



## Review

## Exploring viral infection using single-cell sequencing

Sylvie Rato<sup>a,\*</sup>, Monica Golumbeanu<sup>b</sup>, Amalio Telenti<sup>c</sup>, Angela Ciuffi<sup>a</sup><sup>a</sup> Institute of Microbiology, University Hospital Center and University of Lausanne, Bugnon 48, CH-1011 Lausanne, Switzerland<sup>b</sup> Department of Biosystems Science and Engineering, ETH Zürich and SIB Swiss Institute of Bioinformatics, CH-4058 Basel, Switzerland<sup>c</sup> The J. Craig Venter Institute, 4120 Capricorn Lane, La Jolla, CA 92037, USA

## ARTICLE INFO

## Article history:

Received 7 June 2016

Received in revised form 21 October 2016

Accepted 24 October 2016

Available online 2 November 2016

## Keywords:

Single-cell

Heterogeneity

Single-cell sequencing

Virus

Viral infection

## ABSTRACT

Single-cell sequencing (SCS) has emerged as a valuable tool to study cellular heterogeneity in diverse fields, including virology. By studying the viral and cellular genome and/or transcriptome, the dynamics of viral infection can be investigated at single cell level. Most studies have explored the impact of cell-to-cell variation on the viral life cycle from the point of view of the virus, by analyzing viral sequences, and from the point of view of the cell, mainly by analyzing the cellular host transcriptome. In this review, we will focus on recent studies that use single-cell sequencing to explore viral diversity and cell variability in response to viral replication.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Contents

1. Introduction.....	55
2. Single-cell sequencing (SCS).....	57
2.1. Single-cell isolation.....	57
2.2. Single-cell DNA sequencing (DNA-SCS).....	57
2.3. Single-cell RNA sequencing (RNA-SCS).....	59
2.4. DNA/RNA dual sequencing.....	60
2.5. Computational analysis of single-cell sequencing data.....	60
3. Virus infection at the single-cell level.....	63
3.1. Heterogeneity of cell population reflects differences in viral infection.....	63
3.2. Viral genome analysis by SCS.....	63
3.3. Cellular heterogeneity to virus infection analysis by SCS.....	64
3.3.1. DNA sequencing.....	64
3.3.2. RNA sequencing.....	64
4. Conclusions and perspectives.....	65
Author contributions.....	65
Conflicts of interest.....	65
Acknowledgments.....	65
References.....	65

## 1. Introduction

Cells are the basic biological units of living organisms. They are structurally grouped by function into tissues and organs, and they form the most basic component of these structures (at the transcriptome and proteome levels) and thus their identity. Despite sharing the same DNA content and being exposed to the apparent same conditions, cells display some level of functional hetero-

\* Corresponding author.

E-mail addresses: [Sylvie.Ferreira-Rato@chuv.ch](mailto:Sylvie.Ferreira-Rato@chuv.ch) (S. Rato), [monica.golumbeanu@bsse.ethz.ch](mailto:monica.golumbeanu@bsse.ethz.ch) (M. Golumbeanu), [atelenti@jvci.org](mailto:atelenti@jvci.org) (A. Telenti), [Angela.Ciuffi@chuv.ch](mailto:Angela.Ciuffi@chuv.ch) (A. Ciuffi).

geneity. This heterogeneity can be explained by extrinsic features, such as cell identity (cell type/subpopulation/lineage) and cell state/process (cell cycle, circadian rhythm), or by the intrinsic stochastic nature of gene expression (Battich et al., 2015; Satija and Shalek, 2014; Stoeger et al., 2016). These cell-to-cell variations can impact cell function, cell communication and proliferation or cell fate, and behavior of the cell population at large (Altschuler and Wu, 2010; Huang, 2009; Satija and Shalek, 2014). Moreover, rare cells or small fractions of cells can play an important biological role in specific disorders and environments (e.g. cancer cells), but their contribution can be masked by the larger cell population.

Single-cell technology evolved quickly in the last years, from a simple FACS analysis assessing the expression of a specific protein to the most sophisticated techniques that allow the analysis of single-cell genome, transcriptome and proteome. The breakthrough of single-cell analysis, and specifically single-cell-omics, was an important milestone in some areas such as cancer, stem cells, epigenetics or immunology.

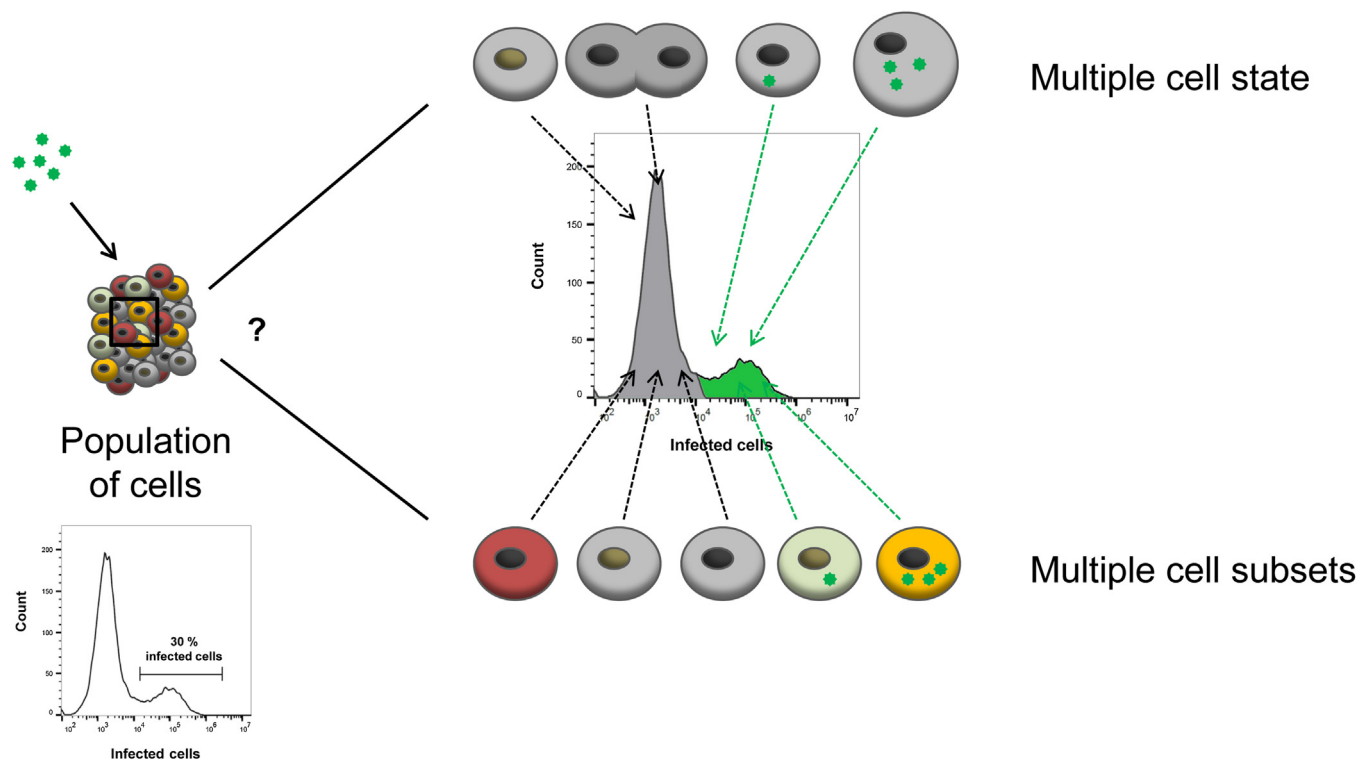
In the immunology field, single-cell technologies were critical for the discovery of new gene networks and novel cell subpopulations, elucidating relationships between cell clonality and their functional phenotypes, and in the discovery of new cell states or cell types within the immune system (Vieira Braga et al., 2016). For example, Stubbington et al., using a new computational method comparing the paired T cell receptor (TCR) sequences from lymphocyte single-cell RNA sequence data, were able to directly correlate T cell clonal origin with the functional phenotype in a mouse Salmonella infection model (Stubbington et al., 2016). Buetner and colleagues developed a single-cell latent variable model (sLVM) allowing identification of otherwise undetectable subpopulations of cells that correspond to different stages during the differentiation of naive T cells into T helper type 2 (Th2) cells (Buetner et al., 2015). In another study, analysis of single-cell messenger

RNA sequencing revealed rare intestinal cell types (Grun et al., 2015). Also, single-cell analysis of CD4<sup>+</sup> T-cell differentiation characterized three major different cell states during Th2 polarization, from the intermediate activated cell state to the mature cytokine-secreting effector state (Proserpio et al., 2016).

In the cancer field, single-cell studies have allowed significant progress in understanding carcinogenesis, progression, metastases and drug resistance (Qian et al., 2016). Single-cell RNA-seq led for example to the identification of distinct tumor subpopulations in lung adenocarcinoma (Min et al., 2015), showed that there was subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells (Kim et al., 2015b), and also led to the identification of distinct gene expression patterns, including candidate biomarkers for melanoma circulating tumor cells (Ramskold et al., 2012).

Single-cell genome, epigenome and transcriptome sequencing led as well to considerable advancement in the stem-cell field, improving the knowledge in both pluripotent and tissue-specific stem cells (Wen and Tang, 2016). Single-cell RNA-seq provided the opportunity to decipher gene expression dynamics during mammalian pre-implantation development, by analyzing transcriptome profiles from both human and mouse cells undergoing pre-implantation (Deng et al., 2014; Yan et al., 2013). Based on single-cell RNA-seq studies, it was also possible to identify new stem cells types (Treutlein et al., 2014) and dissect cell heterogeneity among a stem cell population (reviewed in Wen and Tang, 2016).

Single-cell technology can also be a great asset for the virology field. Viruses are dependent on the host cell to replicate and therefore, heterogeneity in the host cell population will be reflected in viral infection outcome (Fig. 1). In many cases, with many cell types, achieving 100% of infected cells is difficult. This result can be due (i) to the heterogeneity of virus particles, i.e. defective viruses or



**Fig. 1.** Impact of cellular heterogeneity on viral infection. The percentage of infected cells in a population can reflect the cellular heterogeneity in response to viral exposure. The source of this heterogeneity can be due to different cellular states (activation stage or cell cycle) or different cellular subsets of the same cell type (e.g. CD4<sup>+</sup> T cells subsets). Indeed, specific cell states or cell subsets may be permissive to viral infection while other cell states and other cell subsets may be resistant to viral infection. The proportion of infected cells at the population mixed level would mirror the proportion of permissive cells within the total population.

mutations that lead to a decrease of infectivity, or (ii) to the cellular heterogeneity, due to cells in a different cell state when infected (e.g. different state of cell activation or cell cycle), or due to a mixture of different subsets of the same cell type (e.g. CD4<sup>+</sup> T cells subsets). Therefore, single-cell analysis and specifically single-cell sequencing (SCS) allows for investigating the source of heterogeneity, through the joint analysis of virus replication and host cell environment.

To date, single-cell technology has already been applied to study virus evolution and virus-host interaction. The first studies exploiting single-cell analysis used simple techniques to study viral DNA, RNA or protein such as time-lapse microscopy, RT-qPCR, and FACS (recently reviewed in Ciuffi et al., 2016). The development of SCS and its application to the virology field holds great potential to help decipher virus biology and its interaction with the host cell. In this review, we will focus on the main techniques that exist to date to perform SCS as well as on the computational methods for SCS data analysis, and present recent studies that used SCS to explore the impact of host cell heterogeneity on virus infection.

## 2. Single-cell sequencing (SCS)

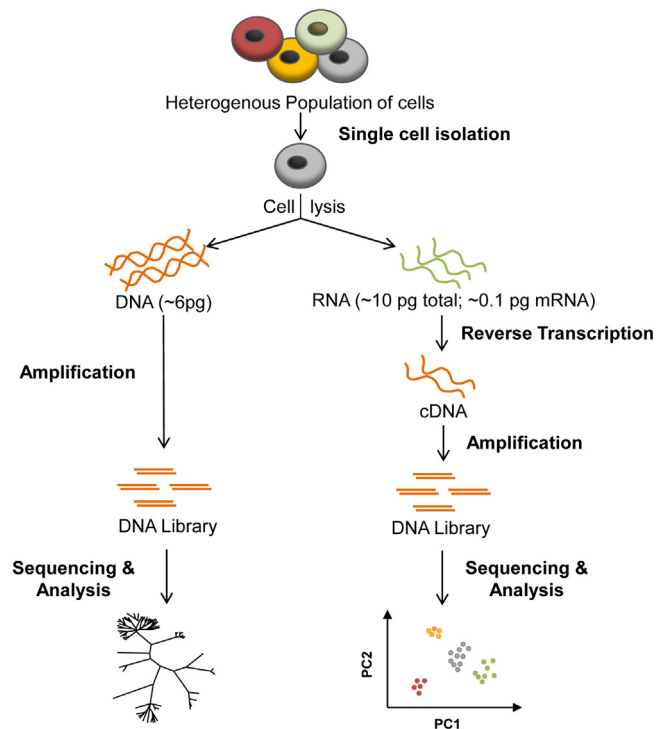
Single-cell sequencing (SCS) has proved to be a powerful tool to study cell heterogeneity through analysis of whole genome and transcriptome of an individual cell. While based on deep-sequencing methods applied to cell population nucleic acid material, SCS had to deal with three main additional challenges: first, how to succeed with the physical isolation of an individual cell, second, how to deal with the limited amount of biological material (6 pg of DNA, 10 pg of total RNA and 0.1 pg of mRNA per cell), and third how to discriminate *bona fide* cell-to-cell variation from noise.

To overcome these issues, SCS techniques include (i) single cell isolation methods, (ii) extensive amplification of DNA or cDNA before library preparation and sequencing, and (iii) development of dedicated data analysis tools (Fig. 2).

### 2.1. Single-cell isolation

Several methods have been used for single-cell isolation from tissue or cell populations in suspension (Table 1, for more detailed information see Gross et al., 2015; Hodne and Weltzien, 2015).

Initially, micromanipulation was often used as gold standard (Brauns and Goos, 2005; Hodne and Weltzien, 2015). It uses a glass micropipette to aspirate one single cell at a time from a cell population under a microscope. Similarly, laser-capture microdissection isolates individual cells from cell population/tissues under direct microscopic visualization (Frost et al., 2015; Kroneis et al., 2016). These methods, although useful when working with tissues or with rare and fragile cells, have significant drawbacks, including time-consuming and labor-intensive manual handling. On the other hand, automated isolation based on Fluorescent-activated cell sorting (FACS) is less time-consuming and allows the isolation of thousands to millions of cells in a couple of hours. Furthermore, cells can also be isolated according to their size, morphology, or expression of specific markers bound to fluorescently labeled antibodies (Gross et al., 2015; Saliba et al., 2014). A more automated isolation method, microfluidics, allows the automated capture of single cells using special Integrated Fluidic Circuits (IFC). Subsequently, single cell capture efficiency can be confirmed by imaging. Each cell undergoes a series of controlled chemical reactions that are carried out in successive microchambers. These reactions include cell lysis, reverse transcription of RNA into cDNA, and PCR amplification (Reece et al., 2016; Wheeler et al., 2003). All reactions are performed in very low volumes (nanoliters) and in a completely automated way. C1™ Single-Cell Auto Prep System,



**Fig. 2.** Single-Cell Sequencing. SCS can be divided in three major steps: (i) single-cell isolation and capture, (ii) nucleic acid amplification and library preparation, and (iii) sequencing and data analysis. First, cells have to be physically isolated prior to cell lysis. Genomic DNA is then amplified while RNA requires a reverse transcription of RNA to cDNA prior to the amplification step. Amplified DNA is then used to prepare libraries that are subsequently sequenced, producing data sets that are computationally analyzed. Therefore, SCS allows analysis of genome and transcriptome of individual cells.

from Fluidigm, is the most commonly used system. Although very convenient, this technique has two major weaknesses: on one hand, the single-cell capture is performed on fixed size chips accommodating three cell size ranges, potentially introducing a size bias in the cells that can be selected; on the other hand, a limited amount of cells (up to 96 cells for RNA-seq and up to 800 cells for Tag-based RNA-seq) can be captured on one chip, which will not be efficient for the isolation of rare cells from a mixed cell population. For that, an additional enrichment step, by FACS for instance, would be required. Similarly, ICELL8 Single-Cell System (Wafer-Gen Biosystems) is able to capture up to 1800 cells without any size bias for Tag-based RNA-Seq. Very recently, microfluidics has evolved to physically isolate individual cells into separated aqueous droplets using hydrogel microsphere emulsion. Each microdroplet constitutes a physical compartment in which cell lysis, reverse transcription, DNA amplification, SCS, or other process can be subsequently carried out (Brouzes et al., 2009; Hummer et al., 2016; Mazutis et al., 2013).

### 2.2. Single-cell DNA sequencing (DNA-SCS)

After isolation, each single-cell is lysed and the nucleic acid material is amplified. Because of the limited amount of starting DNA, an amplification step is first required. For DNA amplification the most commonly used techniques are degenerate-oligonucleotide-primed PCR (DOP-PCR), multiple-displacement-amplification (MDA) and multiple annealing and looping-based amplification cycles (MALBAC) (reviewed in detail in Gawad et al., 2016; Grun and van Oudenaarden, 2015; Huang et al., 2015; Wang and Navin, 2015). These amplification methods exhibit differences in coverage, in efficiency of amplification and in the

**Table 1**  
Single cell isolation techniques.

Method	Equipment and Principle	Advantages	Disadvantages	References
Micromanipulation	<ul style="list-style-type: none"> <li>• Glass micropipette/microscope</li> <li>• Micropipette dispensing</li> </ul>	<ul style="list-style-type: none"> <li>• Good for rare and fragile cells</li> <li>• Low cost</li> </ul>	<ul style="list-style-type: none"> <li>• Labor intensive</li> </ul>	<a href="#">Brauns and Goos (2005)</a>
Laser-capture microdissection	<ul style="list-style-type: none"> <li>• Automated technique</li> <li>• Combination of infrared (IR) laser and microscope visualization</li> <li>• Cell is cut out by laser and isolated</li> </ul>	<ul style="list-style-type: none"> <li>• Good for isolating cells from tissues</li> </ul>	<ul style="list-style-type: none"> <li>• Time consuming</li> <li>• Labor intensive</li> </ul>	<a href="#">Frost et al. (2015)</a>
FACS	<ul style="list-style-type: none"> <li>• FACS sorter</li> <li>• Cells are sorted</li> </ul>	<ul style="list-style-type: none"> <li>• Relative rapid technique</li> <li>• Cells are sorted by fluorescent staining of specific proteins</li> </ul>	<ul style="list-style-type: none"> <li>• Expensive equipment</li> </ul>	<a href="#">Gross et al. (2015)</a>
Microfluidics	<ul style="list-style-type: none"> <li>• Automated capture through Integrated Fluidic Circuit</li> </ul>	<ul style="list-style-type: none"> <li>• Automated capture</li> <li>• Relative rapid technique</li> <li>• Cell isolation to genome amplification</li> <li>• All steps performed in very low volumes</li> </ul>	<ul style="list-style-type: none"> <li>• Expensive equipment</li> <li>• Requirement of specific fixed cell size chambers</li> </ul>	<a href="#">Reece et al. (2016)</a>
Droplets	<ul style="list-style-type: none"> <li>• Automated capture into droplets through Integrated Fluidic Circuits</li> </ul>	<ul style="list-style-type: none"> <li>• Cell isolation to genome amplification performed in the same device</li> <li>• All steps performed in very low volumes</li> </ul>	<ul style="list-style-type: none"> <li>• Expensive equipment</li> </ul>	<a href="#">Mazutis et al. (2013)</a>

**Table 2**  
Single cell DNA sequencing techniques (DNA-SCS).

Method	Main characteristics	Advantages	Disadvantages	References
DOP-PCR (Degenerate-oligonucleotide-primed PCR)	<ul style="list-style-type: none"> <li>• Exponential amplification</li> <li>• Degenerate primers</li> <li>• Changes in annealing temperature</li> </ul>	<ul style="list-style-type: none"> <li>• High reproducibility</li> </ul>	<ul style="list-style-type: none"> <li>• Low genome coverage</li> <li>• High rate of technical errors</li> </ul>	<a href="#">Telenius et al. (1992)</a>
MDA (multiple-displacement-amplification)	<ul style="list-style-type: none"> <li>• Exponential amplification</li> <li>• Random primers</li> <li>• Isothermal conditions</li> </ul>	<ul style="list-style-type: none"> <li>• High genome coverage</li> </ul>	<ul style="list-style-type: none"> <li>• High rate of technical errors</li> </ul>	<a href="#">Spits et al. (2006)</a>
MALBAC (Multiple annealing and looping-based amplification cycles)	<ul style="list-style-type: none"> <li>• Quasi-linear amplification</li> <li>• Random and specific primers</li> <li>• Isothermal conditions</li> </ul>	<ul style="list-style-type: none"> <li>• Intermediate genome coverage</li> <li>• Low rate of technical errors</li> </ul>	<ul style="list-style-type: none"> <li>• Intermediate genome coverage</li> </ul>	<a href="#">Zong et al. (2012)</a>

amount of generated technical errors ([Table 2](#)). Briefly, degenerate-oligonucleotide-primed PCR (DOP-PCR) is a PCR-based technique that uses degenerate primers (containing a known sequence at the 5' end of the primers and terminating with random hexamers at the 3' end) to first amplify DNA at a low annealing temperature for a few cycles. Then, a second round of PCR is carried out with primers specific to the known sequence of the degenerate primers initially used and at a higher annealing temperature. The concentration of primers and polymerase directly affect the efficiency of amplification. DOP-PCR exhibits generally low genome coverage with a high rate of technical errors, which can lead to the over-amplification of small differences and also under-amplification of some regions of the genome ([Telenius et al., 1992](#)).

Similarly to DOP-PCR, multiple-displacement-amplification (MDA) uses a random hexamer primer to launch DNA synthe-

sis, but which is carried out by the Phi29 DNA polymerase, a high-fidelity polymerase with 3'→5' exonuclease and proofreading activity. Annealing and amplification are performed at a constant temperature ([Dean et al., 2001](#); [Spits et al., 2006](#)). MDA has higher genome coverage as compared to DOP-PCR, but still displays some technical errors due to the exponential amplification process.

Multiple annealing and looping-based amplification cycles (MALBAC) technique although combining features from DOP-PCR and MDA has the advantage of performing a quasi-linear amplification of the genome. This technique uses degenerate primers (27 known nucleotides followed by a random octamer) to prime extension by a DNA polymerase through isothermal strand-displacement. Amplicons are terminated at both extremities by the same 27-nt contained in the primers, which can hybridize and form a hairpin thereby preventing further amplification. After 8–12

**Table 3**  
Single cell RNA sequencing techniques (RNA-SCS).

Method	Main characteristics	Advantages	Disadvantages	References
SMART-seq (Switching Mechanism At the end of the 5'-end of the RNA Transcript)	<ul style="list-style-type: none"> <li>• Full-length RNA sequencing</li> <li>• Template switching</li> <li>• Exponential RNA amplification</li> </ul>	<ul style="list-style-type: none"> <li>• Intermediate genome coverage</li> </ul>	<ul style="list-style-type: none"> <li>• Low coverage toward the 5'-end;</li> <li>• Under-representation of transcripts with a high GC-content</li> </ul>	<a href="#">Goetz and Trimarchi (2012)</a>
SMART-seq2, (Switching Mechanism At the end of the 5'-end of the RNA Transcript)	<ul style="list-style-type: none"> <li>• Full-length RNA sequencing</li> <li>• Template switching</li> <li>• Exponential RNA amplification</li> </ul>	<ul style="list-style-type: none"> <li>• High coverage</li> </ul>	<ul style="list-style-type: none"> <li>• High rate of technical errors</li> </ul>	<a href="#">Picelli et al. (2013)</a> and <a href="#">Picelli et al. (2014)</a>
CEL-seq, (Cell Expression by Linear amplification and Sequencing)	<ul style="list-style-type: none"> <li>• 3'-end tag-based</li> <li>• <i>in vitro</i> transcription</li> <li>• Linear RNA amplification</li> </ul>	<ul style="list-style-type: none"> <li>• Low rate of technical errors</li> </ul>	<ul style="list-style-type: none"> <li>• Strong 3'-bias amplification</li> <li>• Low sensitivity for low expressed transcripts</li> </ul>	<a href="#">Hashimshony et al. (2012)</a>
CEL-seq2, (Cell Expression by Linear amplification and Sequencing),	<ul style="list-style-type: none"> <li>• 3'-end tag-based</li> <li>• <i>in vitro</i> transcription</li> <li>• Linear RNA amplification</li> </ul>	<ul style="list-style-type: none"> <li>• High sensitivity</li> <li>• Low costs</li> <li>• Low time-consuming</li> </ul>	<ul style="list-style-type: none"> <li>• 3'-bias amplification</li> <li>• Low sensitivity for low expressed transcripts</li> </ul>	<a href="#">Hashimshony et al. (2016)</a>
STRT-seq (STRT/C1). (Single-cell tagged reverse transcription sequencing)	<ul style="list-style-type: none"> <li>• 5'-end tag-based;</li> <li>• Template switching</li> <li>• Exponential RNA amplification</li> </ul>	<ul style="list-style-type: none"> <li>• High sensitivity</li> <li>• High coverage</li> <li>• Automated approach</li> </ul>	<ul style="list-style-type: none"> <li>• 5'-bias amplification</li> </ul>	<a href="#">Islam et al. (2014)</a>
Cyto-seq (Gene expression cytometry)	<ul style="list-style-type: none"> <li>• Bead-based</li> <li>• CEL-seq amplification method</li> </ul>	<ul style="list-style-type: none"> <li>• Low volume</li> <li>• Closed system (low risk of contaminations)</li> </ul>	<ul style="list-style-type: none"> <li>• 3'-bias amplification</li> </ul>	<a href="#">Fan et al. (2015)</a>
Drop-seq	<ul style="list-style-type: none"> <li>• Droplet and bead-based</li> <li>• Template switching</li> <li>• Exponential RNA amplification</li> </ul>	<ul style="list-style-type: none"> <li>• Low volume</li> <li>• Closed system (low risk of contaminations)</li> </ul>	<ul style="list-style-type: none"> <li>• 3'-bias amplification</li> </ul>	<a href="#">Mazutis et al. (2013)</a>
in-Drop	<ul style="list-style-type: none"> <li>• Droplet and bead-based</li> <li>• CEL-seq</li> </ul>	<ul style="list-style-type: none"> <li>• Low volume</li> <li>• Closed system (low risk of contaminations)</li> </ul>	<ul style="list-style-type: none"> <li>• 3'-bias amplification</li> </ul>	<a href="#">Hummer et al. (2016)</a>

cycles of quasi-linear amplification, there is enough amplified DNA for sequencing. The quasi-linear amplification renders this technique suitable as it reduces significantly the sequence-dependent bias that exists with exponential amplification ([Zong et al., 2012](#)).

### 2.3. Single-cell RNA sequencing (RNA-SCS)

Single-cell RNA sequencing first starts with a step of reverse-transcription (RT) to generate cDNA as it is not yet possible to sequence RNA directly. This step is then followed by cDNA amplification and sequencing. The most current methods used for single-cell RNA sequencing are SMART-seq, SMART-seq2, CEL-seq, CEL-seq2, STRT-seq, Cyto-seq and Drop-seq, differing on the RT (traditional, modified RT-PCR or T7-*in vitro* transcription) and on the full length or tag-based cDNA amplification ([Table 3](#), for more information refer to [Picelli, 2016](#); [Saliba et al., 2014](#)). These differences will be reflected on the length of cDNA amplified, on the coverage efficiency and on 3' or 5' amplification bias.

SMART (Switching Mechanism At the end of the 5'-end of the RNA Transcript)-seq is a full-length RNA sequencing method that exploits the dual activity of the Moloney Murine Leukemia Virus (MMLV) reverse transcriptase, *i.e.* reverse transcription (RT) and template switching (TS). RT reaction requires a primer (often an oligodT, tailed to a known sequence for subsequent PCR amplification) and a RNA template in order to achieve first-strand cDNA synthesis. TS is characterized by the 3' adjunction of 3–4 cytosines

once the RT is completed. This 3' C-overhang is then used to ligate a compatible oligonucleotide duplex with a known sequence useful for subsequent PCR amplification. This ensures that only full-length transcripts are further amplified by PCR. Furthermore, the identification of strand specificity is maintained due to the localization of the added C residues (method detailed in [Goetz and Trimarchi, 2012](#); [Ramskold et al., 2012](#)). SMART-seq allows a high coverage of the transcriptome with an intermediate sensitivity, but with low read coverage toward the 5'-end of the transcripts and an under-representation of transcripts with a high GC-content ([Picelli et al., 2013](#)). SMART-seq2 improved the previous technique using (i) a refinement of the RT and TS steps and (ii) a pre-amplification step to increase the cDNA yield. This results in the increased ability of detecting gene expression, especially for the genes with high GC-content and in lower technical biases. Nevertheless, this method does not allow identification of strand-specificity of mRNAs and is still labor intensive, since the samples can be pooled just prior to sequencing. Nevertheless, this can be now overcome with automated platforms or microfluidics devices such as Fluidigm ([Picelli et al., 2013, 2014](#)). Indeed, recently Clontech™ have implemented the SMART-Seq v4 Ultra. Low Input RNA Kit, which includes all the improvements of SMARTseq2 and that can be used with Fluidigm C1 instrument. Both SMART-seq and SMART-seq2 are full-length RNA-sequencing methods based on exponential-PCR amplification. Different methods were developed using *in vitro* transcription (IVT), allowing for targeted and linear RNA amplification. CEL-Seq (Cell

**Table 4**  
Combined Single-cell sequencing techniques.

Method	Main characteristics	Advantages	Disadvantages	References
G&T-seq (Genome and Transcriptome Sequencing)	<ul style="list-style-type: none"> <li>• Separation of DNA from RNA before amplification</li> <li>• DNA and RNA sequencing</li> <li>• SMART-seq2 + any other DNA amplification technique</li> </ul>	<ul style="list-style-type: none"> <li>• DNA and RNA sequencing from the same single cell</li> </ul>	<ul style="list-style-type: none"> <li>• High rate of technical errors</li> </ul>	Goetz and Trimarchi (2012)
DR-seq (DNA and RNA Sequencing)	<ul style="list-style-type: none"> <li>• Simultaneous DNA and RNA amplification</li> <li>• CEL-seq followed by MALBAC</li> </ul>	<ul style="list-style-type: none"> <li>• DNA and RNA sequencing from the same single cell</li> <li>• Reduced handling (no separation step) and transfer</li> </ul>	<ul style="list-style-type: none"> <li>• 3'-bias amplification</li> </ul>	Picelli et al. (2013) and Picelli et al. (2014)

Expression by Linear amplification and Sequencing) method uses a specific barcode system in addition to IVT. Briefly, single-cell RNA is uniquely bar-coded for reverse transcription and, after second-strand synthesis the reactions are pooled for IVT. The amplified RNA is then fragmented, purified and sequenced (Hashimshony et al., 2012). CEL-seq allows for strand specificity and has high barcoding efficiency. Nevertheless, CEL-seq has a strong 3'-bias and a low sensitivity for lowly expressed transcripts (Hashimshony et al., 2012). An updated version of this method, CEL-seq2, has been developed, allowing for higher sensitivity, lower costs and faster processing (Hashimshony et al., 2016). STRT-seq (Single-cell tagged reverse transcription sequencing) is a method evolved from SMART-seq that was modified to a tag-based method with the introduction of 6-base random barcodes at 3'-end. Additional modifications were implemented to increase the TS efficiency (Islam et al., 2011, 2012). The original STRT-seq protocol was subsequently simplified, improved and adapted to the C1 Single-Cell Auto Prep system (Fluidigm) and is known as STRT/C1 (Islam et al., 2014). STRT-seq and STRT/C1 display a high sensitivity and good genome coverage although is still limited by 5'-end bias.

Single-cell RNA-seq methods were highly improved at levels of efficiency, rapidity and generation of technical errors. Further improvements are still being developed, aiming at scalable approaches compatible with routine analyses, and allowing digital gene expression profiling of thousands of single cells across an arbitrary number of genes, without using robotics or automation. For that reasons, novel methods were combined with established single-cell RNA-seq techniques, such as Cyto-seq which allows for volume reduction (Cytpicoliter wells) (Fan et al., 2015), Drop-seq and in-Drop that are based on physical displacement of emulsion droplet) (Brouzes et al., 2009; Hummer et al., 2016; Mazutis et al., 2013).

#### 2.4. DNA/RNA dual sequencing

Single-cell technology was further improved with the recent development of methods allowing the analysis of both the genome and the transcriptome of the same single cell. Two methods have been described for this dual –omics, Genome and Transcriptome Sequencing (G&T-seq) and DNA and RNA sequencing (DR-seq) (Table 4). In G&T-seq poly(A) RNA is captured by a biotinylated oligo-dT primer, thus separated from genomic DNA, and then both the genome and the transcriptome are amplified and sequenced in parallel (Macaulay et al., 2015). The full-length mRNA is amplified using SMART-seq2 and the genomic DNA can be amplified using different Whole Genome-Amplification (WGA) methods. In contrast, DR-seq does not separate RNA from DNA before amplification, but rather directly lyses and reverse transcribes RNA first (similarly to CEL-seq), and then amplifies both genomic DNA and cDNA by quasi-linear whole genome amplification (MALBAC method) (Dey et al., 2015).

Recently, a step further has been made to uncover all genetic information of a single cell. New exciting methods start to appear combining RNA and/or DNA SCS with epigenome sequencing, through sequencing of single-cell methylome by bisulfite sequencing (scRRBS) (Hou et al., 2016) or by investigating chromatin interactions and chromosomal organization (3D genome structure) through single-cell Hi-C (Furlan-Magaril et al., 2015; Nagano et al., 2015).

#### 2.5. Computational analysis of single-cell sequencing data

The challenging technological manipulation during SCS experiments, such as capture of single cells, dealing with minute amounts of DNA or RNA, and amplification of the biological material, introduce systematic noise and biases in the data. These render data processing complex and analysis difficult to discriminate between *bona fide* cell-to-cell variation and technical noise. Currently, there are several tools available for the analysis of single-cell sequencing data, taking these issues in consideration.

Initial preprocessing of SCS data typically involves similar steps as when dealing with conventional bulk DNA or RNA sequencing. These steps include read trimming, quality control and mapping to the reference genome, using the same tools as for bulk sequencing (Tables 5 and 6). However, SCS datasets contain a high level of artifacts and noise, such as allelic dropout, non-uniform coverage and spurious sequencing errors; therefore, after alignment, they require a tailored downstream analysis (Gawad et al., 2016; Li et al., 2016; Stegle et al., 2015). A necessary quality assessment of each individual single-cell sample is necessary. This step consists of inspecting different data preprocessing results, such as read count, percentage of uniquely aligned reads, genome coverage and allelic dropout. These features allow detecting low-quality samples with broken or dead cells, and select the viable datasets, relevant for further study of biological questions (Bacher and Kendzierski, 2016; Gawad et al., 2016; Grun and van Oudenaarden, 2015). Ilicic et al. propose a complete pipeline for RNA-SCS data processing, including a classification method to discard unreliable single-cell datasets. Their approach accounts for features such as read count, read mappability, proportions of non-exonic and mitochondrial reads and transcriptome variance (Ilicic et al., 2016).

Due to false positive rates specific to the whole-genome amplification methods (de Bourcy et al., 2014), non-uniform coverage, sequencing errors and high rates of allelic dropouts, detecting copy number variants (CNVs) or single nucleotide variants (SNVs) in DNA-SCS data is a difficult task. Currently, in order to ease variant detection, a handful of studies attempt to quantify the different biases in single-cell data and propose normalization approaches to correct for them (Table 5 and (Cheng et al., 2011; Zhang et al., 2015)). Zong et al. incorporate bulk DNA sequencing data to train a mixture model, which identifies SNVs in DNA-SCS, and a hidden Markov Model (HMM) with a pre-defined transition matrix for CNV

**Table 5**  
Computational methods designed for analysis of DNA-SCS data.

Analysis step	Suggested Methods	References	
Preprocessing <ul style="list-style-type: none"> <li>• Trimming</li> <li>• Read quality control</li> <li>• Alignment</li> <li>• PCR duplicate removal</li> </ul>	<ul style="list-style-type: none"> <li>• Cutadapt, Trimmomatic</li> <li>• FastQC</li> <li>• BWA, Bowtie2</li> <li>• PICARD</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Martin (2011) and Bolger et al. (2014)</a></li> <li>• <a href="#">Andrews (2016)</a></li> <li>• <a href="#">McKenna et al. (2010) and Langmead and Salzberg (2012)</a></li> <li>• <a href="https://broadinstitute.github.io/picard/">https://broadinstitute.github.io/picard/</a></li> </ul>	
	Sample quality control	<ul style="list-style-type: none"> <li>• Inspect various alignment statistics such as read count and proportion of mapped reads</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Bacher and Kendzierski (2016), Gawad et al. (2016) and Grün and van Oudenaarden (2015)</a></li> </ul>
	Normalization and variant detection	<ul style="list-style-type: none"> <li>• Amplification bias quantification and coverage adjustment using a statistical model</li> <li>• Normalization method for single-cell CGH arrays</li> <li>• Mixture model combined with GATK for SNV detection and hidden Markov models for CNV identification</li> <li>• Monovar statistical model</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Zhang et al. (2015)</a></li> <li>• <a href="#">Cheng et al. (2011)</a></li> <li>• <a href="#">Zong et al. (2012)</a></li> <li>• <a href="#">Zafar et al. (2016)</a></li> </ul>
		Genome assembly	<ul style="list-style-type: none"> <li>• SPAdes collection of genome assembly tools</li> <li>• IDBA-UD</li> <li>• K-mer correction method for genome assembly</li> </ul>

detection ([Zong et al., 2012](#)). Another approach for SNV calling in DNA-SCS data is the recently proposed method Monovar. By the means of a statistical model, the method takes into account allelic dropout, uneven coverage and false positive effects and calculates posterior probabilities for each locus of incorporating an SNV ([Zafar et al., 2016](#)).

DNA-SCS also allows novel genome assembly of viruses and other organisms that are not easy to culture for subsequent bulk sequencing and assembly analysis. The traditional approaches of many genome assemblers rely on the uniform read coverage assumption, which does not stand for DNA-SCS data. Novel assembly methods such as SPAdes ([Bankevich et al., 2012](#)) and IDBA-UD ([Peng et al., 2012](#)) tackle issues such as uneven coverage, amplification errors and spurious reads and perform adequate k-mer correction. The single-cell assembler SPAdes comes with an arsenal of genome assembly tools, capable of handling diploid genomes with numerous polymorphisms (dipSPAdes [Safonova et al., 2015](#)), Illumina TruSeq barcoded reads (TruSPAdes), or RNA assembly (rnaSPAdes). [Nikolenko et al.](#) propose a novel k-mer error-correction method, which uses connected components in Hamming graphs and Bayesian clustering of these components prior genome assembly ([Nikolenko et al., 2013](#)).

After preprocessing, alignment and cell quality control, the naturally following computational task in RNA-SCS is gene expression quantification. Raw gene expression counts can be obtained with existing tools for bulk RNA-Seq, such as HTSeq ([Anders et al., 2015](#)). Note that relative expression measures such as reads per kilobase per million (RPKM) are not applicable for RNA-SCS data because of unsuitable assumptions such as homogeneous RNA and gene expression across cells ([Bacher and Kendzierski, 2016](#)). Following normalization of the read counts, some of the typical downstream analyses comprise differential expression, identification of cellular subpopulations based on heterogeneous gene

expression patterns, studying splicing heterogeneity across cells, pseudo-temporal ordering of cells and functional analysis ([Table 6](#)).

With approximately 10%–20% of the total number of transcripts being successfully reverse transcribed in RNA-SCS ([Islam et al., 2014; Kolodziejczyk et al., 2015](#)), and with numerous sources of variation and technical noise ([Grün et al., 2014](#)), normalization of raw gene expression counts proves to be challenging. The key necessity in this step is deconvolution of technical and biological variations. Several existing computational tools use spike-ins to quantify technical noise and identify genes showing significant biological variation across cells ([Table 6](#) and ([Brennecke et al., 2013; Kim et al., 2015a; Vallejos et al., 2015, 2016](#))). In absence of spike-ins, normalization methods rely on the use of global scaling factors as proposed in the DESeq tool ([Anders and Huber, 2010](#)). However, this might not be well suited for single cell data, as the idea of global scale factors implies homogeneous RNA amounts across the different cells, which might not be the case ([Bacher and Kendzierski, 2016](#)); furthermore, RNA-SCS data contains a large number of dropout events, thus, the gene expression profiles contain a high proportion of zeros. There are available computational tools that model the dropout events into a separate component to infer the expected gene expression levels in each cell ([Finak et al., 2015; Kharchenko et al., 2014](#)). Another approach combines the gene expression levels by pooling read counts from different cells, and deconvolving size factors characteristic to each cell ([L. Lun et al., 2016](#)).

Clustering of single cells based on their gene expression profiles leads to the stratification of the total population of single cells into subpopulations or subtypes of functional relevance. Several studies have performed this task using conventional clustering methods such as principle component analysis (PCA), hierarchical clustering, or K-Means ([Table 6](#); for a detailed review see [Bacher and Kendzierski, 2016](#)). However, many of the classical clustering

**Table 6**  
Computational methods designed for analysis of RNA-SCS data.

Analysis step	Suggested Methods	References
Preprocessing		
• Trimming	• Cutadapt, Trimmomatic	• <a href="#">Martin (2011)</a> and <a href="#">Bolger et al. (2014)</a>
• Read quality control	• PRINSEQ, FastQC	• <a href="#">Schmieder and Edwards (2011)</a> and <a href="#">Andrews (2016)</a>
• Alignment	• STAR, TopHat	• <a href="#">Dobin et al., (2012)</a> and <a href="#">Kim et al. (2013)</a>
• Expression quantification	• HTSeq	• <a href="#">Anders et al. (2015)</a>
Sample quality control	• Classification of cell samples using support vector machines on a selected collection of sample features	• <a href="#">Ilicic et al. (2016)</a>
Normalization, technical noise quantification	• Linear model for estimating technical noise using a coefficient of variation of spike-ins; $\chi^2$ -test for identifying genes with high biological variability	• <a href="#">Brennecke et al. (2013)</a>
	• Gamma regression model trained on the concentrations of spike-ins for estimating noise-free gene expression levels	• <a href="#">Ding et al. (2015)</a>
	• Normalization based on Poisson bootstrapping using spike-ins; statistical test for differential expression identification	• <a href="#">Katayama et al. (2013)</a>
	• Statistical generative model fitted to spike-ins to learn technically-induced variability	• <a href="#">Kim et al. (2015a)</a>
	• scLVM statistical model which identifies variability associated to unknown effects such as cell-cycle or technical noise	• <a href="#">Buettner et al. (2015)</a>
	• Bayesian hierarchical model BASiCS designed to separate technical and biological variation in gene expression measurements	• <a href="#">Vallejos et al. (2015)</a> and <a href="#">Vallejos et al. (2016)</a>
	• Method which uses linear equations to deconvolve cell-specific normalization factors from global factors corresponding to pools of cells	• <a href="#">Lun et al. (2016)</a>
Clustering and differential expression analysis	• Bayesian mixture model accounting for allelic dropout events	• <a href="#">Kharchenko et al. (2014)</a>
	• Zero-inflated factor analysis (ZIFA) method which mitigates the effect of dropout events for dimensionality reduction and clustering	• <a href="#">Pierson and Yau (2015)</a>
	• MAST method consisting of a generalized linear model for analyzing time-series RNA-SCS data, performing differential expression assessment and gene set enrichment analysis	• <a href="#">Finak et al. (2015)</a>
	• Shared nearest neighbor clustering approach based on clique joining (SNN-Cliq)	• <a href="#">Xu and Su (2015)</a>
	• RaceID, a k-means-based clustering method with focus on identifying rare cells	• <a href="#">Grun et al. (2015)</a>
	• BackSPIN single-cell biclustering	• <a href="#">Zeisel et al. (2015)</a>
	• PcaReduce method, combines PCA and hierarchical clustering to identify novel cell types	• <a href="#">Žurauskienė and Yau (2016)</a>
	• Method that applies classical clustering methods to transcript compatibility counts from SCS data	• <a href="#">Ntranos et al. (2016)</a>
	• Pathway and gene set overdispersion analysis (PAGODA)	• <a href="#">Fan et al. (2016)</a>
Pseudo-temporal ordering, cell-state hierarchies identification	• Single-cell clustering using bifurcation analysis (SCUBA) identifies a cell-state hierarchy tree from time-series SCS data	• <a href="#">Marco et al. (2014)</a>
	• Sincell – a comprehensive workflow for identification of cell-state hierarchies	• <a href="#">Juliá et al. (2015)</a>
	• Monocle: pseudo-temporal ordering of single cells based on their transcriptome profiles used to build a minimum spanning tree	• <a href="#">Trapnell et al. (2014)</a>
	• Oscope oscillation model for identifying clusters of oscillating genes and ordering single cells according to recurring cyclic states;	• <a href="#">Leng et al. (2015)</a>
	• Waterfall pipeline for ordering single cells according to an inferred expression trajectory using k-means and PCA	• <a href="#">Shin et al. (2015)</a>
Characterization of gene expression kinetics	• Statistical Bayesian model for estimating kinetics parameters describing gene expression dynamics	• <a href="#">Kim and Marioni (2013)</a>
Splicing heterogeneity analysis	• SingleSplice method, constructs a weighted splice graph and employs Gamma regression models fitted to spike-ins measurements to estimate technical variability due to noise and dropout events; the method then uses bootstrapping to identify isoform biological variation across cells	• <a href="#">Welch et al. (2016)</a>



methods may be vulnerable to the existence of confounding factors, such as cell cycle, which should be corrected for. For this reason, novel clustering methods, tailored for single-cell gene expression data, tackle the confounding, hidden, effects and robustly identify different subgroups of cells (Buettner et al., 2015). Subsequently, the clustered individual cells can be further organized in so-called cellular hierarchies, depending on their transcriptional state profiles. This type of analysis provides a glimpse of the heterogeneity in transcription and regulatory dynamics. Juliá et al. introduced the Sincell R package, which comprises a comprehensive and rigorous workflow for cell-state hierarchies identification. Sincell gives the user the opportunity to continuously adapt the analysis pipeline and choose between alternative approaches to identify, evaluate and interpret cell-state hierarchies (Juliá et al., 2015).

RNA-SCS gives the opportunity of studying other interesting biological questions that are computationally challenging, such as identifying gene expression kinetics (Kim and Marioni, 2013) and analysis of splicing heterogeneity across different cells (Shalek et al., 2013; Welch et al., 2016).

### 3. Virus infection at the single-cell level

#### 3.1. Heterogeneity of cell population reflects differences in viral infection

When studying viral infection, we cannot separate the virus from the host cell and subsequently, to the intrinsic heterogeneity that exists in the cell population. Differences in cell subpopulations/subsets or in cell cycle/differentiation stages reflect into the specific composition of cells, both in the nature and the amount of RNA molecules and proteins, thereby influencing viral replication and pathogenesis. This also defines the cell susceptibility and permissiveness to viral infection.

A large number of examples are known to correlate cell heterogeneity with differences in susceptibility to viral infection. For example, differentiation of monocytes to macrophages leads to different susceptibilities to viral infection. Indeed, macrophages are more susceptible than monocytes to human influenza A virus (IAV) (Hoeve et al., 2012) or to human immunodeficiency virus (HIV) infection. This differential susceptibility was shown to be dependent on the cell differentiation status in case of HIV, correlating with differences in gene expression, including APOBEC3A and 3G restriction factors (Peng et al., 2007), CCR5 co-receptor (Tuttle et al., 1998), or with the expression of miRNA that target the 3'UTR of HIV-1 transcripts (Wang et al., 2009).

Cell cycle can also determine viral infection efficiency, as illustrated by retrovirus infections (Katz et al., 2005). Upon cell entry, common to all retroviruses, the viral RNA genome (positive, single-stranded) is first reverse-transcribed to double-stranded DNA and then integrated in the host cell genome. To achieve this last step, the murine leukemia virus (MLV), a gammaretrovirus, needs the cell to be dividing so that the nuclear membrane is disrupted and the viral DNA can be tethered to the host chromatin (reviewed in Ciuffi, 2016; Suzuki and Craigie, 2007). In contrast, the avian sarcoma virus (ASV), an alpharetrovirus, can integrate its genetic material into the host genome of non-dividing cells, but requires cells to be in S-phase to do so efficiently (Humphries et al., 1981). The human foamy virus (HFV), a spumavirus, is unable to infect productively G1/S or G2 growth-arrested cells, thereby requiring cell cycle progression in order to successfully replicate (Bieniasz et al., 1995; Patton et al., 2004). HIV, the prototypical lentivirus, is able to infect both dividing and non-dividing cells. However, depending on cellular division, HIV DNA reaches nuclear chromatin differently, due to the absence or presence of an intact nuclear envelope, and thus to the interaction with nuclear pore components. Thus, the cell cycle

determines the entry route of the viral genome to the chromatin, contributing to integration site selection and potentially impacting viral gene expression (Ciuffi, 2016; Demeulemeester et al., 2015).

In addition, virus infection itself triggers the expression of numerous cellular genes, which can be different from cell to cell and that adds to the cellular heterogeneity. This is illustrated in a study from Mohammadi et al. where HIV replication in SupT1 cells (a CD4<sup>+</sup> T cell line) is followed over time (Mohammadi et al., 2013). Kinetic differences appeared very quickly during viral progression, where most cells reverse transcribed the viral RNA genome within 7 to 10 h post-virus exposure and expressed virus-encoded green fluorescent protein (GFP) within 22 to 25 h. However, a minority of cells did so earlier or later, *i.e.* reverse-transcribing as quickly as 3 h or as lately as 23 h post-exposure and expressing GFP as quickly as 16 h and as lately as 32 h post-exposure. Also, Holmes et al. investigated the dynamics of the HIV-1 life cycle using time-to-inhibitor-addition experiments in MT4 cells (CD4<sup>+</sup>T cell line) and observed similar dynamics of viral progression although within a different time frame kinetics (Holmes et al., 2015). These studies clearly exemplify the different abilities of individual cells to support and conduct productive infection.

At last, differences in innate immune response, due to the cell type or cell state, can affect the outcome of viral infection. Differential expression of effector molecules, at least *in vitro*, can contribute to cell permissiveness to viral infection. Several studies comparing the permissiveness of cell lines to HIV infection showed critical differences in the expression of such proteins and led to the discovery of Interferon (IFN)-induced anti-viral proteins, also called restriction factors, like MX2 (Kane et al., 2013), APOBEC3G (Sheehy et al., 2002) or BST2/tetherin (Kluge et al., 2015; Neil et al., 2008). Similarly, Ye et al. compared the proteome composition of two hepatocyte cell lines differing on their permissiveness to hepatitis C virus (HCV) infection (Ye et al., 2015). They observed that, compared to the other cell line (Huh7.5.1), the less permissive cell line (HepG2) expressed high levels of the major vault protein (MPV) protein, an IFN-inducible gene, thereby leading to an inhibition of HCV replication.

Moreover, innate immune defenses that block viral replication are not expressed similarly in all cell types, thus contributing to the cell susceptibility to virus infection. For instance, primary CD4<sup>+</sup> T cells are more or less susceptible to HIV infection depending on whether they are in an activated or quiescent state, respectively (Doitsh and Greene, 2016; Jakobsen et al., 2015; Pan et al., 2013; Rausell et al., 2016). This was shown to be due, at least in part, to the differential expression of cellular restriction factors, such as SAMHD1, in the different cell states. Similarly, transformed cell lines are usually more permissive to viral infection because they display various defects in expression of the innate immune defense (Rausell et al., 2016).

Studies at the population level revealed cellular heterogeneity and a subsequent impact on the viral outcome. The advances of single-cell technologies allow to go further studying viral infection in the context of this cellular heterogeneity. Specifically, single-cell sequencing (SCS) allows for evaluating viral diversity independently of the cell population bias, and for studying the impact of the cellular heterogeneity on viral replication. SCS applied to virology is an emerging field. Studies so far investigated viral diversity at the single-cell level, through viral genome analysis, or explored cellular variability in response to viral infection by evaluating the genome and transcriptome of a single infected cell.

#### 3.2. Viral genome analysis by SCS

Several studies have used SCS to better understand viral diversity and to identify different quasispecies at a single-cell level. McWilliam Leitch and McLauchlan analyzed viral RNA heterogene-

ity. They established a system to amplify the genome of HCV from single infected cells and analyzed the viral quasispecies diversity (McWilliam Leitch and McLauchlan, 2013). Using RT-qPCR and viral RNA deep-sequencing, they determined the quasispecies composition of 16 individual single cells as compared to the total cell population. They determined that, on average, one single cell contains 113 copies of replicon RNA (ranging from 84 to 160 copies). Moreover, results showed that HCV quasispecies composition differed significantly among cells, indicating a “cellular compartmentalization” of replicon HCV RNA with distinct sequences, which can probably lead to divergent evolution. Although the wild-type sequence was the most prevalent sequence identified in both single and total cells, different variants were found in single cells that appeared as minor variants at the population level.

Combe et al. studied the genetic diversity of vesicular stomatitis virus (VSV), a RNA virus from the *Rhabdoviridae* family, by combining single-cell and ultra-deep sequencing (Combe et al., 2015). Baby Hamster Kidney (BHK)-21 cells were infected with VSV for 45 min before being separated by micromanipulation and further incubated for 24 h to allow completion of viral infection. Supernatants were collected, used to quantify infectious virion progeny by plaque assay and to investigate viral genome sequence variation by deep sequencing. Full-length sequences from 881 viral plaques derived from 90 individual cells and corresponding to three viral generations were compared to the initial viral genome input: in total, 532 single nucleotide polymorphisms were identified, 36 polymorphisms were originally present in the viral stock and 496 newly arising polymorphisms were due to spontaneous mutations, suggesting that new sequence variants emerge rapidly. The authors exploited the pre-existing genetic diversity of viruses to track viral infection at single-cell level. They showed that in most cases single cells were infected by more than one sequence variant of the virus even at low multiplicity of infection. Furthermore, the viral progeny showed variable rates of spontaneous mutation between single cells.

Based on a previous SCS assay (Josefsson et al., 2011), Josefsson et al. investigated the HIV genetic composition in individual memory and naïve CD4<sup>+</sup> T cells from lymph node tissue and peripheral blood. Single cells were isolated from patients by FACS sorting followed by PCR amplification of HIV DNA and sequencing (Josefsson et al., 2013). They showed that the majority of infected CD4<sup>+</sup> T-cells (90%) located in lymph node tissue and peripheral blood contained only one HIV-1 DNA molecule. Furthermore, when they compared the viral sequences from the population of cells of memory and naïve T-cells from lymph node tissue and peripheral blood, obtained by the SCS assay with the HIV-1 RNA sequences from the populations of cells found in plasma, they observed that viral sequences were similar to each other and to HIV-1 RNA from contemporaneous plasma from these patients, implying an exchange of virus and/or infected cells between these compartments in untreated chronic infection.

### 3.3. Cellular heterogeneity to virus infection analysis by SCS

#### 3.3.1. DNA sequencing

Some studies have used DNA-SCS to investigate the heterogeneity of immune responses in infected single cells. Tanaka et al. studied T-cell receptor (TCR) repertoire of human T-cell lymphotropic virus type 1 (HTLV-1)-specific cytotoxic T lymphocytes (Tanaka et al., 2010). HTLV-1 is the major causative agent of adult T-cell leukemia (ATL), an aggressive lymphoproliferative malignancy, which has been associated with specific HLA alleles. Allo-hematopoietic stem cell transplantation has proved to be an effective treatment for adult T-cell leukemia. Tanaka et al. used single-cell approach to study the TCR repertoire of cytotoxic T cells of patients with specific HLA haplotypes undergoing or

not allo-hematopoietic stem cell transplantation. They performed single-cell isolation of T cells by FACS followed by RT-PCR, and direct sequencing of the V-D-J CDR3 region of TCR. They used a total of 160 cells from two patients before treatment and 282 cells from three patients after transplantation. Results showed highly restricted oligoclonal diversity of cytotoxic T cells in bone marrow and peripheral blood at single-cell level before and after allo-transplantation. They also observed that the amino acid motif P-D/P-R of CDR3 was conserved between unrelated ATL patients and in the same patient before and after transplant.

Tsioris et al. used single-cell analysis to study the humoral response against West Nile Virus (WNV) in individuals recently infected with WNV (Tsioris et al., 2015). Antibody secreting single cells and stimulated memory B single cells were identified and isolated by microengraving, and the corresponding antibodies were recovered, cloned and analyzed by next generation sequencing (NGS). Single-cell analysis allowed the identification of WNV-specific antibodies. Moreover, data showed that WNV-specific memory B cells and antibody-secreting cells persisted in post-convalescent individuals and that antibody response was independent of an asymptomatic versus symptomatic disease outcome.

Similarly, Cox et al. isolated dengue-neutralizing antibodies from single cell-FACS sorted human antigen-specific memory B-cell (Cox et al., 2016). Antibody sequencing allowed the characterization of antibody diversity produced by human memory B cells and also led to the identification of a novel epitope for dengue 2 virus serotype.

#### 3.3.2. RNA sequencing

Most transcriptome studies at the single cell level used RNA-Seq technology and investigated cell-to-cell variability in the context of viral infection. Wu et al. studied the cell heterogeneity of a human papilloma virus (HPV)-infected cell line derived from a cervical cancer (HeLa S3 cells) by single-cell RNA-seq (Wu et al., 2015). They developed a customized full-length RNA sequencing pipeline, mRNA amplification and library construction system (MIRALCS followed by modified SMART-seq2), which is based on a high throughput platform where single-cell RNA and subsequent cDNA were prepared at nanoliter scale based on a customized microwell chip. Using this method, they amplified full-length transcripts of 669 single HeLa S3 cells and selected 40 random individual cells for single-cell RNA sequencing. They identified significant differences in the number of total transcripts among single cells (n = 67,000–233,000). Cellular heterogeneity was observed at the level of expression, alternative splicing, and at RNA chimeric transcripts (originated from fusion genes or two different genes by subsequent trans-splicing, and translated into chimeric proteins). Moreover, they identified a set of genes, which are potential interactors with or regulated by E6 and E7 oncogenes.

A single-cell RNA-seq approach was used with a dedicated bioinformatic pipeline (Juliá et al., 2015) to investigate cellular heterogeneity and identify biomarkers of HIV permissiveness (Ciuffi et al., 2015; Rato et al., 2016). Non-infected activated CD4<sup>+</sup> T cells from a highly and a poorly susceptible individual were selected and used for single-cell RNA-seq analysis (SMARTseq) using Fluidigm C1 technology. RNA-seq profiles from 85 highly permissive and 81 poorly permissive single cells were successfully obtained, with ~25 million reads per single cell. Transcriptional heterogeneity translated in a continuum of intermediary cell states in both highly and poorly permissive donor cells, which was mainly driven by TCR-mediated cell activation. Genes whose expression was differential were further investigated as candidate biomarkers of HIV permissiveness.

Martin-Gayo et al. studied at the single-cell level a subset of dendritic cells (DCs) from Elite Controller patients (EC) previ-

ously described to have a more efficient immune response against HIV-1 infection (Martin-Gayo et al., 2015). DCs were first challenged with HIV-1 before being isolated by FACS and processed to single-cell RNA-seq. The whole transcriptome analysis of 85 cells demonstrated different transcriptional patterns on the expression of interferon stimulated genes, cytokines, cytokine receptors and co-stimulatory molecules. Furthermore, from single-cell analyses the authors were able to identify specific markers that characterize a highly functional subset of DCs, with improved abilities to induce T cell proliferation (Martin-Gayo et al., 2016).

Single-cell analysis was also used to study the very recent outbreak of Zika virus (ZIKV). ZIKV is a mosquito-borne flavivirus. It is known to enter human skin fibroblasts, keratinocytes, and immature dendritic cells (Hamel et al., 2015). Very recently, reports of infants born with microcephaly from infected mothers have markedly increased, establishing a pathological link between ZIKV and the central nervous system. Nowakowski et al. used single-cell RNA-Seq (SMART-seq) to identify cell populations that could be susceptible to ZIKV (Nowakowski et al., 2016). The authors studied the expression of different cell receptors that can be used by ZIKV across diverse cell types of the developing brain. Using Fluidigm technology, they perform single-cell RNA-seq on these different cell types and observed that the receptor AXL, known to mediate ZIKV and dengue virus entry in human skin, was also highly expressed in human radial glial cells, astrocytes, endothelial cells, and microglia in developing human cortex and by progenitor cells in developing retina. These results suggested that all these cell types may be particularly vulnerable to ZIKV infection.

#### 4. Conclusions and perspectives

Single-cell technologies have the potential to change the experimental approach to the study of viral infection. Specifically, single-cell sequencing allows examining the impact of cell-to-cell variability on the outcome of viral infection. The analysis can capture viral diversity and evolution to identify sequence variation in viral quasispecies. The approach can also examine cellular heterogeneity through the changes of cell transcriptome. Additional applications can explore the immune response to viral infection in the infected cell.

The current challenge in the field is the joint study of viral and cellular heterogeneity. Ideally this would need the ability to analyze the same single cell before and after viral infection, as the cellular state and gene expression evolves in response to the incoming virus. This would also support the identification of the effect of cellular heterogeneity on the outcome of infection. Future improvements of single-cell technologies, from isolation to biological sampling and analysis, should aim at a more comprehensive snapshot of the single-cell composition.

Ultimately, single-cell-based technologies could be used as a “personalized” tool to identify individual cells in a cell population, in the same way as personalized medicine deals with different individuals. Indeed, personalized medicine originated from the initial identification of outlier individuals, with a higher or lower response to a defined drug treatment for example, as compared to the average individual population. A better characterization of these outlier individuals, at the level of genotyping, transcriptome or proteome profiling, provided the foundation for understanding individual specificities leading to the observed differential response to drug treatment, including antivirals, thereby allowing refinement of the drug treatment and thus personalized medicine. Similarly, SCS is setting the ground for understanding the cellular individual response, by characterizing individual cells at the level of DNA, RNA and protein signatures.

Finally, novel technologies aiming at a controlled and specific treatment of individual cells, followed by SCS may help identifying cellular features characterizing the quality of the individual cellular response to a given stimulus, drug or virus exposure. This should prove useful for future clinical applications, including personalized medicine.

#### Author contributions

SR, MG, AT and AC wrote the manuscript.

#### Conflicts of interest

The authors declare no conflict of interest.

#### Acknowledgments

This work was supported by the European Union’s Seventh Framework Program FP7/2007-2013/under grant agreement no. 305762/Hit Hidden HIV, by the Swiss National Science Foundation grant no. 166412 and by the Novartis Foundation for medical-biological Research grant no. 16B115.

#### References

- Altschuler, S.J., Wu, L.F., 2010. Cellular heterogeneity: do differences make a difference? *Cell* 141 (4), 559–563.
- Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11 (10), 1–12.
- Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31 (2), 166–169.
- S. Andrews, Fast QC A Quality Control tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bacher, R., Kendziorski, C., 2016. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* 17 (1), 1–14.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshtkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-Cell sequencing. *J. Comput. Biol.* 19 (5), 455–477.
- Battich, N., Stoeger, T., Pelkmans, L., 2015. Control of transcript variability in single mammalian cells. *Cell* 163 (7), 1596–1610.
- Bieniasz, P.D., Weiss, R.A., McClure, M.O., 1995. Cell cycle dependence of foamy retrovirus infection. *J. Virol.* 69 (11), 7295–7299.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120.
- Brauns, T.C., Goos, M., 2005. Micromanipulation of single cells from tissue imprints is an alternative to laser-assisted microdissection. *J. Cutan. Pathol.* 32 (7), 461–466.
- Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* 10.
- Brouzes, E., Medkova, M., Savenelli, N., Marran, D., Twardowski, M., Hutchison, J.B., Rothberg, J.M., Link, D.R., Perrimon, N., Samuels, M.L., 2009. Droplet microfluidic technology for single-cell high-throughput screening. *Proc. Natl. Acad. Sci. U. S. A.* 106 (34), 14195–14200.
- Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., 2015. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33.
- Cheng, J., Vanneste, E., Konings, P., Voet, T., Vermeesch, J.R., Moreau, Y., 2011. Single-cell copy number variation detection. *Genome Biol.* 12 (8), 1–14.
- Ciuffi, A., Quenneville, S., Rausell, A., Rato, S., Muñoz, M., Telenti, A., 2015. Single-cell analysis identifies biomarkers for HIV permissiveness Abstract OP2.1), Seventh International Workshop on HIV Persistence during Therapy. Vol. 1. Session 2: Basic science of HIV latency vols. Journal of Virus Eradication, Miami (FL), USA, pp. 1–18.
- Ciuffi, A., Rato, S., Telenti, A., 2016. Single-Cell genomics for virology. *Viruses* 8 (5).
- Ciuffi, A., 2016. The benefits of integration. *Clin. Microbiol. Infect.* 22 (4), 324–332.
- Combe, M., Garijo, R., Geller, R., Cuevas, J.M., Sanjuan, R., 2015. Single-cell analysis of RNA virus infection identifies multiple genetically diverse viral genomes within single infectious units. *Cell Host Microbe* 18 (4), 424–432.
- Cox, K.S., Tang, A., Chen, Z., Horton, M.S., Yan, H., Wang, X.M., Dubey, S.A., DiStefano, D.J., Ettenger, A., Fong, R.H., Doranz, B.J., Casimiro, D.R., Vora, K.A., 2016. Rapid isolation of dengue-neutralizing antibodies from single cell-sorted human antigen-specific memory B-cell cultures. *mAbs* 8 (1), 129–140.

- Dean, F.B., Nelson, J.R., Giesler, T.L., Lasken, R.S., 2001. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* 11 (6), 1095–1099.
- Demeulemeester, J., De Rijck, J., Gijssbers, R., Debyser, Z., 2015. Retroviral integration: site matters: mechanisms and consequences of retroviral integration site selection. *Bioessays* 37 (11), 1202–1214.
- de Bourcy, C.F.A., De Vlaminck, I., Kanbar, J.N., Wang, J., Gawad, C., Quake, S.R., 2014. A quantitative comparison of single-cell whole genome amplification methods. *PLoS One* 9 (8), e105585.
- Deng, Q., Ramskold, D., Reinius, B., Sandberg, R., 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343 (6167), 193–196.
- Dey, S.S., Kester, L., Spanjaard, B., Bienko, M., van Oudenaarden, A., 2015. Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* 33 (3), 285–289.
- Ding, B., Zheng, L., Zhu, Y., Li, N., Jia, H., Ai, R., Wildberg, A., Wang, W., 2015. Normalization and noise reduction for single cell RNA-seq experiments. *Bioinformatics* 31 (13), 2225–2227.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*.
- Doitsh, G., Greene, W.C., 2016. Dissecting how CD4T cells are lost during HIV infection. *Cell Host Microbe* 19 (3), 280–291.
- Fan, H.C., Fu, G.K., Fodor, S.P., 2015. Expression profiling: combinatorial labeling of single cells for gene expression cytometry. *Science* 347 (6222), 1258367.
- Fan, J., Salathia, N., Liu, R., Kaeser, G.E., Yung, Y.C., Herman, J.L., Kaper, F., Fan, J.-B., Zhang, K., Chun, J., Kharchenko, P.V., 2016. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* 13 (3), 241–244.
- Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., Linsley, P.S., Gottardo, R., 2015. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* 16 (1), 1–13.
- Frost, A.R., Eltoum, I.E., Siegal, G.P., Emmert-Buck, M.R., Tangrea, M.A., 2015. Laser microdissection. *Curr. Protoc. Mol. Biol.* 112 (25A 21), 21–30.
- Furlan-Magaril, M., Varnai, C., Nagano, T., Fraser, P., 2015. 3D genome architecture from populations to single cells. *Curr. Opin. Genet. Dev.* 31, 36–41.
- Gawad, C., Koh, W., Quake, S.R., 2016. Single-cell genome sequencing: current state of the science. *Nature reviews. Genetics* 17 (3), 175–188.
- Goetz, J.J., Trimarchi, J.M., 2012. Transcriptome sequencing of single cells with Smart-Seq. *Nat. Biotechnol.* 30 (8), 763–765.
- Grün, D., van Oudenaarden, A., 2015. Design and analysis of single-cell sequencing experiments. *Cell* 163 (4), 799–810.
- Grün, D., Kester, L., van Oudenaarden, A., 2014. Validation of noise models for single-cell transcriptomics. *Nat. Methods* 11.
- Gross, A., Schoendube, J., Zimmermann, S., Steeb, M., Zengerle, R., Koltay, P., 2015. Technologies for single-cell isolation. *Int. J. Mol. Sci.* 16 (8), 16897–16919.
- Grun, D., van Oudenaarden, A., 2015. Design and analysis of single-cell sequencing experiments. *Cell* 163 (4), 799–810.
- Grun, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., van Oudenaarden, A., 2015. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525 (7568), 251–255.
- Hamel, R., Dejarnac, O., Wichit, S., Ekchariyawat, P., Neyret, A., Luplertlop, N., Perera-Lecoin, M., Surasombatpattana, P., Talignani, L., Thomas, F., Cao-Lormeau, V.M., Choumet, V., Briant, L., Despres, P., Amara, A., Yssel, H., Misse, D., 2015. Biology of zika virus infection in human skin cells. *J. Virol.* 89 (17), 8880–8896.
- Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., Dor, Y., Regev, A., Yanai, I., 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* 17, 77.
- Hashimshony, T., Wagner, F., Sher, N., Yanai, I., 2012. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2 (3), 666–673.
- Hodne, K., Weltzien, F.A., 2015. Single-cell isolation and gene analysis: pitfalls and possibilities. *Int. J. Mol. Sci.* 16 (11), 26832–26849.
- Hoeve, M.A., Nash, A.A., Jackson, D., Randall, R.E., Dransfield, I., 2012. Influenza virus A infection of human monocyte and macrophage subpopulations reveals increased susceptibility associated with cell differentiation. *PLoS One* 7 (1), e29443.
- Holmes, M., Zhang, F., Bieniasz, P.D., 2015. Single-cell and single-cycle analysis of HIV-1 replication. *PLoS Pathog.* 11 (6), e1004961.
- Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y., Peng, J., 2016. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* 26 (3), 304–319.
- Huang, L., Ma, F., Chapman, A., Lu, S., Xie, X.S., 2015. Single-cell whole-genome amplification and sequencing: methodology and applications. *Annu. Rev. Genomics Hum. Genet.* 16, 79–102.
- Huang, S., 2009. Non-genetic heterogeneity of cells in development: more than just noise. *Development* 136 (23), 3853–3862.
- Hummer, D., Kurth, F., Naredi-Rainer, N., Dittich, P.S., 2016. Single cells in confined volumes: microchambers and microdroplets. *Lab Chip* 16 (3), 447–458.
- Humphries, E.H., Glover, C., Reichmann, M.E., 1981. Rous sarcoma virus infection of synchronized cells establishes provirus integration during S-phase DNA synthesis prior to cellular division. *Proc. Natl. Acad. Sci. U. S. A.* 78 (4), 2601–2605.
- Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., Teichmann, S.A., 2016. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* 17 (1), 1–15.
- Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J.B., Lonnerberg, P., Linnarsson, S., 2011. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* 21 (7), 1160–1167.
- Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J.B., Lonnerberg, P., Linnarsson, S., 2012. Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat. Protoc.* 7 (5), 813–828.
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lonnerberg, P., Linnarsson, S., 2014. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11 (2), 163–166.
- Jakobsen, M.R., Olgner, D., Hiscott, J., 2015. Innate immune sensing of HIV-1 infection. *Curr. Opin. HIV AIDS* 10 (2), 96–102.
- Josefsson, L., King, M.S., Makitalo, B., Brannstrom, J., Shao, W., Maldarelli, F., Kearney, M.F., Hu, W.S., Chen, J., Gaines, H., Mellors, J.W., Albert, J., Coffin, J.M., Palmer, S.E., 2011. Majority of CD4+ T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. *Proc. Natl. Acad. Sci. U. S. A.* 108 (27), 11199–11204.
- Josefsson, L., Palmer, S., Faria, N.R., Lemey, P., Casazza, J., Ambrozak, D., Kearney, M., Shao, W., Kottlil, S., Sneller, M., Mellors, J., Coffin, J.M., Maldarelli, F., 2013. Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4+ T-cells contain one HIV-1 DNA molecule. *PLoS Pathog.* 9 (6), e1003432.
- Juliá, M., Telenti, A., Raussell, A., 2015. Sincell: an R/Bioconductor package for statistical assessment of cell-state hierarchies from single-cell RNA-seq. *Bioinformatics* 31 (20), 3380–3382.
- Kane, M., Yadav, S.S., Bitzegeio, J., Kutluay, S.B., Zang, T., Wilson, S.J., Schoggins, J.W., Rice, C.M., Yamashita, M., Hatzioannou, T., Bieniasz, P.D., 2013. MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* 502 (7472), 563–566.
- Katayama, S., Töhönen, V., Linnarsson, S., Kere, J., 2013. SAMstr: statistical test for differential expression in single-cell transcriptome with spike-in normalization. *Bioinformatics* 29 (22), 2943–2945.
- Katz, R.A., Greger, J.G., Skalka, A.M., 2005. Effects of cell cycle status on early events in retroviral replication. *J. Cell. Biochem.* 94 (5), 880–889.
- Kharchenko, P.V., Silberstein, L., Scadden, D.T., 2014. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* 11.
- Kim, J.K., Marioni, J.C., 2013. Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol.* 14 (1), R7.
- Kim, D., Perte, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14 (4), R36.
- Kim, J.K., Kolodziejczyk, A.A., Ilicic, T., Teichmann, S.A., Marioni, J.C., 2015a. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat. Commun.* 6, 8687.
- Kim, K.T., Lee, H.W., Lee, H.O., Kim, S.C., Seo, Y.J., Chung, W., Eum, H.H., Nam, D.H., Kim, J., Joo, K.M., Park, W.Y., 2015b. Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. *Genome Biol.* 16, 127.
- Kluge, S.F., Sauter, D., Kirchhoff, F., 2015. SnapShot: antiviral restriction factors. *Cell* 163 (3), 774–774 e771.
- Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., Teichmann, S.A., 2015. The technology and biology of single-cell RNA sequencing. *Mol. Cell* 58 (4), 610–620.
- Kroneis, T., Ye, J., Gillespie, K., 2016. Laser capture and single cell genotyping from frozen tissue sections. *Methods Mol. Biol.* 1433, 159–167.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9 (4), 357–359.
- Leng, N., Chu, L.-F., Barry, C., Li, Y., Choi, J., Li, X., Jiang, P., Stewart, R.M., Thomson, J.A., Kendzierski, C., 2015. Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments. *Nat. Methods* 12 (10), 947–950.
- Li, D., Wang, Z., Ren, L., Zhang, J., Feng, G., Hong, K., Hao, Y., Qi, Z., Liang, H., Shao, Y., 2016. Study of antibody repertoires to the CD4 binding site of gp120 of a Chinese HIV-1-infected elite neutralizer, using 454 sequencing and single-cell sorting. *Arch. Virol.* 161 (4), 789–799.
- Lun, L., Bach, A.T., Marioni, K., J.C., 2016. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* 17 (1), 1–14.
- Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., Smith, M., Van der Aa, N., Banerjee, R., Ellis, P.D., Quail, M.A., Swerdlow, H.P., Zernicka-Goetz, M., Livesey, F.J., Ponting, C.P., Voet, T., 2015. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* 12 (6), 519–522.
- Marco, E., Karp, R.L., Guo, G., Robson, P., Hart, A.H., Trippa, L., Yuan, G.-C., 2014. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl. Acad. Sci.* 111 (52), E5643–E5650.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17 (1), Next Generation Sequencing Data Analysis.
- Martin-Gayo, E., Buzon, M.J., Ouyang, Z., Hickman, T., Cronin, J., Pimenova, D., Walker, B.D., Lichtenfeld, M., Yu, X.G., 2015. Potent cell-intrinsic immune responses in dendritic cells facilitate HIV-1-specific t cell immunity in HIV-1 elite controllers. *PLoS Pathog.* 11 (6), e1004930.

- Martin-Gayo, E., Cole, M., Kolb, K.E., Ouyang, Z., Kazer, S.W., Walker, B.D., Yosef, N., Shalek, A.K., Yu, X.G., 2016. Identification of a highly functional DC subset in controllers by single-cell RNA-Seq (Abstract 19). In: Conference on Retroviruses and Opportunistic Infections, Boston.
- Mazutis, L., Gilbert, J., Ung, W.L., Weitz, D.A., Griffiths, A.D., Heyman, J.A., 2013. Single-cell analysis and sorting using droplet-based microfluidics. *Nat. Protoc.* 8 (5), 870–891.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303.
- McWilliam Leitch, E.C., McLauchlan, J., 2013. Determining the cellular diversity of hepatitis C virus quasiespecies by single-cell viral sequencing. *J. Virol.* 87 (23), 12648–12655.
- Min, J.W., Kim, W.J., Han, J.A., Jung, Y.J., Kim, K.T., Park, W.Y., Lee, H.O., Choi, S.S., 2015. Identification of distinct tumor subpopulations in lung adenocarcinoma via single-cell RNA-seq. *PLoS One* 10 (8), e0135817.
- Mohammadi, P., Desfarges, S., Bartha, I., Joos, B., Zangger, N., Munoz, M., Gunthard, H.F., Beerenwinkel, N., Telenti, A., Ciuffi, A., 2013. 24 hours in the life of HIV-1 in a T cell line. *PLoS Pathog.* 9 (1), e1003161.
- Nagano, T., Lubling, Y., Yaffe, E., Winget, S.W., Dean, W., Tanay, A., Fraser, P., 2015. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat. Protoc.* 10 (12), 1986–2003.
- Neil, S.J., Zang, T., Bieniasz, P.D., 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* 451 (7177), 425–430.
- Nikolenko, S.I., Korobeynikov, A.I., Alekseyev, M.A., 2013. BayesHammer: bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* 14 (1), 1–11.
- Nowakowski, T.J., Pollen, A.A., Di Lullo, E., Sandoval-Espinosa, C., Bershteyn, M., Kriegstein, A.R., 2016. Expression analysis highlights AXL as a candidate zika virus entry receptor in neural stem cells. *Cell Stem Cell* 18 (5), 591–596.
- Ntranos, V., Kamath, G.M., Zhang, J., Pachter, L., Tse, D.N., 2016. Fast and accurate single-cell RNA-Seq analysis by clustering of transcript-compatibility counts. *bioRxiv*.
- Pan, X., Baldauf, H.M., Keppler, O.T., Fackler, O.T., 2013. Restrictions to HIV-1 replication in resting CD4+ T lymphocytes. *Cell Res.* 23 (7), 876–885.
- Patton, G.S., Erlwein, O., McClure, M.O., 2004. Cell-cycle dependence of foamy virus vectors. *J. Gen. Virol.* 85 (Pt 10), 2925–2930.
- Peng, C., Greenwell-Wild, T., Nares, S., Jin, W., Lei, K.J., Rangel, Z.G., Munson, P.J., Wahl, S.M., 2007. Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression. *Blood* 110 (1), 393–400.
- Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28 (11), 1420–1428.
- Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., Sandberg, R., 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10 (11), 1096–1098.
- Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., Sandberg, R., 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9 (1), 171–181.
- Picelli, S., 2016. Single-cell RNA-sequencing: the future of genome biology is now. *RNA Biol.* 1–14.
- Pierson, E., Yau, C., 2015. ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome Biol.* 16 (1), 241.
- Proserpio, V., Piccolo, A., Haim-Vilmovsky, L., Kar, G., Lonnberg, T., Svensson, V., Pramanik, J., Natarajan, K.N., Zhai, W., Zhang, X., Donati, G., Kayikci, M., Kotar, J., McKenzie, A.N., Montandon, R., Billker, O., Woodhouse, S., Cicuta, P., Nicodemí, M., Teichmann, S.A., 2016. Single-cell analysis of CD4+ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation. *Genome Biol.* 17, 103.
- Qian, M., Wang, D.C., Chen, H., Cheng, Y., 2016. Detection of single cell heterogeneity in cancer. *Semin. Cell Dev. Biol.*
- Ramskold, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtkova, I., Loring, J.F., Laurent, L.C., Schroth, G.P., Sandberg, R., 2012. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* 30 (8), 777–782.
- Rato, S., Rausell, A., Munoz, M., Telenti, A., Ciuffi, A., 2016. Proceedings of the frontiers of retrovirology conference 2016. *Retrovirology* 13 (1), 68.
- Rausell, A., Munoz, M., Martinez, R., Roger, T., Telenti, A., Ciuffi, A., 2016. Innate immune defects in HIV permissive cell lines. *Retrovirology* 13 (1), 43.
- Reece, A., Xia, B., Jiang, Z., Noren, B., McBride, R., Oakey, J., 2016. Microfluidic techniques for high throughput single cell analysis. *Curr. Opin. Biotechnol.* 40, 90–96.
- Safonova, Y., Bankevich, A., Pevzner, P.A., 2015. dipSPAdes: assembler for highly polymorphic diploid genomes. *J. Comput. Biol.* 22 (6), 528–545.
- Saliba, A.E., Westermann, A.J., Gorski, S.A., Vogel, J., 2014. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* 42 (14), 8845–8860.
- Satija, R., Shalek, A.K., 2014. Heterogeneity in immune responses: from populations to single cells. *Trends Immunol.* 35 (5), 219–229.
- Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27 (6), 863–864.
- Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., Trombetta, J.J., Gennert, D., Gnirke, A., Goren, A., Hacohen, N., Levin, J.Z., Park, H., Reggev, A., 2013. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498 (7453), 236–240.
- Sheehy, A.M., Gaddis, N.C., Choi, J.D., Malim, M.H., 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418 (6898), 646–650.
- Shin, J., Berg Daniel, A., Zhu, Y., Shin Joseph, Y., Song, J., Bonaguidi Michael, A., Enikolopov, G., Nauen David, W., Christian Kimberly, M., Ming, G.-l., Song, H., 2015. Single-cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* 17 (3), 360–372.
- Spits, C., Le Caignec, C., De Ryck, M., Van Haute, L., Van Steirteghem, A., Liebaers, I., Sermon, K., 2006. Whole-genome multiple displacement amplification from single cells. *Nat. Protoc.* 1 (4), 1965–1970.
- Stegle, O., Teichmann, S.A., Marioni, J.C., 2015. Computational and analytical challenges in single-cell transcriptomics. *Nature reviews. Genetics* 16 (3), 133–145.
- Stoeger, T., Battich, N., Pelkmans, L., 2016. Passive noise filtering by cellular compartmentalization. *Cell* 164 (6), 1151–1161.
- Stubbington, M.J., Lonnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., Teichmann, S.A., 2016. T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* 13 (4), 329–332.
- Suzuki, Y., Craigie, R., 2007. The road to chromatin – nuclear entry of retroviruses. *Nat. Rev. Microbiol.* 5 (3), 187–196.
- Tanaka, Y., Nakasone, H., Yamazaki, R., Sato, K., Sato, M., Terasako, K., Kimura, S., Okuda, S., Kako, S., Oshima, K., Tanihara, A., Nishida, J., Yoshikawa, T., Nakatsura, T., Sugiyama, H., Kanda, Y., 2010. Single-cell analysis of T-cell receptor repertoire of HTLV-1 Tax-specific cytotoxic T cells in allogeneic transplant recipients with adult T-cell leukemia/lymphoma. *Cancer Res.* 70 (15), 6181–6192.
- Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjold, M., Ponder, B.A., Tunnacliffe, A., 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13 (3), 718–725.
- Trappell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., Rinn, J.L., 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32 (4), 381–386.
- Treutlein, B., Brwnfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., Quake, S.R., 2014. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509 (7500), 371–375.
- Tsioris, K., Gupta, N.T., Ogunniyi, A.O., Zimmisky, R.M., Qian, F., Yao, Y., Wang, X., Stern, J.N., Chari, R., Briggs, A.W., Clouser, C.R., Vigneault, F., Church, G.M., Garcia, M.N., Murray, K.O., Montgomery, R.R., Kleinstein, S.H., Love, J.C., 2015. Neutralizing antibodies against West Nile virus identified directly from human B cells by single-cell analysis and next generation sequencing. *Integr. Biol.* 7 (12), 1587–1597.
- Tuttle, D.L., Harrison, J.K., Anders, C., Sleasman, J.W., Goodenow, M.M., 1998. Expression of CCR5 increases during monocyte differentiation and directly mediates macrophage susceptibility to infection by human immunodeficiency virus type 1. *J. Virol.* 72 (6), 4962–4969.
- Vallejos, C.A., Marioni, J.C., Richardson, S., 2015. BASiCS: Bayesian analysis of single-cell sequencing data. *PLoS Comput. Biol.* 11.
- Vallejos, C.A., Richardson, S., Marioni, J.C., 2016. Beyond comparisons of means: understanding changes in gene expression at the single-cell level. *Genome Biol.* 17 (1), 1–14.
- Vieira Braga, F.A., Teichmann, S.A., Chen, X., 2016. Genetics and immunity in the era of single-cell genomics. *Hum. Mol. Genet.* 25 (R2), R141–R148.
- Wang, Y., Navin, N.E., 2015. Advances and applications of single-cell sequencing technologies. *Mol. Cell* 58 (4), 598–609.
- Wang, X., Ye, L., Hou, W., Zhou, Y., Wang, Y.J., Metzger, D.S., Ho, W.Z., 2009. Cellular microRNA expression correlates with susceptibility of monocytes/macrophages to HIV-1 infection. *Blood* 113 (3), 671–674.
- Welch, J.D., Hu, Y., Prins, J.F., 2016. Robust detection of alternative splicing in a population of single cells. *Nucleic Acids Res.*
- Wen, L., Tang, F., 2016. Single-cell sequencing in stem cell biology. *Genome Biol.* 17 (1), 71.
- Wheeler, A.R., Thronset, W.R., Whelan, R.J., Leach, A.M., Zare, R.N., Liao, Y.H., Farrell, K., Manger, I.D., Daridon, A., 2003. Microfluidic device for single-cell analysis. *Anal. Chem.* 75 (14), 3581–3586.
- Wu, L., Zhang, X., Zhao, Z., Wang, L., Li, B., Li, G., Dean, M., Yu, Q., Wang, Y., Lin, X., Rao, W., Mei, Z., Li, Y., Jiang, R., Yang, H., Li, F., Xie, G., Xu, L., Wu, K., Zhang, J., Chen, J., Wang, T., Kristiansen, K., Zhang, X., Li, Y., Yang, H., Wang, J., Hou, Y., Xu, X., 2015. Full-length single-cell RNA-seq applied to a viral human cancer: applications to HPV expression and splicing analysis in HeLa S3 cells. *GigaScience* 4, 51.
- Xu, C., Su, Z., 2015. Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics* 31 (12), 1974–1980.
- Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., Huang, J., Li, M., Wu, X., Wen, L., Lao, K., Li, R., Qiao, J., Tang, F., 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* 20 (9), 1131–1139.
- Ye, F., Xin, Z., Han, W., Fan, J., Yin, B., Wu, S., Yang, W., Yuan, J., Qiang, B., Sun, W., Peng, X., 2015. Quantitative proteomics analysis of the hepatitis C virus replicon high-permissive and low-permissive cell lines. *PLoS One* 10 (11), e0142082.
- Zafar, H., Wang, Y., Nakhleh, L., Navin, N., Chen, K., 2016. Monovar: single-nucleotide variant detection in single cells. *Nat. Methods* 13 (6), 505–507.

- Zeisel, A., Machado, A.B.M., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., Rolny, C., Castelo-Branco, G., Hjerling-Leffler, J., Linnarsson, S., 2015. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*.
- Zong, C., Lu, S., Chapman, A.R., Xie, X.S., 2012. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 338 (6114), 1622–1626.
- Zhang, C.-Z., Adalsteinsson, V.A., Francis, J., Cornils, H., Jung, J., Maire, C., Ligon, K.L., Meyerson, M., Love, J.C., 2015. Calibrating genomic and allelic coverage bias in single-cell sequencing. *Nat. Commun.* 6, 6822.
- žuraskienė, J., Yau, C., 2016. *pcaReduce*: hierarchical clustering of single cell transcriptional profiles. *BMC Bioinf.* 17 (1), 140.