

REVIEW

The use of single-cell RNA-seq to study heterogeneity at varying levels of virus–host interactions

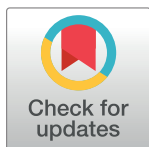
Sharmada Swaminath, Alistair B. Russell¹*

School of Biological Sciences, University of California, San Diego, La Jolla, California, United States of America

* a5russell@ucsd.edu

Abstract

The outcome of viral infection depends on the diversity of the infecting viral population and the heterogeneity of the cell population that is infected. Until almost a decade ago, the study of these dynamic processes during viral infection was challenging and limited to certain targeted measurements. Presently, with the use of single-cell sequencing technology, the complex interface defined by the interactions of cells with infecting virus can now be studied across the breadth of the transcriptome in thousands of individual cells simultaneously. In this review, we will describe the use of single-cell RNA sequencing (scRNA-seq) to study the heterogeneity of viral infections, ranging from individual virions to the immune response between infected individuals. In addition, we highlight certain key experimental limitations and methodological decisions that are critical to analyzing scRNA-seq data at each scale.



OPEN ACCESS

Citation: Swaminath S, Russell AB (2024) The use of single-cell RNA-seq to study heterogeneity at varying levels of virus–host interactions. *PLoS Pathog* 20(1): e1011898. <https://doi.org/10.1371/journal.ppat.1011898>

Editor: Daniel Blanco-Melo, Fred Hutchinson Cancer Center, UNITED STATES

Published: January 18, 2024

Copyright: © 2024 Swaminath, Russell. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the NIGMS of the NIH under grant R35GM147031 awarded to ABR. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Viral disease is fundamentally driven by interactions of populations of viruses, often comprising a diverse set of genotypes, with populations of host cells, themselves comprised of a variety of cell types and epigenetic states. Infection outcome is dependent upon these complex, heterogeneous interactions. Until relatively recently, it was difficult, labor-intensive, and limited in terms of available experimental measurements to study infection at any level other than bulk measurements averaging effects across many different infected (and uninfected) cells. With the advent of single-cell RNA sequencing (scRNA-seq), it has now become much easier to explore complexities within viral infection.

In this review, we will discuss examples of how scRNA-seq has been used to characterize heterogeneity and stochasticity in viral infection, ranging from individual virions to even understanding differences in the immune response between infected individuals ([Fig 1](#)). While a powerful new approach, as with any tool, there are important experimental limitations and methodological adjustments that are critical to understanding and interpreting scRNA-seq experiments. As we describe each level at which scRNA-seq has been applied to understanding viral infection, we also provide some considerations in terms of known limitations to the technology.

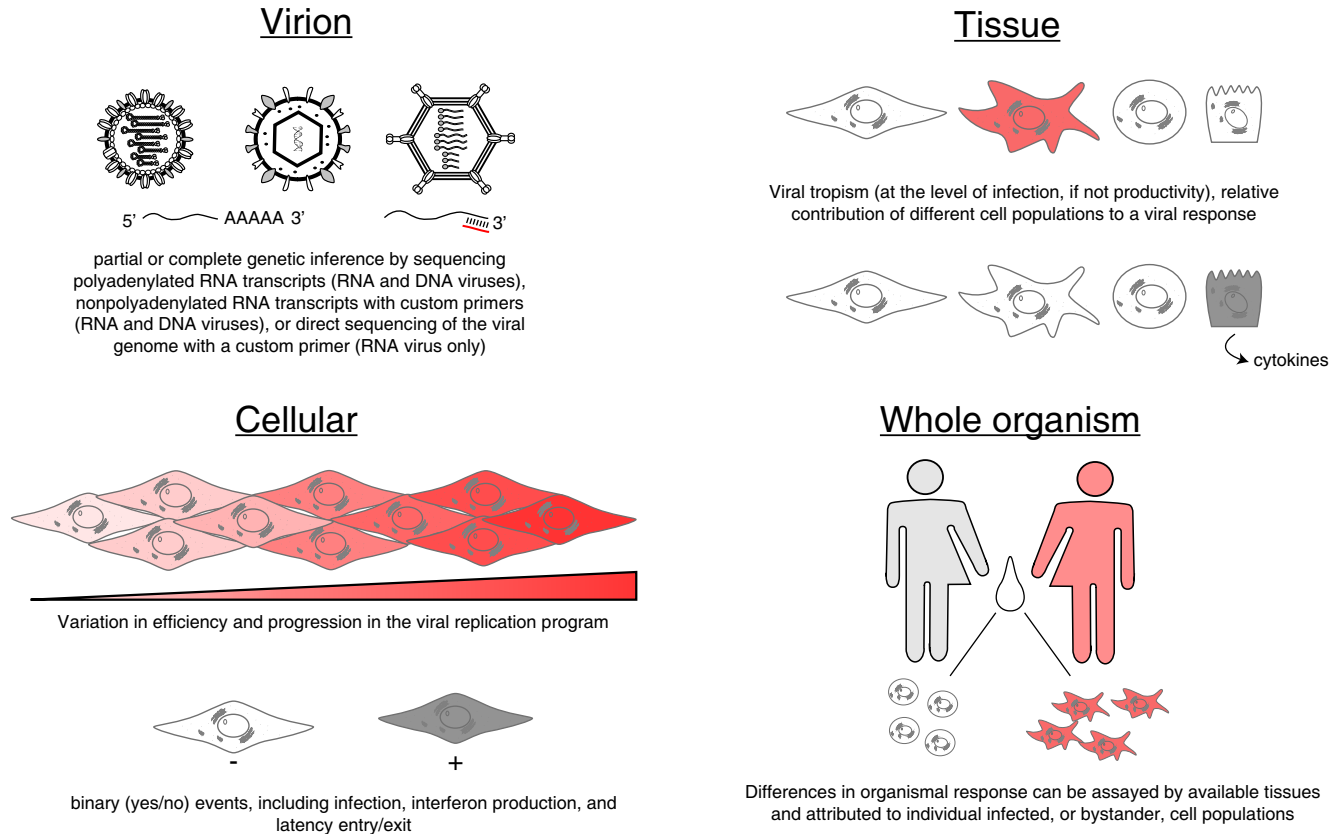


Fig 1. Broad examples of heterogeneity probed by scRNA-seq in the context of viral infection. Specific instances of each provided in the text of the appropriate section.

<https://doi.org/10.1371/journal.ppat.1011898.g001>

Single-cell RNA sequencing methods

scRNA-seq has been a powerful and widely used technique to study biological heterogeneity since its development in 2009 [1]. Since then, scRNA-seq methods have advanced in sensitivity, throughput, and cost, leading to increased adoption in research and clinical settings. Existing scRNA-seq platforms are generally based on the same fundamental steps: (1) cells isolated as a single-cell suspension; (2) mRNA is captured, either in situ or from lysed cells; (3) cDNA is generated from captured mRNA and barcoded with cellular identity; (4) cDNA is prepared for sequencing; and (5) sequencing is performed (Table 1).

scRNA-seq experiments were made possible in part due to key innovations in both cDNA synthesis and sequencing library preparation: the highly efficient Smart-Seq method, which uses template switching to amplify full-length cDNA from picogram amounts of RNA, and tagmentation using Tn5 transposase, which permits the fragmentation of much smaller amounts of material than classical shearing methods [2–10]. For the former, Smart-Seq uses 2 properties of the reverse-transcriptase derived from Moloney murine leukemia virus, the untemplated addition of cytosines at the 5' end of mRNA transcripts, and the ability to template switch, generating cDNA that is a chimera of 2 different molecules [11]. By adding a template switching oligonucleotide, an oligonucleotide containing riboguanosine or a locked-nucleic acid form of the same at the 3' end, this oligonucleotide can hybridize to the untemplated cytosines and be transcribed as part of the cDNA [12]. Broadly, this is a highly efficient means to add molecular handles for downstream PCR amplification of full-length cDNA.

Table 1. Single-cell RNA sequencing methods discussed in this review.

Methods	Examples	Description	References
Individual, sorted cells	Smart-Seq followed by tagmentation	Individual cells isolated in microliter droplets in a standard 96- or 384-well plate, sequencing libraries generated from each independent cell. Low throughput.	Ramsköld et al., 2012 [6]
Well-based	Seq-Well	Cells captured in individual, subnanoliter wells with barcoded beads. Cellular barcoding of transcripts occurs during reverse transcription, material pooled for library preparation. Does not require specialized equipment beyond standard molecular biology instruments.	Gierahn et al., 2017 [16]
Droplet-based	InDrop, Drop-Seq, 10x Genomics Chromium	Cells co-encapsidated in droplets alongside barcoded beads in a microfluidic device. In-droplet reverse transcription provides cellular barcodes, after which droplets are combined and pooled for library preparation. Most used technology in the studies discussed in this review.	Macosko et al., 2015; Klein et al., 2015; Zhang et al. 2019 [20,21,116]
Split-Pool	SPLiT-Seq	Cells fixed, distributed between standard wells in a 96- or 384-well plate, and an initial barcode added to permeabilized cells via in situ reverse transcription. Subsequent ligation reactions after repooling and resplitting cells produces a combinatorial cell barcode, after which standard library preparation follows. Does not require specialized equipment beyond standard molecular biology instruments.	Cao et al., 2017; Rosenberg et al., 2018 [22,23]

<https://doi.org/10.1371/journal.ppat.1011898.t001>

While an incredible advancement, this early effort was limited in terms of cost and throughput, as each cell was processed in an individual well of a 96- or 384-well plate through library preparation. However, we should add that, owing to the indexing of each cell performed at the tagmentation step, this method has the advantage of capturing all portions of all transcripts without the use of long-read sequencing.

One of the first efforts to parallelize and reduce the cost of this process was the use of microfluidic chips such as the Fluidigm C1 platform [13,14]. This instrument uses a microfluidic chip to capture, lyse, and perform library preparation in volumes much smaller than achievable using pipetting. However, this technology has largely been supplanted by methods that instead use droplet- or well-based platforms that accomplish even higher throughput at lower cost. A key innovation in both methodologies is the use of beads coupled to barcoded oligonucleotides, which, using a split-pool synthesis strategy, allows for every oligonucleotide on a bead to contain the same barcode, but oligonucleotides from different beads to possess different barcodes [15]. This introduces a cellular index during reverse transcription, allowing for downstream library preparation steps to be performed with cDNA from all individual cells in a single pool. By combining processing steps after initial barcoding, higher throughput, lower cost, and reduced technical noise have all been achieved relative to older technologies. However, a key limitation is that, without modification to use long-read technology, sequencing is generally limited to either 5' or 3' end of each transcript.

Well-based scRNA-seq methods such as Seq-Well use arrays of subnanoliter wells and a gravity-based capture of both beads and individual cells [16]. Once loaded in wells, a semipermeable membrane is applied, cells are lysed, and mRNA captured by oligo(dT) beads containing cell-specific barcodes. Beads are then recovered followed by reverse transcription of the bead-bound mRNA using the Smart-seq method and sequencing libraries with cell-specific identifiers are generated. While currently not as broadly used as droplet-based methods, discussed below, Seq-Well is relatively gentle on cells and may be more advantageous in difficult-to-handle cell types such as neutrophils [17,18]. In addition, similar to a split-pool method introduced below, this method does not require equipment beyond an array of subnanoliter wells and standard molecular biology instruments, making it easier to deploy in virological studies without the dedication of expensive equipment to a biocontainment facility.

Drop-seq, InDrop, and the commercially available Chromium platform from 10x Genomics, collectively, are droplet-based methods of scRNA-seq (Fig 2) [19–21]. Microfluidic chips

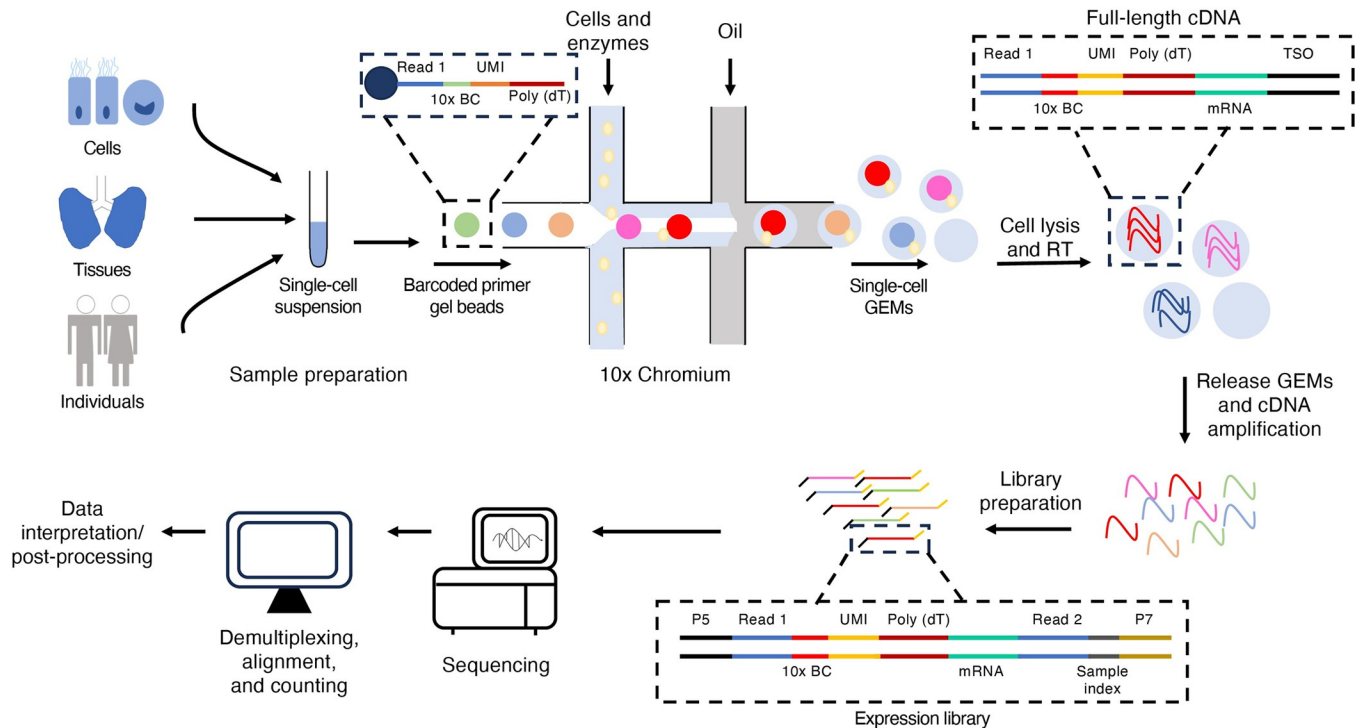


Fig 2. Single-cell sequencing workflow for droplet-based approaches exemplified by the 10x Genomics Chromium platform. Infected cells from cell culture, tissues/organoids, or infected individuals are dissociated into a single-cell suspension. The cell suspension is loaded onto a microfluidic chip, and cells are partitioned into nanoliter-scale Gel Beads-in-emulsion (GEMs) droplets containing barcoded gel beads and reagents for reverse transcription (RT). Following cell lysis, the beads capture the mRNA molecules. Reverse transcription (RT) by template switching using a template switching oligonucleotide (TS) generates cDNA tagged with a 10x barcode (BC) to identify the cell and a unique molecular identifier (UMI) to label the mRNA transcript. The pooled cDNA is amplified in bulk, fragmented by enzymatic fragmentation, and sequencing adapters (P5 and P7) including a sample index, are added to the fragments by PCR to generate sequencing libraries. The sequencing libraries are sequenced, and the data are analyzed by alignment and demultiplexing, following which the data are interpreted.

<https://doi.org/10.1371/journal.ppat.1011898.g002>

are used to generate picoliter droplets containing cells (loaded probabilistically and assumed to follow a Poisson function of occupancy) and barcoded beads. These droplets contain cell lysis and reverse transcription reagents. While maintaining droplet integrity, cells are lysed and reverse transcribed to generate cDNA in a standard thermocycler. As the cDNAs are already tagged with a cell-specific barcode, emulsions are broken and sequencing libraries are prepared by standard methods. Cell-specific barcodes added during reverse transcription are from oligonucleotides coupled to beads, either as an oligo(dT) and marking the 3' end of transcripts, or in the template-switching oligonucleotide, labeling the 5' end of transcripts.

Using a similar split-pool philosophy as barcoded bead generation to identify individual cells, but using the cell itself as a reaction vessel, is split-pool sequencing, such as SPLiT-seq or SCI RNA-seq [22–24]. Rather than lysing the cell, cells are first fixed and permeabilized. Then, cells are distributed between individual wells of a 96-well plate, with multiple cells per well. A barcode can be added at this point by reverse transcription, with each cell within a well sharing a barcode, but distinct barcodes between wells. Following within-cell reverse transcription, cells are pulled from wells, mixed, and redistributed again between wells. An additional well-specific barcode is added at this point by ligation. By performing multiple additional rounds of barcoding after “reshuffling” the cells, each cell is tagged with a unique combination of barcodes, and, thus, cDNA are reassigned to an individual cell after sequencing. After barcoding, cells are lysed, and standard library preparation followed. Split-pool sequencing has the potential to be highly scalable, with theoretical barcode diversity increasing exponentially with the

addition of each barcoding step. Additionally, the use of fixed cells and, as with well-based approaches, the lack of specialized equipment, makes this approach potentially more appealing when working with infectious agents.

While we generally make no distinctions in the remainder of the review on which technology was used to generate observations in terms of viral infection and heterogeneity, we do note that by far the most frequent technology used has been droplet-based sequencing (described in Fig 2). We therefore largely restrict our discussions of experimental considerations and caveats to that technology.

Virion-level heterogeneity

It has been appreciated for quite some time that viral populations are heterogeneous [25]. Some of the earliest observations in virology were that the number of physical particles in viral preparations greatly exceeds that of infection-competent particles alone. While some of these particles are merely empty capsids, not relevant to the technologies we describe here, many bear mutations or other genetic defects that render them incapable of completing the viral life cycle. Curiously, despite outnumbering their replication-competent counterparts, the consensus sequence across a viral population is generally fit (safe for populations enriched for defective viral genomes, discussed briefly below), indicating a myriad of different genetic defects underlie the failure to replicate rather than a single stereotyped defect that rises to consensus.

These unfit particles have represented a sort of “dark matter” in experiments [26–29]. They may still enter cells and can influence features such as the induction of an innate immune response but are difficult to assess using bulk methodologies that largely focus on replication-competent components, particularly those that rise to consensus, of a viral population. As viral genomes are highly compact, or themselves can serve as the messages for protein production, sequencing of mRNA pools in a cell can provide insight into the totality of viral genotypes, replication competent and otherwise. This assumes the defect itself does not prevent transcription—or that the genome itself is an RNA molecule—however, this limitation still provides access to biological features that were generally difficult to access by older methodologies.

The capacity to explore individual virion genetic variation using modern scRNA-seq, replication-competent or otherwise, has been perhaps most extensively (but not exclusively) realized in the study of influenza A virus [30–36]. As this virus possess a relatively small genome of approximately 13.5 kb, split across 8 RNA molecules, the longest of which is still relatively short at 2.3 kb in length, and nearly all of which is transcribed into capped, polyadenylated mRNA, very few modifications are required to use preexisting single-cell approaches to access viral genetic heterogeneity in a standard scRNA-seq workflow [37]. Indeed, nonsegmented RNA viruses would present a considerable challenge for genetic inference as it is difficult to achieve full-length cDNA synthesis from long, structured RNA molecules, and DNA viruses would present their own challenge as regions that are not transcribed cannot be measured by these approaches. While only 2 manuscripts thus far have sequenced the entire 5′ to 3′ ends of viral transcript in infected cells, several have used partial genetic information to inform studies of viral variation, including features such as reassortment [38].

The most commonly used workflow in scRNA-seq is to sequence either the 5′ or 3′ ends of transcripts. This provides the ability to recover both the cell-identifying barcode (in either the template-switching oligonucleotide or the oligo(dT) primer) as well as sufficient sequence information to assign transcripts. The key limitation requiring fragmentation and end-choice is due to both the read-length limitation of Illumina sequencing, as well as length requirements for efficient bridge amplification in an Illumina flow cell [39]. From this partial genetic information, several studies nevertheless were able to identify architectural changes to viral

genomes, for instance, large internal deletions in influenza A virus [31,33,34]. These deletions form the basis of defective viral genomes (DVGs, also called nonstandard viral genomes or nsVGs), which are associated with a general loss of viral fitness as well as the induction of an innate immune response [40,41]. Critically, these studies not only identified these species but also were able to assay them with respect to a heterogeneous host response, indicating that while they are associated with induction of an innate immune response, they are neither necessary nor sufficient to induce interferon but instead increase the probability of its expression nondeterministically.

Experimental considerations for the study of virions; read-length limitations and alternate priming strategies

Several considerations should be taken into account if using single-cell transcriptomics to assay viral genetic heterogeneity alongside other features of interest. The first is the limitation of the standard workflow to 5' or 3' end sequencing. In addition to the obvious—that variant capture will largely be limited to sequences in these regions—the “width” of the coverage window is largely dependent on sequencing depth. For viral reads, this will, in turn, depend on the level of viral transcripts in an infected cell. Thus, successful genotyping, including architectural variation, is impacted by the level of viral transcripts [34]. This can lead to certain incorrect conclusions; for instance, if the probability that a large deletion in the viral genome is detected is proportional to the level of coverage, then that deletion will be found to be correlated with high levels of viral transcription regardless of its actual biological impact.

Solving the issue of partial coverage, during the preparation of sequencing libraries, full-length, tagged cDNA is generated prior to fragmentation. This material can be sequenced by long-read technologies, such as PacBio SMRT or nanopore sequencing. This methodology has been termed “ISOseq” (ScISOr-Seq in single cells) and has been used to annotate isoform differences in both bulk and single-cell sequencing [42,43]. A newer method, named HIT-ScISOr-Seq, further uses concatemerization to increase the throughput of long-read sequencing to more comprehensively measure full-length mRNA across an entire scRNA-seq experiment [44,45]. Use of a targeted ScISOr-Seq workflow to call viral genotypes has been used in 2 different manuscripts for influenza A virus, and similar approaches could be extended to other viral populations [35,36]. One key artifact, though, that should be considered is that some genotypes will be erroneously assigned to certain cells owing to template switching during reverse transcription or chimera formation during PCR [46]. One possible solution is to use populations with known variants that differ at the 5' and 3' end of transcripts, thus allowing for an empirical measurement of these artifacts [35].

Additionally, not all viruses generate capped, polyadenylated messages. This is particularly true of RNA viruses outside of the Orthomyxoviruses. As current scRNA-seq protocols almost exclusively use oligo(dT) priming, these messages would be missed. One potential solution is to use a splinted oligonucleotide, appending a unique sequence to capture virus-specific transcripts. This was used by Saikia and colleagues to sequence Reovirus genomes, a dsRNA virus that does not generate polyadenylated transcripts [47]. Regarding cap structures, while a wealth of literature suggests that the template-switching activity is limited to canonical 7-methylguanosine cap structures, more mechanistic studies have demonstrated that while this behavior is more efficient on 5' methylated caps, it can still work on alternative 5' ends, extending full-length cDNA synthesis and amplification out to alternate RNA structures that may be present in viral transcripts or genomes [48].

Lastly, the standard workflow used for most scRNA-seq approaches uses readmappers that, while appropriate for mapping spliced host transcripts with little variation between reference

and reads, are inappropriate for mapping viral diversity. For instance, the most commonly used readmapper for scRNA-seq data, STAR, attempts to map disjoint (gapped) reads using a combination of preassigned splice junctions or scoring models that take into account general features of splice junctions [49]. When applied to architectural variation in viral populations, such as recombination or large deletions, these read mappers will inappropriately apply these scoring models. It is therefore critical to use specific, purpose-built pipelines, such as ViReMa or VODKA2, for mapping viral variation, and potentially regenerating viral consensus sequences when diversity significantly diverges from the chosen reference sequence [50,51].

Intercellular variation

Some of the first uses of scRNA-seq in virology were to assay heterogeneity between cells in tissue culture infections of either immortalized or patient-derived primary cells [30,32,35,52–54]. Despite the relative uniformity of these cells, there exists considerable variation in even one of the simplest models of infection. Examples in terms of viral measurements include the amount of viral transcript made, entry and exit into latency, and progression through complex viral life cycles. On the host response to viral infection, measurements can be made between bystander and infected cells, correlations of host response and viral abundance, and, lastly, engagement of the host innate immune response.

With respect to viral measurements, one of the most stark observations across multiple viral families is the extreme variation in the fraction of virally derived reads within a cell. Cells often vary in terms of viral transcript abundance by as much as several orders of magnitude, as observed in flavivirus, influenza, and coronavirus infections, to name a few [32,52,54]. Similar heterogeneity has also been observed in viral burst size, from early measurements in bacteriophage out to more recent measurements in influenza, although the relationship between transcript variance and burst size is nonlinear, in part due to the fact that many infections are nonproductive despite completing a large portion of the viral replication program [36,55,56]. So, while population-level bulk measurements can establish per cell productivity, as well as the average fraction of transcripts that are derived from viruses, a “boots-on-the-ground” view demonstrates these averages miss the high degree of heterogeneity in actual infections. This has implications in appropriate modeling of viral replication and, if burst size is variable even within a given genetic variant, has implications for effective viral population size and the evolutionary rates within individual infections.

Using this heterogeneity as a means to understand the host response, the massive breadth in viral transcription can then be correlated with host gene expression. After all, one would hardly expect a cell wherein >50% of the transcripts are derived from a viral infection to be equivalent to one where only one out of every thousand transcripts is viral. The most straightforward analysis is to look at genes that vary in expression in a manner correlated with viral expression. This type of analysis has been broadly useful for at least 2 reasons: that these genes themselves may be causal with respect to either pro- or antiviral strategies, and, independent of direct impacts on the virus, that they represent stressors on host cells imposed by viral infection. An example of the former, high expression of interferon-stimulated antiviral genes correlates with lower viral transcripts for HSV-1 and West Nile virus infections, indicating successful engagement of host defenses as well as demonstrating that the breadth of viral outcomes are driven, at least in part, by host restriction mechanisms [53,57,58]. For the latter, the massive metabolic burden placed on cells by viral infection can leave a distinct signature that is readily revealed by correlating such pressures with viral burden even within a single experiment. For instance, oxidative and ER stress pathways have been observed to correlate with viral infection across multiple phylogenetically distinct viral families, perhaps due to the

relatively high energetic and protein-folding demands placed on a cell by the massive increases of nucleic acid and protein biosynthesis required for viral replication [32,52,57,58].

Beyond these straightforward correlations, scRNA-seq has also allowed the study of highly stochastic, bimodal (yes/no) events in viral infection. On the viral side, a fantastic example of this is the study of latency. The earliest example of which we are aware of this use was in the study of HCMV infection in primary cells, wherein more precise correlations of latency, host gene expression, and viral programs were able to be established by the capacity to separate out subpopulations of cells to ensure that correlated features coexisted within the same infected cell—critical as while such features may correlate at the population level, that does not necessarily mean that such programs are occurring within the same cell [59].

On the host side, one of the best examples of a bimodal event is the engagement of an anti-viral response. The production of interferons in response to viral infection has been known to be highly stochastic for quite some time, but the features that lead to this stochasticity have been very difficult to study [60–62]. Using scRNA-seq, individual cells producing interferon can be identified [63,64]. This, in turn, can be correlated with viral infection, levels of viral transcripts, and, as discussed in the prior section, viral genetic variation to determine what features may lead to the induction of a response. Thus far, it has been illuminating that while there is a considerable contribution of viral variation to the induction of an innate immune response, even stimulatory variants within a population tend to induce a stochastic, if more common, production of interferons, indicating a role for chance events or preexisting host cell heterogeneity [34,65].

Lastly, in addition to profiling standing variation within a single tissue culture infection, single-cell transcriptomics permits the simultaneous profiling of many CRISPR modulations within a single experiment, allowing one to deliberately introduce variation and perform a highly parallel genetic analysis. One method that accomplishes this is Perturb-seq [66]. Perturb-seq has been used to identify host factors that influence viral replication and transcription, determining how particular perturbations may enhance, or repress, the viral replication program, as has been performed for Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2) [67]. Moreover, while likely restricted to only DNA viruses, Perturb-seq has also been used to directly profile CRISPR perturbations to a herpesvirus genome, human cytomegalovirus, reducing an otherwise laborious process of individually mutating and assessing the effect of each individual viral gene to a single one-pot experiment [68].

Experimental considerations for the study of intercellular variation; lysis and batch effects

A number of modifications to the standard single-cell workflow are even useful in interpreting these relatively simple models of infection. One problem that particularly plagues an understanding of viral infection models by scRNA-seq is the contribution of extracellular/misassigned RNA (also referred to as ambient RNA). As indicated in the prior section, there is considerable interest in bimodal events in viral infection, which can even include the probability that a cell itself is detectably infected. Droplet sequencing methodologies have significant amounts of reads contributed from the “soup,” a mixture of misassigned reads due to template switching and PCR chimeras, as well as extracellular RNA from cell lysis, membrane vesicles, and potentially even viral particles themselves. This may particularly pose a problem for cytopathic viruses, which may be overrepresented in the “soup.” This creates a challenge when identifying whether cellular responses are directly due to viral infection or due to signaling from infected cells in terms of the “bystander” population, as uninfected cells may nevertheless be observed to be associated with viral RNA.

While true negatives can never really be identified in scRNA-seq data, it is critical to set some kind of threshold for true positives. In other words, where is the limit of detection? This limit cannot be extrapolated across even individual sequencing runs from the same laboratory, as the degree of contamination is highly idiosyncratic and sample dependent. There exist some methods to interpolate these contamination levels in datasets, such as SoupX, although caution should be used as these methods rely on distinct transcriptional clusters of cells that may not reliably exist in tissue culture models of infection [69–72]. Another method is to actually measure the contamination rate, which can be achieved by adding in a cell type that possesses none of the transcripts of interest and measuring their rate of presence (for instance, by adding uninfected canine cells to infected human cells immediately prior to processing) [35].

Secondly, batch effects can also play a role when comparing across experimental treatments [73]. Regression analyses often find that replicate number is a significant source of signal in single cell analyses [74]. This is a trickier problem to overcome, and there are likely multiple nonoverlapping reasons for batch-to-batch variation, which may include, but is unlikely to be limited to, ambient RNA contamination. Inclusion of multiple replicates of different treatment conditions, and use of a regression analysis to identify genes that tend to vary based on replicate rather than experimental treatment, can partially address these issues with scRNA-seq [75–78]. Alternatively, as batch effects tend to be limited to a particular scRNA-seq run, cells under different experimental conditions can be labeled using “hashing” reagents, which are either an antibody coupled to a short oligonucleotide or a compound coupled to a short oligonucleotide that can embed in cell membranes, permitting multiplexing of experimental conditions in a single experimental run by marking cells from each condition in a way that is recoverable during sequencing [79].

Tissue-level heterogeneity

While simple, homogenous tissue culture models of infection can be informative, real infections are significantly more complex. The fact that infections occur in the presence of multiple divergent cell populations, which themselves communicate extensively with one another is missing. This type of complexity lends itself well to scRNA-seq, as prior studies of tissue-level phenomena were either limited by averaging over potentially wildly divergent individual cellular responses or, when cell-level heterogeneity could be resolved by microscopy or flow cytometry, limited to a relatively small number of targeted observations.

Unlike in homogenous tissue culture experiments, we would anticipate tissue infections to have significant variability in terms of not only what cells are present but also what cells are infected. This, for instance, can provide insight into how infections are successfully controlled by cell intrinsic innate immunity. For instance, when comparing infected and bystander cells during influenza and Ebola infections in *in vivo* animal models, bystander cells exhibit significantly higher expression of interferon-stimulated genes [33,80,81]. This could represent both viral suppression in infected cells, as well as an inability of these viruses to spread within tissues engaged in active interferon signaling. Similarly, changes in cell infectivity can also be assessed by scRNA-seq during antiviral drug treatment [82].

Comparing infected, and bystander, cells in tissues also can provide insights into viral tropism. While scRNA-seq measurements do not meet the classical definition of tropism, which is the productive infection of a cell type, they provide identification of cells that support successful viral entry and engagement of some portion of the viral transcriptional program. This can also be seen as an advantage, as both productively, and nonproductively, infected cells can contribute to overall pathology and disease progression. This method was used to support tropism of SARS-CoV2 in both human and animal tissues, including *ex vivo* human lung models,

not only confirming angiotensin converting enzyme 2 (ACE2) as the predominant receptor, and identifying cell-specific expression of this receptor, but also demonstrating ACE2-independent entry into certain cell populations [83–85]. Not limited to SARS-CoV2, cells susceptible to infection have been studied using scRNA-seq for astrovirus, rubella, and influenza, among others [80,86,87].

Experimental considerations for the study of tissue-level variation; isolation effects and biases

While generation of single-cell suspensions required for scRNA-seq is often straightforward in homogenous tissue culture infections, there is a considerable challenge to generating them from complex animal or human tissues. For instance, neurons are incredibly fragile and possess a large number of intercellular contacts and are incredibly difficult to isolate without their destruction. One solution for such tissues is to perform single-nuclei RNA-seq (snRNA-seq), rather than scRNA-seq [88–91]. Nuclei are morphologically simple, making their isolation for RNA-seq easier than intact cells. snRNA-seq compares well with scRNA-seq data, although reads should be mapped to intronic as well as exonic regions owing to the increased presence of unprocessed mRNA in snRNA-seq [92,93]. A caveat to this approach somewhat unique to virology is that it would be limited to viruses that replicate in the nucleus, as cytoplasmic RNA would fail to be captured.

In addition to the challenges of generating a single-cell suspension from a complex tissue, choice of dissociation protocol can influence experimental results [94,95]. There is no perfect way to capture all cells, and protocols may fail to solubilize all populations equivalently, or else, if using harsh detergents to address tough extracellular matrix, may result in the loss of more sensitive cells. There is no universal solution to this issue, but analysis of scRNA-seq data should be performed with an awareness that cell distributions may not reflect their true biological abundance but only their representation in the processed sample.

With regard to identifying different cell populations, cell type annotation can be accomplished using known markers or unsupervised clustering algorithms [96–100]. However, these processes possess their own biases. Cells with unusually low or high gene content, amplification biases, and ambient RNA contamination, all impacting downstream clustering [101]. Pre-filtering, that is removing cells with unusually low or high gene counts, or unusually large amounts of mitochondrial transcripts or a high degree of intron retention, can address some of these issues but may also remove true biological signal unique to viral infections such as general host transcriptome suppression. Another means of addressing these difficulties is to use multimodal datasets with included cell surface marker data, using oligonucleotide labeled antibodies that are resolvable in scRNA-seq in order to combine targeted, hypothesis-driven measurements with massive scRNA-seq datasets [102]. Rare cell types, which themselves can have a massive impact on disease progression, are also difficult to identify using standard clustering algorithms. Preenrichment of rare cells using fluorescence activated cell sorting can also help to address this issue but requires preexisting markers for populations of interest.

Interindividual variation

Disease outcome can vary widely between individuals. While this variation in outcome is broadly accessible to bulk methods, a key component of this variation is differences in individual cellular response as well as relative proportions of different immune cells during infection. scRNA-seq provides sufficient resolution to unravel these differences, providing deeper knowledge into disease severity and progression.

A key component of interindividual variation is the immune response to infection. For instance, in HIV infection, a minority of HIV⁺ individuals, termed “elite controllers,” largely control viral growth. scRNA-seq of these individuals was able to identify elite controller-specific signatures in HIV-specific CD8⁺ T cells in the lymph node, which indicated a role for noncytolytic functions in control of this virus [103]. An additional example, *ex vivo* infection of peripheral blood mononuclear cells (PBMCs) of individuals with divergent genetic ancestry was able to find significant differences between a population of European ancestry and a population of African ancestry in terms of the interferon response when challenged with influenza virus [104]. With an eye towards a potential clinical application, scRNA-seq has also been applied to biomarker study, with the identification of monocyte and B cell-specific signatures associated with progression to severe dengue [105].

Similarly, scRNA-seq has also been used to understand progression of Coronavirus Disease 2019 (COVID-19) disease. scRNA-seq on PBMCs from COVID-19 patients have demonstrated lymphopenia, T-cell exhaustion, expanded myeloid compartments, and an association of expanded classical monocyte population with severe disease [17,106]. In addition, scRNA-seq analysis of bronchiolar lavage fluid samples from COVID-19 patients found a shift in cell populations consistent with increased inflammation in those individuals whose course of disease was more severe [107]. Moving from immune responses, scRNA-seq has also been used to explore the response to antiviral therapy, such as in HIV⁺ patients [108].

Experimental considerations for interindividual variation; batch effects and tissue accessibility

An issue common to both bulk methodologies and scRNA-seq when exploring clinical disease is tissue accessibility. Many tissues cannot be accessed ethically *premortem*, limiting our ability to study disease. Work on peripheral cells, such as PBMCs, can give insight into systemic disease as well as immune state (and provide biomarker information), but we remain limited in terms of studying disease at the site of infection for many different viruses. Organoids, or organ-on-a-chip approaches, are attempting to address this current shortcoming [109].

Batch effects, discussed in a prior section, are also a significant concern when comparing across individual responses. However, interindividual samples have an additional means of sample hashing unavailable in tissue-culture experiments. As scRNA-seq captures genetic information on not only the virus but also the host, this genotype-level variation can be used to identify which cells derived from which individual. This markerless batching of cells from multiple individuals into a single RNA-seq dataset can help to aid in variation in processing and sequencing [110].

Conclusions

There has been an increased adoption of scRNA-seq in virology, as well as broadly in the field of molecular cell biology and even in clinical science. Continued development of these methods and decreases in sequencing cost will likely continue, making this yet another tool for use in molecular virology. Despite the limitations in assessing the heterogeneity at the various levels during viral infection, a number of experimental adjustments and specific pipelines have been devised to carefully evaluate viral and cellular heterogeneity (Tables 2 and 3). With the use of carefully considered experimental pipelines, scRNA-seq can serve as both a powerful hypothesis generation tool, as well as a means of uncoupling whether features correlated at the population level co-occur in individual infected cells. Use of scRNA-seq on clinical samples has also provided significant insight into the nature of viral disease, expanding from

Table 2. Table describing the caveats and experimental modifications that can aid in scRNA-seq to assess heterogeneity at different levels.

Experimental questions	Limitations	Controls
Virion-level heterogeneity	Partial coverage	Generation of full-length transcripts by template switching and sequencing full-length by scISO-seq and HIT-scISO-seq
	Lack of polyadenylated viral transcripts to capture	Use of splinted oligos with virus-specific sequences
	Using read mappers like STAR is inappropriate for mapping viral diversity	Use of more purpose-built pipelines like ViReMa and VODKA2 for mapping viral variation
Cellular-level heterogeneity	Misassigned reads due to template switching, PCR chimeras, and extracellular RNA from cell lysis	Using control cell lines that do not have transcripts of interest Methods like SoupX to remove ambient RNA
	Batch variations between experiments (impacts all experiments but addressed in this section of this review)	Regression analysis of genes varying between replicates Cell Hashing with barcoded antibodies enabling multiplexing of experimental conditions
Tissue-level heterogeneity	Isolation of intact single cells from complex tissues	Single-nuclear RNA sequencing for complex tissues
	Cell clustering—affected by the quality of isolated cells, amplification biases, and misassigned reads	Pre-filtering the data Use of oligonucleotide-labelled antibody to generate multimodal dataset
	Rare cell groups	Preenrichment of rare cell types by using fluorescent-associated cell sorting
Individual-level heterogeneity	Accessibility of tissues from infected individuals	Humanized organ-on-a-chip models from relevant individual cells
	Batch variation between experiments	Use genetic variation between individuals to assign cells to the individual

<https://doi.org/10.1371/journal.ppat.1011898.t002>

transcriptional and cytokine signals out to how individual populations of circulating immune cells are reshaped during and after viral infection.

Looking to the future, barriers to performing scRNA-seq are increasingly becoming lower, with methods being developed that do not even require instrumentation or microfluidics to perform cell encapsidation [111]. Besides these engineering and chemical advances, there is increasing development of bioinformatic pipelines and tools that are appropriate for understanding viral infection. In addition to scRNA-seq, there is increasing use of spatial transcriptomics, with technologies beginning to approach single-cell resolution, permitting the same explorations described here but with an additional aspect of understanding where, in complex tissues, events are occurring [112–115]. We hope that the examples, and considerations, we have provided here aid the field in proceeding to use both scRNA-seq as well as newer technologies to continue to explore fundamental questions in the field of virology.

Table 3. Example computational tools useful in the analysis scRNA-seq datasets during the study of virus–host interactions.

Software	Purpose/Description	References
STAR	Read mapping software with support to specifically map spliced reads.	Dobin et al., 2013 [49]
ViReMa, VODKA	Used to map nonstandard viral genomes, such as DVGs and recombination within viral species	Routh et al., 2014; Achouri et al., 2023 [50,51]
CellRanger	Commercial software from 10x genomics for parsing single-cell sequencing.	10xgenomics.com
SoupX, CellBender	Software used for elimination of technical artifacts in scRNA-seq such as ambient RNA and chimeric PCR artifacts.	Young et al., 2020; Fleming et al., 2023 [71,117]
Viral-Track	It relies on the STAR aligner to map the reads of scRNA-seq data to both the host reference genome and an extensive list of high-quality viral genomes. Further, specific annotation of infected versus bystander cells enables the identification of DEGs between infected and bystander cells.	Bost et al., 2020 [118]
Seurat, Monocle	Used to analyze scRNA-seq data, including dimensionality reduction, batch correction, and differential gene expression. It also uses unsupervised clustering algorithms to assess cell states and is useful in detecting novel cell states.	Hao et al., 2023; Cao et al., 2019 [119,120]

DEG, differentially expressed gene; DVG, defective viral genome; scRNA-seq, single-cell RNA sequencing.

<https://doi.org/10.1371/journal.ppat.1011898.t003>

Author Contributions

Conceptualization: Sharmada Swaminath, Alistair B. Russell.

Visualization: Sharmada Swaminath.

Writing – original draft: Sharmada Swaminath, Alistair B. Russell.

Writing – review & editing: Sharmada Swaminath, Alistair B. Russell.

References

1. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*. 2009; 6:377–382. <https://doi.org/10.1038/nmeth.1315> PMID: 19349980
2. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*. 2014; 9:171–181. <https://doi.org/10.1038/nprot.2014.006> PMID: 24385147
3. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*. 2013; 10:1096–1098. <https://doi.org/10.1038/nmeth.2639> PMID: 24056875
4. Picelli S, Björklund ÅK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res*. 2014; 24:2033–2040. <https://doi.org/10.1101/gr.177881.114> PMID: 25079858
5. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol*. 2010; 11:R119. <https://doi.org/10.1186/gb-2010-11-12-r119> PMID: 21143862
6. Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol*. 2012; 30:777–782. <https://doi.org/10.1038/nbt.2282> PMID: 22820318
7. Weinreich MD, Gasch A, Reznikoff WS. Evidence that the cis preference of the Tn5 transposase is caused by nonproductive multimerization. *Genes Dev*. 1994; 8:2363–2374. <https://doi.org/10.1101/gad.8.19.2363> PMID: 7958902
8. Zhou M, Bhasin A, Reznikoff WS. Molecular genetic analysis of transposase-end DNA sequence recognition: cooperativity of three adjacent base-pairs in specific interaction with a mutant Tn5 transposase¹¹ Edited by G. Smith. *J Mol Biol*. 1998; 276:913–925. <https://doi.org/10.1006/jmbi.1997.1579> PMID: 9566196
9. Zhou M, Reznikoff WS. Tn 5 transposase mutants that alter DNA binding specificity¹¹ Edited by G. Smith *J Mol Biol*. 1997; 271:362–373. <https://doi.org/10.1006/jmbi.1997.1188> PMID: 9268665
10. Berg DE, Davies J, Allet B, Rochaix JD. Transposition of R factor genes to bacteriophage lambda. *Proc Natl Acad Sci*. 1975; 72:3628–3632. <https://doi.org/10.1073/pnas.72.9.3628> PMID: 1059152
11. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse Transcriptase Template Switching: A SMART Approach for Full-Length cDNA Library Construction. *Biotechniques*. 2001; 30:892–897. <https://doi.org/10.2144/01304pf02> PMID: 11314272
12. Petersen M, Wengel J. LNA: a versatile tool for therapeutics and genomics. *Trends Biotechnol*. 2003; 21:74–81. [https://doi.org/10.1016/S0167-7799\(02\)00038-0](https://doi.org/10.1016/S0167-7799(02)00038-0) PMID: 12573856
13. Marcus JS, Anderson WF, Quake SR. Microfluidic Single-Cell mRNA Isolation and Analysis. *Anal Chem*. 2006; 78:3084–3089. <https://doi.org/10.1021/ac0519460> PMID: 16642997
14. DeLaughter DM. The Use of the Fluidigm C1 for RNA Expression Analyses of Single Cells. *Curr Protoc Mol Biol*. 2018; 122:e55. <https://doi.org/10.1002/cpmb.55> PMID: 29851244
15. Delley CL, Abate AR. Modular barcode beads for microfluidic single cell genomics. *Sci Rep*. 2021; 11:10857. <https://doi.org/10.1038/s41598-021-90255-x> PMID: 34035349
16. Gierahn TM, Wadsworth MH, Hughes TK, Bryson BD, Butler A, Satija R, et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods*. 2017; 14:395–398. <https://doi.org/10.1038/nmeth.4179> PMID: 28192419
17. Wilk AJ, Rustagi A, Zhao NQ, Roque J, Martínez-Colón GJ, McKechnie JL, et al. A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat Med*. 2020; 26:1070–1076. <https://doi.org/10.1038/s41591-020-0944-y> PMID: 32514174

18. Wilk AJ, Lee MJ, Wei B, Parks B, Pi R, Martínez-Colón GJ, et al. Multi-omic profiling reveals widespread dysregulation of innate immunity and hematopoiesis in COVID-19. *J Exp Med*. 2021; 218: e20210582. <https://doi.org/10.1084/jem.20210582> PMID: 34128959
19. Mazutis L, Gilbert J, Ung WL, Weitz DA, Griffiths AD, Heyman JA. Single-cell analysis and sorting using droplet-based microfluidics. *Nat Protoc*. 2013; 8:870–891. <https://doi.org/10.1038/nprot.2013.046> PMID: 23558786
20. Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015; 161:1202–1214. <https://doi.org/10.1016/j.cell.2015.05.002> PMID: 26000488
21. Zhang X, Li T, Liu F, Chen Y, Yao J, Li Z, et al. Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol Cell*. 2019; 73:130–142.e5. <https://doi.org/10.1016/j.molcel.2018.10.020> PMID: 30472192
22. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. 2017; 357:661–667. <https://doi.org/10.1126/science.aam8940> PMID: 28818938
23. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science*. 2018; 360:176–182. <https://doi.org/10.1126/science.aam8999> PMID: 29545511
24. Tran V, Papalexi E, Schroeder S, Kim G, Sapre A, Pangallo J, et al. High sensitivity single cell RNA sequencing with split pool barcoding. *bioRxiv*. 2022; 2022.08.27.505512. <https://doi.org/10.1101/2022.08.27.505512>
25. Lauring AS, Andino R. Quasispecies Theory and the Behavior of RNA Viruses. *PLoS Pathog*. 2010; 6: e1001005. <https://doi.org/10.1371/journal.ppat.1001005> PMID: 20661479
26. Vignuzzi M, López CB. Defective viral genomes are key drivers of the virus–host interaction. *Nat Microbiol*. 2019; 4:1075–1087. <https://doi.org/10.1038/s41564-019-0465-y> PMID: 31160826
27. Ngunjiri JM, Sekellick MJ, Marcus PI. Clonogenic Assay of Type A Influenza Viruses Reveals Noninfectious Cell-Killing (Apoptosis-Inducing) Particles ∇ †. *J Virol*. 2008; 82:2673–2680. <https://doi.org/10.1128/jvi.02221-07> PMID: 18184709
28. Brooke CB. Biological activities of “noninfectious” influenza A virus particles. *Future Virol*. 2014; 9:41–51. <https://doi.org/10.2217/fvl.13.118> PMID: 25067941
29. Huang AS, Baltimore D. Defective Viral Particles and Viral Disease Processes. *Nature*. 1970; 226:325–327. <https://doi.org/10.1038/226325a0> PMID: 5439728
30. Sun J, Vera JC, Drnevich J, Lin YT, Ke R, Brooke CB. Single cell heterogeneity in influenza A virus gene expression shapes the innate antiviral response to infection. *PLoS Pathog*. 2020; 16:e1008671. <https://doi.org/10.1371/journal.ppat.1008671> PMID: 32614923
31. Wang C, Forst CV, Chou T-W, Geber A, Wang M, Hamou W, et al. Cell-to-Cell Variation in Defective Virus Expression and Effects on Host Responses during Influenza Virus Infection. *MBio*. 2020; 11. <https://doi.org/10.1128/mBio.02880-19> PMID: 31937643
32. Russell AB, Trapnell C, Bloom JD. Extreme heterogeneity of influenza virus infection in single cells. *Elife*. 2018; 7. <https://doi.org/10.7554/eLife.32303> PMID: 29451492
33. Hamele CE, Russell AB, Heaton NS. In Vivo Profiling of Individual Multiciliated Cells during Acute Influenza A Virus Infection. *J Virol*. 2022; 96:e00505–e00522. <https://doi.org/10.1128/jvi.00505-22> PMID: 35867557
34. Vicary AC, Mendes M, Swaminath S, Lekbua A, Reddan J, Rodriguez ZK, et al. Maximal interferon induction by influenza lacking NS1 is infrequent owing to requirements for replication and export. *PLoS Pathog*. 2023; 19:e1010943. <https://doi.org/10.1371/journal.ppat.1010943> PMID: 37068114
35. Russell AB, Elshina E, Kowalsky JR, Velthuis AJWT, Bloom JD. Single-Cell Virus Sequencing of Influenza Infections That Trigger Innate Immunity. *J Virol*. 2019; 93. <https://doi.org/10.1128/jvi.00500-19> PMID: 31068418
36. Bacsik DJ, Dadonaite B, Butler A, Greaney AJ, Heaton NS, Bloom JD. Influenza virus transcription and progeny production are poorly correlated in single cells. *bioRxiv*. 2023; 2022.08.30.505828. <https://doi.org/10.7554/eLife.86852> PMID: 37675839
37. Fodor E, Velthuis AJW te. Structure and Function of the Influenza Virus Transcription and Replication Machinery. *Csh Perspect Med*. 2019; 10:a038398. <https://doi.org/10.1101/cshperspect.a038398> PMID: 31871230
38. Chen K-Y, Karuppusamy J, O'Neill MB, Opuu V, Bahin M, Foulon S, et al. High-throughput droplet-based analysis of influenza A virus genetic reassortment by single-virus RNA sequencing. *Proc National Acad Sci*. 2023; 120:e2211098120. <https://doi.org/10.1073/pnas.2211098120> PMID: 36730204

39. Quail MA, Swerdlow H, Turner DJ. Improved Protocols for the Illumina Genome Analyzer Sequencing System. *Curr Protoc Hum Genet.* 2009; 62:18.2.1–18.2.27. <https://doi.org/10.1002/0471142905.hg1802s62> PMID: 19582764
40. Alnaji FG, Brooke CB. Influenza virus DI particles: Defective interfering or delightfully interesting? *PLoS Pathog.* 2020; 16:e1008436. <https://doi.org/10.1371/journal.ppat.1008436> PMID: 32437428
41. López CB. Defective Viral Genomes: Critical Danger Signals of Viral Infections. *J Virol.* 2014; 88:8720–8723. <https://doi.org/10.1128/JVI.00707-14> PMID: 24872580
42. Karlsson K, Linnarsson S. Single-cell mRNA isoform diversity in the mouse brain. *BMC Genomics.* 2017; 18:126. <https://doi.org/10.1186/s12864-017-3528-6> PMID: 28158971
43. Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, Floyd T, et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat Biotechnol.* 2018; 36:1197–1202. <https://doi.org/10.1038/nbt.4259> PMID: 30320766
44. Schlecht U, Mok J, Dallett C, Berka J. ConcatSeq: A method for increasing throughput of single molecule sequencing by concatenating short DNA fragments. *Sci Rep.* 2017; 7:5252. <https://doi.org/10.1038/s41598-017-05503-w> PMID: 28701704
45. Shi Z-X, Chen Z-C, Zhong J-Y, Hu K-H, Zheng Y-F, Chen Y, et al. High-throughput and high-accuracy single-cell RNA isoform analysis using PacBio circular consensus sequencing. *Nat Commun.* 2023; 14:2631. <https://doi.org/10.1038/s41467-023-38324-9> PMID: 37149708
46. Verwilt J, Mestdagh P, Vandesompele J. Artifacts and biases of the reverse transcription reaction in RNA sequencing. *RNA.* 2023; 29:889–897. <https://doi.org/10.1261/rna.079623.123> PMID: 36990512
47. Saikia M, Burnham P, Keshavjee SH, Wang MFZ, Heyang M, Moral-Lopez P, et al. Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. *Nat Methods.* 2019; 16:59–62. <https://doi.org/10.1038/s41592-018-0259-9> PMID: 30559431
48. Wulf MG, Maguire S, Dai N, Blondel A, Posfai D, Krishnan K, et al. Chemical capping improves template switching and enhances sequencing of small RNAs. *Nucleic Acids Res.* 2021; 50:e2–e2. <https://doi.org/10.1093/nar/gkab861> PMID: 34581823
49. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
50. Routh A, Johnson JE. Discovery of functional genomic motifs in viruses with ViReMa—a Virus Recombination Mapper—for analysis of next-generation sequencing data. *Nucleic Acids Res.* 2014; 42:e11–e11. <https://doi.org/10.1093/nar/gkt916> PMID: 24137010
51. Achouri E, Felt SA, Hackbart M, López CB. VODKA2: An accurate method to detect copy-back and deletion viral genomes from next-generation sequencing data. *bioRxiv.* 2023; 2023.04.25.537842. <https://doi.org/10.1101/2023.04.25.537842> PMID: 37163001
52. Zanini F, Pu S-Y, Bekerman E, Einav S, Quake SR. Single-cell transcriptional dynamics of flavivirus infection. *Elife.* 2018; 7:e32942. <https://doi.org/10.7554/eLife.32942> PMID: 29451494
53. Drayman N, Patel P, Vistain L, Tay S. HSV-1 single cell analysis reveals anti-viral and developmental programs activation in distinct sub-populations. *Elife.* 2019; 8:e46339. <https://doi.org/10.7554/elife.46339> PMID: 31090537
54. Wyler E, Mösbauer K, Franke V, Diag A, Gottula LT, Arsiè R, et al. Transcriptomic profiling of SARS-CoV-2 infected human cell lines identifies HSP90 as target for COVID-19 therapy. *iScience.* 2021; 24:102151. <https://doi.org/10.1016/j.isci.2021.102151> PMID: 33585804
55. Loveday EK, Sanchez HS, Thomas MM, Chang CB. Single-Cell Infection of Influenza A Virus Using Drop-Based Microfluidics. *Microbiol Spectr.* 2022; 10:e00993–e00922. <https://doi.org/10.1128/spectrum.00993-22> PMID: 36125315
56. Delbrück M. The Burst Size Distribution in the Growth of Bacterial Viruses (Bacteriophages). *J Bacteriol.* 1945; 50:131–135. <https://doi.org/10.1128/jb.50.2.131-135.1945>
57. Wyler E, Franke V, Menegatti J, Kocks C, Boltengagen A, Praktikno S, et al. Single-cell RNA-sequencing of herpes simplex virus 1-infected cells connects NRF2 activation to an antiviral program. *Nat Commun.* 2019; 10:4878. <https://doi.org/10.1038/s41467-019-12894-z> PMID: 31653857
58. O'Neal JT, Upadhyay AA, Wolabaugh A, Patel NB, Bosinger SE, Suthar MS. West Nile Virus-Inclusive Single-Cell RNA Sequencing Reveals Heterogeneity in the Type I Interferon Response within Single Cells. *J Virol.* 2019; 93. <https://doi.org/10.1128/JVI.01778-18> PMID: 30626670
59. Shnyder M, Nachshon A, Krishna B, Poole E, Boshkov A, Binyamin A, et al. Defining the Transcriptional Landscape during Cytomegalovirus Latency with Single-Cell RNA Sequencing. *MBio.* 2018; 9:e00013–e00018. <https://doi.org/10.1128/mBio.00013-18> PMID: 29535194

60. Chen S, Short JAL, Young DF, Killip MJ, Schneider M, Goodbourn S, et al. Heterocellular induction of interferon by negative-sense RNA viruses. *Virology*. 2010; 407:247–255. <https://doi.org/10.1016/j.virol.2010.08.008> PMID: 20833406
61. Kalfass C, Lienenklaus S, Weiss S, Staeheli P. Visualizing the Beta Interferon Response in Mice during Infection with Influenza A Viruses Expressing or Lacking Nonstructural Protein 1. *J Virol*. 2013; 87:6925–6930. <https://doi.org/10.1128/JVI.00283-13> PMID: 23576514
62. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*. 2014; 510:363–369. <https://doi.org/10.1038/nature13437> PMID: 24919153
63. Ramos I, Smith G, Ruf-zamojski F, Martinez-Romero C, Fribourg M, Carbajal EA, et al. Innate immune response to influenza virus at single-cell resolution in human epithelial cells revealed paracrine induction of interferon lambda 1. *J Virol*. 2019. <https://doi.org/10.1128/JVI.00559-19> PMID: 31375585
64. Triana S, Metz-Zumaran C, Ramirez C, Kee C, Doldan P, Shahraz M, et al. Single-cell analyses reveal SARS-CoV-2 interference with intrinsic immune response in the human gut. *Mol Syst Biol*. 2021; 17: e10232. <https://doi.org/10.15252/msb.202110232> PMID: 33904651
65. Kelly JN, Laloli L, V'kovski P, Holwerda M, Portmann J, Thiel V, et al. Comprehensive single cell analysis of pandemic influenza A virus infection in the human airways uncovers cell-type specific host transcriptional signatures relevant for disease progression and pathogenesis. *Front Immunol*. 2022; 13:978824. <https://doi.org/10.3389/fimmu.2022.978824> PMID: 36268025
66. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*. 2016; 167:1853–1866. e17. <https://doi.org/10.1016/j.cell.2016.11.038> PMID: 27984732
67. Sunshine S, Puschnik AS, Replogle JM, Laurie MT, Liu J, Zha BS, et al. Systematic functional interrogation of SARS-CoV-2 host factors using Perturb-seq. *Nat Commun*. 2023; 14:6245. <https://doi.org/10.1038/s41467-023-41788-4> PMID: 37803001
68. Hein MY, Weissman JS. Functional single-cell genomics of human cytomegalovirus infection. *Nat Biotechnol*. 2022; 40:391–401. <https://doi.org/10.1038/s41587-021-01059-3> PMID: 34697476
69. Alvarez M, Rahmani E, Jew B, Garske KM, Miao Z, Benhammou JN, et al. Enhancing droplet-based single-nucleus RNA-seq resolution using the semi-supervised machine learning classifier DIEM. *Sci Rep*. 2020; 10:11019. <https://doi.org/10.1038/s41598-020-67513-5> PMID: 32620816
70. Muskovic W, Powell JE. DropletQC: improved identification of empty droplets and damaged cells in single-cell RNA-seq data. *Genome Biol*. 2021; 22:329. <https://doi.org/10.1186/s13059-021-02547-0> PMID: 34857027
71. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience*. 2020; 9:giaa151. <https://doi.org/10.1093/gigascience/giaa151> PMID: 33367645
72. Yang S, Corbett SE, Koga Y, Wang Z, Johnson WE, Yajima M, et al. Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol*. 2020; 21:57. <https://doi.org/10.1186/s13059-020-1950-6> PMID: 32138770
73. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol*. 2018; 36:421–427. <https://doi.org/10.1038/nbt.4091> PMID: 29608177
74. Tung P-Y, Blischak JD, Hsiao CJ, Knowles DA, Burnett JE, Pritchard JK, et al. Batch effects and the effective design of single-cell gene expression studies. *Sci Rep*. 2017; 7:39921. <https://doi.org/10.1038/srep39921> PMID: 28045081
75. Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M, et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol*. 2020; 21:12. <https://doi.org/10.1186/s13059-019-1850-9> PMID: 31948481
76. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019; 177:1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031> PMID: 31178118
77. Polański K, Park J-E, Young MD, Miao Z, Meyer KB, Teichmann SA. BBKNN: Fast Batch Alignment of Single Cell Transcriptomes. *Bioinformatics*. 2019; 36:964–965. <https://doi.org/10.1093/bioinformatics/btz625> PMID: 31400197
78. Hie B, Bryson B, Berger B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat Biotechnol*. 2019; 37:685–691. <https://doi.org/10.1038/s41587-019-0113-3> PMID: 31061482

79. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM, et al. Cell Hashing with bar-coded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* 2018; 19:224. <https://doi.org/10.1186/s13059-018-1603-1> PMID: 30567574
80. Steuerma Y, Cohen M, Peshes-Yaloz N, Valadarsky L, Cohn O, David E, et al. Dissection of Influenza Infection In Vivo by Single-Cell RNA Sequencing. *Cell Syst.* 2018; 6:679–691.e4. <https://doi.org/10.1016/j.cels.2018.05.008> PMID: 29886109
81. Kotliar D, Lin AE, Logue J, Hughes TK, Khoury NM, Raju SS, et al. Single-Cell Profiling of Ebola Virus Disease In Vivo Reveals Viral and Host Dynamics. *Cell.* 2020; 183:1383–1401.e19. <https://doi.org/10.1016/j.cell.2020.10.002> PMID: 33159858
82. Medaglia C, Kolpakov I, Zwygart AC-A, Zhu Y, Constant S, Huang S, et al. An anti-influenza combined therapy assessed by single cell RNA-sequencing. *Commun Biol.* 2022; 5:1075. <https://doi.org/10.1038/s42003-022-04013-4> PMID: 36216966
83. Qi F, Qian S, Zhang S, Zhang Z. Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. *Biochem Biophys Res Commun.* 2020; 526:135–140. <https://doi.org/10.1016/j.bbrc.2020.03.044> PMID: 32199615
84. Fiege JK, Thiede JM, Nanda HA, Matchett WE, Moore PJ, Montanari NR, et al. Single cell resolution of SARS-CoV-2 tropism, antiviral responses, and susceptibility to therapies in primary human airway epithelium. *PLoS Pathog.* 2021; 17:e1009292. <https://doi.org/10.1371/journal.ppat.1009292> PMID: 33507952
85. Wu TT-H, Travaglini KJ, Rustagi A, Xu D, Zhang Y, Andronov L, et al. Activated interstitial macrophages are a predominant target of viral takeover and focus of inflammation in COVID-19 initiation in human lung. *bioRxiv.* 2022; 2022.05.10.491266. <https://doi.org/10.1101/2022.05.10.491266>
86. Triana S, Stanifer ML, Metz-Zumaran C, Shahraz M, Mukenhirn M, Kee C, et al. Single-cell transcriptomics reveals immune response of intestinal cell types to viral infection. *Mol Syst Biol.* 2021; 17:e9833. <https://doi.org/10.15252/msb.20209833> PMID: 34309190
87. Popova G, Retallack H, Kim CN, Wang A, Shin D, DeRisi JL, et al. Rubella virus tropism and single-cell responses in human primary tissue and microglia-containing organoids. *Elife.* 2023; 12:RP87696. <https://doi.org/10.7554/eLife.87696> PMID: 37470786
88. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods.* 2017; 14:955–958. <https://doi.org/10.1038/nmeth.4407> PMID: 28846088
89. Krishnaswami SR, Grindberg RV, Novotny M, Venepally P, Lacar B, Bhutani K, et al. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc.* 2016; 11:499–524. <https://doi.org/10.1038/nprot.2016.015> PMID: 26890679
90. Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science.* 2016; 352:1586–1590. <https://doi.org/10.1126/science.aaf1204> PMID: 27339989
91. Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol.* 2018; 36:70–80. <https://doi.org/10.1038/nbt.4038> PMID: 29227469
92. Lake BB, Codeluppi S, Yung YC, Gao D, Chun J, Kharchenko PV, et al. A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci Rep.* 2017; 7:6031. <https://doi.org/10.1038/s41598-017-04426-w> PMID: 28729663
93. Bakken TE, Hodge RD, Miller JA, Yao Z, Nguyen TN, Aevermann B, et al. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS ONE.* 2018; 13:e0209648. <https://doi.org/10.1371/journal.pone.0209648> PMID: 30586455
94. Denisenko E, Guo BB, Jones M, Hou R, de Kock L, Lassmann T, et al. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* 2020; 21:130. <https://doi.org/10.1186/s13059-020-02048-6> PMID: 32487174
95. Venema WTCU, Ramírez-Sánchez AD, Bigaeva E, Withoff S, Jonkers I, McIntyre RE, et al. Gut mucosa dissociation protocols influence cell type proportions and single-cell gene expression levels. *Sci Rep.* 2022; 12:9897. <https://doi.org/10.1038/s41598-022-13812-y> PMID: 35701452
96. Duò A, Robinson MD, Sonesson C. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Res.* 2018; 7:1141. <https://doi.org/10.12688/f1000research.15666.3> PMID: 30271584
97. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell.* 2018; 172:1091–1107.e17. <https://doi.org/10.1016/j.cell.2018.02.001> PMID: 29474909

98. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature*. 2020; 587:619–625. <https://doi.org/10.1038/s41586-020-2922-4> PMID: 33208946
99. Hao Y, Stuart T, Kowalski M, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal, and scalable single-cell analysis. *bioRxiv*. 2022; 2022.02.24.481684. <https://doi.org/10.1101/2022.02.24.481684>
100. Andrews TS, Hemberg M. Identifying cell populations with scRNASeq. *Mol Aspects Med*. 2018; 59:114–122. <https://doi.org/10.1016/j.mam.2017.07.002> PMID: 28712804
101. Zhang S, Li X, Lin J, Lin Q, Wong K-C. Review of Single-cell RNA-seq Data Clustering for Cell Type Identification and Characterization. *RNA*. 2023; 29:rna.078965.121. <https://doi.org/10.1261/rna.078965.121> PMID: 36737104
102. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*. 2017; 14:865–868. <https://doi.org/10.1038/nmeth.4380> PMID: 28759029
103. Nguyen S, Deleage C, Darko S, Ransier A, Truong DP, Agarwal D, et al. Elite control of HIV is associated with distinct functional and transcriptional signatures in lymphoid tissue CD8+ T cells. *Sci Transl Med*. 2019; 11. <https://doi.org/10.1126/scitranslmed.aax4077> PMID: 31852798
104. Randolph HE, Fiege JK, Thielen BK, Mickelson CK, Shiratori M, Barroso-Batista J, et al. Genetic ancestry effects on the response to viral infection are pervasive but cell type specific. *Science*. 2021; 374:1127–1133. <https://doi.org/10.1126/science.abg0928> PMID: 34822289
105. Zanini F, Robinson ML, Croote D, Sahoo MK, Sanz AM, Ortiz-Lasso E, et al. Virus-inclusive single-cell RNA sequencing reveals the molecular signature of progression to severe dengue. *Proc Natl Acad Sci*. 2018; 115:E12363–E12369. <https://doi.org/10.1073/pnas.1813819115> PMID: 30530648
106. Lee JS, Park S, Jeong HW, Ahn JY, Choi SJ, Lee H, et al. Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Sci Immunol*. 2020; 5:eabd1554. <https://doi.org/10.1126/sciimmunol.abd1554> PMID: 32651212
107. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med*. 2020; 26:842–844. <https://doi.org/10.1038/s41591-020-0901-9> PMID: 32398875
108. Pollara J, Khanal S, Edwards RW, Hora B, Ferrari G, Haynes BF, et al. Single-cell analysis of immune cell transcriptome during HIV-1 infection and therapy. *BMC Immunol*. 2022; 23:48. <https://doi.org/10.1186/s12865-022-00523-2> PMID: 36175869
109. Shahabipour F, Satta S, Mahmoodi M, Sun A, de Barros NR, Li S, et al. Engineering organ-on-a-chip systems to model viral infections. *Biofabrication*. 2023; 15. <https://doi.org/10.1088/1758-5090/ac6538> PMID: 35390777
110. Xu J, Falconer C, Nguyen Q, Crawford J, McKinnon BD, Mortlock S, et al. Genotype-free demultiplexing of pooled single-cell RNA-seq. *Genome Biol*. 2019; 20:290. <https://doi.org/10.1186/s13059-019-1852-7> PMID: 31856883
111. Clark IC, Fontanez KM, Meltzer RH, Xue Y, Hayford C, May-Zhang A, et al. Microfluidics-free single-cell genomics with templated emulsification. *Nat Biotechnol*. 2023; 1–10. <https://doi.org/10.1038/s41587-023-01685-z> PMID: 36879006
112. Rodrigues SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*. 2019; 363:1463–1467. <https://doi.org/10.1126/science.aaw1219> PMID: 30923225
113. Zhuang X. Spatially resolved single-cell genomics and transcriptomics by imaging. *Nat Methods*. 2021; 18:18–22. <https://doi.org/10.1038/s41592-020-01037-8> PMID: 33408406
114. Wang F, Flanagan J, Su N, Wang L-C, Bui S, Nielson A, et al. RNAscope A Novel in Situ RNA Analysis Platform for Formalin-Fixed, Paraffin-Embedded Tissues. *J Mol Diagn*. 2012; 14:22–29. <https://doi.org/10.1016/j.jmoldx.2011.08.002> PMID: 22166544
115. Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Schapiro D, et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods*. 2019; 16:987–990. <https://doi.org/10.1038/s41592-019-0548-y> PMID: 31501547
116. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*. 2015; 161:1187–1201. <https://doi.org/10.1016/j.cell.2015.04.044> PMID: 26000487
117. Fleming SJ, Chaffin MD, Arduini A, Akkad A-D, Banks E, Marioni JC, et al. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat Methods*. 2023; 20:1323–1335. <https://doi.org/10.1038/s41592-023-01943-7> PMID: 37550580

118. Bost P, Giladi A, Liu Y, Bendjelal Y, Xu G, David E, et al. Host-Viral Infection Maps Reveal Signatures of Severe COVID-19 Patients. *Cell*. 2020; 181:1475–1488.e12. <https://doi.org/10.1016/j.cell.2020.05.006> PMID: 32479746
119. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. 2023:1–12. <https://doi.org/10.1038/s41587-023-01767-y> PMID: 37231261
120. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. 2019; 566:496–502. <https://doi.org/10.1038/s41586-019-0969-x> PMID: 30787437