types (e.g., neurons, adipocytes), are not amenable to standard organ dissociation protocols, since enzymatic and mechanical forces easily disrupt the cytoplasmic contents (Habib et al., 2017). In these cases, an option could be to isolate intact nuclei for single-nucleus RNAseq (snRNAseq) (Grindberg et al., 2013; Habib et al., 2016, 2017; Krishnaswami et al., 2016; Lacar et al., 2016; Lake et al., 2016). To prepare single nuclei, cells are lysed with detergent and dounce homogenized to expel cytoplasmic contents and nuclei from the cellular membrane, (Habib et al., 2016), which may avoid transcriptomic changes (Van Den Brink et al., 2017). Nuclei can then be purified by flow cytometry or gradient centrifugation (Grindberg et al., 2013; Ambati et al., 2016; Habib et al., 2016). When cell-type specific nuclear proteins exist, they can be used for nuclei isolation from specific cell types using antibody labeling (Lacar et al., 2016; Habib et al., 2017).

Single-nucleus RNAseq is not only amenable for difficult to isolate cell types, but can also be used for archived tissues such as flash-frozen clinical samples. Individual nuclei isolated from frozen adult mouse and human brain tissues have been successfully sequenced, demonstrating that snRNAseq has sufficient resolution to identify many different cell types from frozen and post-mortem tissue (Grindberg et al., 2013). With the rapid development of many applications for snRNAseq, nuclei are amenable to other studies not easily done by scRNAseq.

An important question remains: To what degree is the nuclear transcriptome representative of the whole cell? Recent studies have demonstrated that many transcripts of cell

and nucleus are equally represented and that nuclear RNA represents an important and significant population of transcripts that contribute greatly to the overall diversity of transcripts (Barthelson et al., 2007; Trask et al., 2009). Comparative studies of scRNAseq and snRNAseq in neural progenitor cells have also demonstrated that genes are expressed in equal proportion between whole cell and nuclei (Grindberg et al., 2013). Nanogrid single-cell and nuclei RNA sequencing studies in the same breast cancer lines found that overall copy number, expression level, and abundance had a high ($r_s = 0.95$) Spearman's correlation (Gao et al., 2017). Similarly, the transcriptomes of single cells and nuclei of 3T3 cells have also demonstrated high correlation (Pearson, $r = 0.87$) (Habib et al., 2017). Together these results suggest that nuclei and cells have highly correlated relative gene expression.

Despite the similarities between single-cell and nuclei transcriptomic profiles there remain notable differences. Not surprisingly, nuclear transcriptomes are enriched for several types of nuclear RNAs (Grindberg et al., 2013; Habib et al., 2016, 2017; Krishnaswami et al., 2016; Gao et al., 2017). Since ncRNAs are only polyadenylated in the nucleus, snRNAseq provides a feasible strategy to capture the heterogeneity of ncRNA transcription in single-cell resolution (Krishnaswami et al., 2016). In addition, nuclear transcriptomes are enriched for lncRNAs and nuclear-function genes (Gao et al., 2017). Another difference between cell and nuclear RNAseq is the higher abundance of intronic sequences in snRNAseq, which ranged between 10–40% of mapped reads (Grindberg et al., 2013; Gao et al., 2017; Habib et al., 2017). These features need to be accounted for when comparing datasets from cellular versus nuclear transcriptome analyses.
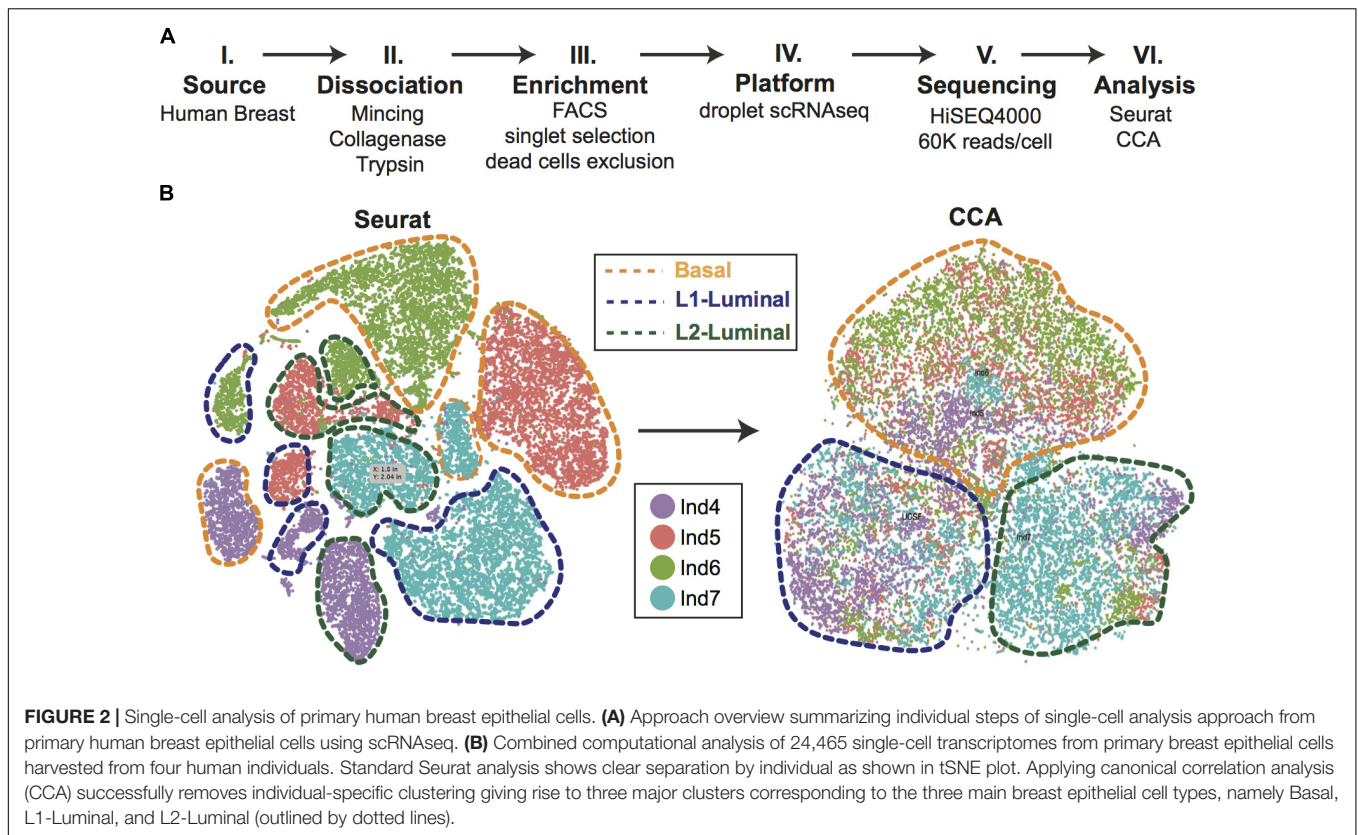
In conclusion, snRNAseq has emerged as a promising avenue for profiling archived samples or cell types that are hard to viably isolate from tissues.

SINGLE-CELL LIBRARY SEQUENCING

The next critical part of designing single-cell workflows is to align the analysis pipeline with the respective NGS platform and sequencing depth. It is important to confirm that the chemistry used for library construction is compatible with the sequencing technology. Currently, there are two major outputs for libraries from scRNAseq: full-length transcript or 3′-end counted libraries, which each require different read depths (Haque et al., 2017). Full-length transcript libraries are typically sequenced at a depth of 10^6 reads per cell, but may still yield important biological information at as low as 5×10^4 reads per cell (Pollen et al., 2014). For specific applications such as alternative splicing analysis on the single-cell level, much higher sequencing depth up to $15–25 \times 10^6$ reads per cell is necessary. On the other hand, 3′-end counting libraries are sequenced at much lower depth of around 10^4 or 10^5 reads per cells (Haque et al., 2017). Reaching the optimal sequencing depth can be an iterative process and may require multiple rounds of optimization. Sequencing saturation can be estimated by plotting down-sampled sequencing depth in mean reads per cell (e.g., 10× Genomics Cell Ranger).

[1]http://scotty.genetics.utah.edu/

[2]http://satijalab.org/howmanycells

FIGURE 2 | Single-cell analysis of primary human breast epithelial cells. (A) Approach overview summarizing individual steps of single-cell analysis approach from primary human breast epithelial cells using scRNAseq. (B) Combined computational analysis of 24,465 single-cell transcriptomes from primary breast epithelial cells harvested from four human individuals. Standard Seurat analysis shows clear separation by individual as shown in tSNE plot. Applying canonical correlation analysis (CCA) successfully removes individual-specific clustering giving rise to three major clusters corresponding to the three main breast epithelial cell types, namely Basal, L1-Luminal, and L2-Luminal (outlined by dotted lines).

STUDY DESIGN AND DATA ANALYSIS

In the following section, we highlight several key considerations from a data analysis perspective for adequately designing a successful scRNAseq study. As mentioned, many single-cell technologies can be greatly affected by technical variation, and without proper study design the results can be difficult to interpret. One critical aspect of this is the separation of *batch* and *condition*. *Batch* refers to a library that was singularly generated in a contained workflow (i.e., harvesting tissue specimen, disassociating into single-cell suspension, and generating scRNAseq library). *Condition* refers to a biological state or experimental treatment that is being analyzed in the study. Technical variation can be difficult to separate from relevant biological variation when conditions are interrogated individually. To help correct for this, the generation of replicates (biological or technical) whenever possible is strongly recommended.

In addition to replicates, an option is to mix samples and conditions within a batch, such that they can be treated without confounding each other (Hicks et al., 2015). One example is the Demuxlet workflow, where samples from genetically distinct individuals can be processed within the same library generation protocol and sequenced together (Kang et al., 2018). Prior to library generation, genotyping of distinct samples is performed and subsequently used in conjunction with the scRNAseq library to demultiplex the mixed cell sample into the samples of origin. In situations where genetically identical samples are used, or

genotypic data is not readily available, cellular hashing can be employed (Stoeckius et al., 2017). This involves oligo-tagged antibodies specific to each sample in the study and then pooling and generating the scRNAseq library from the sample mixture. The antibodies labeled with unique barcodes can be traced back to its sample of origin (Stoeckius et al., 2017).

Efforts can be made computationally to mitigate batch-to-batch variation. Batch effects are not unique to scRNAseq data, but the assumptions made by correction algorithms are not always appropriate for the bimodality of gene expression in zero-inflated scRNAseq data. Here, we highlight recent analytical frameworks that may be used to correct for this phenomenon. A recently developed approach by Haghverdi et al. (2018) builds a mixed nearest neighbor model for cells between datasets or samples that does not require known or equal proportions of cell types between data sets. In addition, the widely used Seurat pipeline for scRNAseq analysis recently employed canonical correlation analysis (CCA) that allows for discovery of co-correlated gene modules between datasets that can then be used to cluster upon (Butler et al., 2018). This approach identifies the cell types common between datasets and samples, as well as those that are unique to an individual set by finding common sources of variation in gene expression. As an illustration of this method, we applied CCA to our recently published droplet-enabled scRNAseq dataset from four individual primary human breast tissue samples (**Figure 2**). Finally, the single-cell batch correction framework MAST (Finak et al., 2015) models the positive expression mean and the over-the-background

expression of transcripts, and calculates a fraction of detected genes per cell and uses this as a covariate that is independent of a previously specified control set of genes. Together, these methods serve as recent examples to handle batch-to-batch variation computationally, resulting in improved dimensionality reduction and clustering for meaningful scRNAseq data analysis.

Beyond accounting for technical variation, a common question that researchers address is the relatedness of described cell populations through the lens of a differentiation processes. The key assumption of pipelines that seek to address this is that the tissue sample analyzed using scRNAseq contains cell types/states that represent not only the ends of a differentiation process, but also stem/progenitor cells and transitional cell states along the path of differentiation. Common analysis suites that seek to reconstruct these differentiation trajectories are Monocle (Qiu et al., 2017), TSCAN (Ji and Ji, 2016), and CellTree (duVerle et al., 2016). Each use different methods, but their goal is to visualize differentiation trajectories and identify expression signatures that change through pseudotime.

CONCLUSION

To fully harness the potential of single-cell analysis tools to decipher complex biological systems on the level of individual cells, careful study design and rigorous optimization of every step along the experimental procedure are mandatory. Here, we delineate a step-wise experimental approach for optimizing tissue handling, cell dissociation and enrichment, single-cell platform selection, library sequencing, and data analysis for designing single-cell workflows. A move toward standardized and automated processing of tissues will minimize changes introduced by tissue handling that may obscure biologically relevant transcriptomic profiles. For tissues that are problematic to dissociate into high-quality and viable single-cell suspensions, snRNAseq offers a solution to this problem, and can be used to achieve uniform extraction and sequencing of multiple cell types for cross comparison. Numerous computational frameworks are currently emerging that help mitigate batch effects to separate biological variation from unwanted technical variation.

AUTHOR CONTRIBUTIONS

KK outlined concept and overview of review. QN, NP, and KN wrote the manuscript. KK and QN designed and prepared the figures.

FUNDING

REFERENCES

Alles, J., Karaiskos, N., Praktiknjo, S. D., Grosswendt, S., Wahle, P., Ruffault, P. L., et al. (2017). Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* 15:44. doi: 10.1186/s12915-017-0383-5

Ambati, S., Yu, P., McKinney, E. C., Kandasamy, M. K., Hartzell, D., Baile, C. A., et al. (2016). Adipocyte nuclei captured from VAT and SAT. *BMC Obes.* 3:35. doi: 10.1186/S40608-016-0112-6

Baldan, V., Griffiths, R., Hawkins, R. E., and Gilham, D. E. (2015). Efficient and reproducible generation of tumour-infiltrating lymphocytes for renal cell carcinoma. *Br. J. Cancer* 112, 1510–1518. doi: 10.1038/bjc.2015.96

Barthelson, R. A., Lambert, G. M., Vanier, C., Lynch, R. M., and Galbraith, D. W. (2007). Comparison of the contributions of the nuclear and cytoplasmic compartments to global gene expression in human cells. *BMC Genomics* 8:340. doi: 10.1186/1471-2164-8-340

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., et al. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–490. doi: 10.1038/nature14590

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420. doi: 10.1038/nbt.4096

Christaki, E., Opal, S. M., Keith, J. C., Kessimian, N., Palardy, J. E., Parejo, N. A., et al. (2011). A monoclonal antibody against RAGE alters gene expression and is protective in experimental models of sepsis and pneumococcal pneumonia. *Shock* 35, 492–498. doi: 10.1097/SHK.0b013e31820b2e1c

duVerle, D. A., Yotsukura, S., Nomura, S., Aburatani, H., and Tsuda, K. (2016). CellTree: an R/bioconductor package to infer the hierarchical structure of cell populations from single-cell RNA-seq data. *BMC Bioinformatics* 17:363. doi: 10.1186/s12859-016-1175-6

Fan, H. C., Fu, G. K., and Fodor, S. P. A. (2015). Combinatorial labeling of single cells for gene expression cytometry. *Science* doi: 10.1126/science.1258367 [Epub ahead of print].

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* 16:278. doi: 10.1186/s13059-015-0844-5

Gao, R., Kim, C., Sei, E., Foukakis, T., Crosetto, N., Chan, L. K., et al. (2017). Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in breast cancer. *Nat. Commun.* 8:228. doi: 10.1038/s41467-017-00244-w

Grindberg, R. V., Yee-Greenbaum, J. L., McConnell, M. J., Novotny, M., and O'Shaughnessy, A. L. (2013). RNA-sequencing from single nuclei. *Proc. Natl. Acad. Sci. U.S.A.* 110, 19802–19807. doi: 10.1073/pnas.1319700110

Habib, N., Avraham-davidi, I., Basu, A., and Burks, T. (2017). Massively-parallel single nucleus RNA-seq with DroNc-seq. *Nat. Methods* 14, 955–958. doi: 10.1038/nmeth.4407

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

Haghverdi, L., Lun, A. T. L., Morgan, M. D., and Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* 36, 421–427. doi: 10.1038/nbt.4091

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., et al. (2018). Mapping the mouse cell atlas by microwell-Seq. *Cell* 172, 1091.e17–1097.e17. doi: 10.1016/j.cell.2018.02.001

Haque, A., Engel, J., Teichmann, S. A., and Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* 9, 1–12. doi: 10.1186/s13073-017-0467-4

Heath, J. R., Ribas, A., and Mischel, P. S. (2016). Single-cell analysis tools for drug discovery and development. *Nat. Rev. Drug Discov.* 15, 204–216. doi: 10.1038/nrd.2015.16

Hicks, S. C., Teng, M., and Irizarry, R. A. (2015). On the widespread and critical impact of systematic bias and batch effects in single-cell RNA-Seq data. *bioRxiv* [Preprint]. doi: 10.1101/025528

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., et al. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11, 163–166. doi: 10.1038/nmeth.2772

Ji, Z., and Ji, H. (2016). TSCAN: pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* 44:e117. doi: 10.1093/nar/gkw430

Kang, H. M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* 36, 89–94. doi: 10.1038/nbt.4042

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

Kornbluth, J., and Hoover, R. G. (1989). "Anti-HLA class i antibodies alter gene expression in human natural killer cells," in *Immunobiology of HLA*, ed. B. Dupont (New York, NY: Springer), 150–152.

Krishnaswami, S. R., Grindberg, R. V., Novotny, M., Venepally, P., Lacar, B., Bhutani, K., et al. (2016). Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* 11, 499–524. doi: 10.1038/nprot.2016.015

Lacar, B., Linker, S. B., Jaeger, B. N., Krishnaswami, S. R., Barron, J. J., Kelder, M. J. E., et al. (2016). Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat. Commun.* 7:11022. doi: 10.1038/ncomms11022

Lake, B., Shen, R., Ronaghi, M., Fan, J., Wang, W., and Zhang, K. (2016). Neuronal subtypes and diverstiy revealed by single-nucleus RNA sequencing of human brain. *Science* 352, 1586–1590. doi: 10.1126/science.aaf1204

Lawson, D. A., Bhakta, N. R., Kessenbrock, K., Prummel, K. D., Yu, Y., Takai, K., et al. (2015). Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* 526, 131–135. doi: 10.1038/nature15260

Lescroart, F., Wang, X., Lin, X., Swedlund, B., Gargouri, S., Sànchez-Dànes, A., et al. (2018). Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science* 359, 1177–1181. doi: 10.1126/science.aao4174

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

Meeson, A., Fuller, A., Breault, D. T., Owens, W. A., and Richardson, G. D. (2013). Optimised protocols for the identification of the murine cardiac side population. *Stem Cell Rev. Rep.* 9, 731–739. doi: 10.1007/s12015-013-9440-9

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., et al. (2011). Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90–95. doi: 10.1038/nature09807

Nguyen, Q. H., Pervolarakis, N., Blake, K., Ma, D., Davis, R., James, N., et al. (2018). Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat. Commun.* 9:2028. doi: 10.1038/s41467-018-04334-1

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 8–13. doi: 10.1126/science.1254257

Picelli, S., Faridani, O. R., Björklund, ÅK., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using smart-seq2. *Nat. Protoc.* 9, 171–181. doi: 10.1038/nprot.2014.006

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

Qiu, X., De Jesus, J., Pennell, M., Troiani, M., and Haun, J. B. (2014). Microfluidic device for mechanical dissociation of cancer cell aggregates into single cells. *Lab Chip* 15, 339–350. doi: 10.1039/c4lc01126k

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., et al. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982. doi: 10.1038/nmeth.4402

Radbruch, A., and Recktenwald, D. (1995). Detection and isolation of rare cells. *Curr. Opin. Immunol.* 7, 270–273. doi: 10.1016/0952-7915(95)80014-X

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

Romero-santacreu, L., Moreno, J., Perez-Ortin, J. E., and Alepuz, P. (2009). Specific and global regulation of mRNA stability during osmotic stress in *Saccharomyces cerevisiae*. *RNA* 15, 1110–1120. doi: 10.1261/rna.1435709

Rozenblatt-Rosen, O., Stubbington, M. J. T., Regev, A., and Teichmann, S. A. (2017). The human cell atlas: from vision to reality. *Nature* 550, 451–453. doi: 10.1038/550451a

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., et al. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* 7:3. doi: 10.1186/1471-2199-7-3

Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B., Smibert, P., et al. (2017). Cell "hashing" with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *bioRxiv* [Preprint]. doi: 10.1101/237693

Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338. doi: 10.1038/nature21350

Tirosh, I., Izar, B., Prakadan, S. M., Ii, M. H. W., Treacy, D., Trombetta, J. J., et al. (2016). Dissecting the multicellular exosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196. doi: 10.1126/science.aad0501. Dissecting

Trask, H. W., Cowper-Sal-lari, R., Sartor, M. A., Gui, J., Heath, C. V., Renuka, J., et al. (2009). Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs. *RNA* 15, 1917–1928. doi: 10.1261/rna.1677409

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

Tung, P. Y., Blischak, J. D., Hsiao, C. J., Knowles, D. A., Burnett, J. E., Pritchard, J. K., et al. (2017). Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* 7:39921. doi: 10.1038/srep39921

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

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

Will, B., and Steidl, U. (2010). Multi-parameter fluorescence-activated cell sorting and analysis of stem and progenitor cells in myeloid malignancies. *Best Pract. Res. Clin. Haematol.* 23, 391–401. doi: 10.1016/j.beha.2010.06.006

Xiong, L., Lee, H., Ishitani, M., and Zhu, J. K. (2002). Regulation of osmotic stress-responsive gene expression by the LOS6/ABA1 locus in *Arabidopsis*. *J. Biol. Chem.* 277, 8588–8596. doi: 10.1074/jbc.M109275200

Yuan, G. C., Cai, L., Elowitz, M., Enver, T., Fan, G., Guo, G., et al. (2017). Challenges and emerging directions in single-cell analysis. *Genome Biol.* 18:84. doi: 10.1186/s13059-017-1218-y

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