Lab on a Chip

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: R. Salomon, D. C. Kaczorowski, F. Valdes-Mora, R. Nordon, A. Neild, N. Farbehi, N. Bartonicek and D. Gallego Ortega, *Lab Chip*, 2019, DOI: 10.1039/C8LC01239C.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/loc

Droplet-based single cell RNAseq tools: a practical guide

View Article Online
DOI: 10.1039/C8LC01239C

Robert Salomon[a,b] *, Dominik Kaczorowski[a], Fatima Valdes-Mora[c,d], Robert E. Nordon[e,f], Adrian Neild[g], Nona Farbehi[a], Nenad Bartonicek[d,h] and David Gallego-Ortega[d,h] *

AFFILIATIONS
a. Garvan-Weizmann Centre for Cellular Genomics. Garvan Institute of Medical Research. Sydney, NSW. Australia.
b. Current address: Institute for Biomedical Materials and Devices, University of Technology Sydney, Ultimo 2007 NSW. Australia.
c. Histone Variants Lab, Genomics and Epigenetics Division. Garvan Institute of Medical Research. Sydney, NSW. Australia.
d. St Vincent's School of Medicine. Faculty of Medicine. University of New South Wales Sydney. Sydney, NSW Australia.
e. Graduate School of Biomedical Engineering, University of New South Wales Sydney, Sydney NSW Australia.
f. Australian Research Council Special Research Initiative in Stem Cell Science – Stem Cells, Australia
g. Department of Mechanical and Aerospace Engineering. Monash University. Clayton, VIC Australia.
h. Tumour Development, The Kinghorn Cancer Centre. Garvan Institute of Medical Research. Sydney NSW Australia.

* Corresponding authors: robert.salomon@uts.edu.au, d.gallego@garvan.org.au

Word count: 12,295

Table of contents caption:

A step-by-step guide for droplet-based single cell RNAseq experiments, practical considerations and technical notes.

Abstract

Droplet based scRNA-seq systems such as Drop-seq, inDrop and Chromium 10X have been the catalyst for the wide adoption of high-throughput scRNA-seq technologies in the research laboratory. In order to understand the capabilities of these systems to deeply interrogate biology; here we provide a practical guide through all the steps involved in a typical scRNA-seq experiment. Through comparing and contrasting these three main droplet based systems (and their derivatives), we provide an overview of all critical considerations in obtaining high quality and biologically relevant data. We also discuss the limitations of these systems and how they fit into the emerging field of Genomic Cytometry.

View Article Online
DOI: 10.1039/C8LC01239C

Introduction

Cells are the smallest functional unit in the body. They are the base of all life and seed of all disease. The human body consists of approximately 37 trillion cells (1) and yet it starts as a single cell at the time of conception. Just as a single cell can give rise to a whole organism so too may a single cell give rise to a cancer or a raft of other diseases. It is therefore imperative to develop technologies that can routinely, accurately and simply look at all aspects of the cell. Since the application of microscopy and flow cytometry, we have long been able to look at individual cells but a true understanding has been hampered as these technologies have limited dimensionality and are only capable of looking at a very small fraction of cell properties at a given time.

From the first single-cell digital gene expression profiling method (2, 3) the field of single-cell RNA sequencing (scRNA-seq) has exploded over the last few years and it is now possible to simultaneously analyse thousands of transcripts per cell and thousands of cells per evaluation. All these technologies partition a cell suspension into nanoliter reaction chambers containing individual cells (either in standard SBS plates, micro-wells or in the form of a water-in-oil emulsion) for subsequent multi-parametrical analysis of cell lysate. The development of a one-step cell-specific barcoding strategy (STRT-seq) enabled high multiplexing analysis of pooled samples increasing the throughput and reducing the cost (4). Subsequent technologies followed similar barcoding strategies but reduced the intense labor required by micromanipulation using Fluorescent Activated Cell Sorting (FACS) to deposit single cells into micro-well plates. This improvement together with new molecular strategies crystalised in the development of SMART-seq (5), CEL-seq (6), MARS-seq (7) or SCRB-seq (8) (and their subsequent refined versions SMART-seq2 (9), CEL-seq2 (10) and mcSCRB-seq (11)) to individually barcode and amplify the transcriptome of each cell. The introduction of microfluidic instruments such the Fluidigm C1™ allowed the automation of single cell docking and the molecular processes, reducing the consumption of reagents and improving sensitivity (reviewed in (12). More recently, microfluidics has heavily contributed to improve scRNA-seq further reducing cost and dramatically increasing cellular throughput, both using dynamic microfluidic systems or high-density microwell based. Drop-seq (13), inDrop (14) and GemCode/Chromium 10X™ (15) use microfluidic droplet chemistry with encapsulation of cells and barcoding beads (16). Microfluidic-controlled high-density microwell plates also use barcoding beads as in the droplet-based systems (17), Seq-well (18), Microwell-seq (19) and commercial variants such as BD Rhapsody™.

View Article Online
DOI: 10.1039/C8LC01239C

Functional genomics plays a central role in understanding molecular processes of tissues in homeostatic conditions and during disease. Conventional bulk RNA expression has been limited in the analysis of heterogeneous cell populations as it provides an average of profiles of the cell mixture, with the consequent loss of information on cellular heterogeneity. In these circumstances, information at the single cell level could only be achieved by inferring patterns of expression in combination with other 'omics approaches and with limited resolution (20, 21). The adoption of scRNA-seq methods enabled deep cellular characterization of the RNA profiles one cell at the time, demonstrating a huge advance over traditional flow cytometric analysis which is restricted to the simultaneous analysis of less than 30 measured parameters per cell event even with high-end cytometers such as BD FACSymphony™ A5 and the Cytek™ Aurora. The capacity to sample tens of thousands of cells has fundamentally reshaped our ability to identify novel cell types and states. In sampling a tissue it is important to analyse a large enough number of cells to resolve subsets of cells with very similar properties (22). By producing detailed molecular cellular landscapes, droplet-based scRNA-seq has arisen as the preferred method to unravel cell diversity in tissues (13, 23-26) and enabling deeper understanding of mechanisms of disease (7, 13, 14, 23, 24, 27-31). Some of the current applications for scRNA-seq include:

1) Identification of the cell composition of tissues, including non-abundant and unknown populations, generating a reference cellular atlas (13, 14, 32-34).

2) Understanding how cells develop from stem and progenitor cells by transit through differentiation and activation states or cell polarisation (35-41)

3) Analysis of the clonal evolution of cancer, identifying the cell of origin or clone that results in cancer spread, and development of resistance to chemotherapy (27, 35, 42-45).

4) Understanding cell communication through cell receptor-ligand networks (46).

5) Understanding the cellular response to genetic manipulation (33) or drugs (47) that act on specific molecular pathways, a key step for the development of new treatments.

The process of scRNA-seq from a practical perspective.

Droplet-based scRNA-seq experiments are complex multistep processes which require the preparation of a cell suspension, rapid cell and barcoding bead encapsulation, followed by unbiased transcript barcoding during reverse transcription, amplification of the barcoded material, and library preparation for DNA sequencing. The bioinformatic tools for analysis of the big multidimensional data generated by scRNA-seq is evolving rapidly with development of more standardised pipelines and analysis methods.

The biological question and biological material will dictate the scRNA-seq analysis pathway, however the general workflow of all droplet-based scRNA-seq methods will remain remarkably similar and can be categorized in 5 steps (**Fig. 1**).

(1) **Sample Preparation**. The sample must be dissociated into a single cell suspension unless meaningful biological information is provided by cell-cell interactions. For example, the interaction of immune cells with tumour cells or circulating tumor cells that travel in clusters. The sample is then purified to remove dead cells or to enrich for cells of interest using FACS, MACS or other procedures.

(2) **Encapsulation in droplets.** Following sample preparation, cells are individually encapsulated into a water-in-oil droplet. Platforms differ in their capture efficiency and cost, thus the encapsulation step is one of the main factors to consider when choosing a specific system.

(3) **Barcode incorporation and Molecular Amplification.** Each cell contains minute quantities of mRNA, thus the transcriptional content of each cell must be amplified for sequencing. The two methods used in droplet-based scRNA-seq systems are inherited from the molecular workflows of previous scRNA-seq systems, and are based on in vitro transcription (IVT) or template switching amplification (TSO), reviewed in (12, 48).

(4) **Sequencing**. Regardless of the method of amplification all material undergoes a molecular library preparation, which involves the incorporation of adaptors compatible with next generation sequencing (NGS).

(5) **Data Processing and Informatics**. The vast amounts of data generated by NGS must be then processed. An increasing number of bioinformatic analytical frameworks have been developed to specifically manage, interrogate and interpret scRNA-seq data (49-51).

Small changes in methodology at each step can dramatically affect the outcome of a scRNA-seq experiment, as such, large international consortia like the Human Cell Atlas (32) are currently working to identify optimal conditions for each sample type and to generate large high quality standardised data sets that will serve as a reference. The practical aspects of each of these steps are discussed in detail below.

1. Sample preparation

Cell preparation remains the most challenging task for generating biological representative data. It is critical to perform a cell preparation that produces healthy live cells in a single cell suspension that uniformly samples the heterogeneous population. However it is difficult to identify the true composition of the original sample, it is for this reason we recommend the incorporation of a high dimensional flow cytometry analysis on all sample digests that are processed through scRNA-seq. This is particularly relevant for the preparation

View Article Online
DOI: 10.1039/C8LC01239C

of solid tissues that require enzymatic digestion of the extracellular matrix and disruption of cell-cell junctions as digestion may be incomplete and cells may rapidly lose viability, both of which will lead to biases in the data produced by scRNA-seq experiments. One of the most important consequences of cell death is the release of cellular RNAs into the extracellular 'soup'. These cell-free RNAs will be encapsulated within droplets alongside capture beads and cells and be indistinguishably processed thereby skewing the expression profile of captured cells and confounding the biological interpretation. Despite recent efforts to computationally detect and account for such contamination (52, 53), the core assumption remains that droplet-based systems capture the RNA from a single-cell, so great care should be taken to ensure as best as possible that the manipulation of tissues and cells prior to encapsulation is as gentle as possible. Ideally, including high dimensional and high-throughput immunohistological analysis of the tissue prior to digestion can be considered as a strategy to assess how faithfully the tissue is represented by the scRNA-seq analysis.

Pre- analysis of cells prior to encapsulation

Cell viability and dissociation can be assessed by bright-field microscopy or flow cytometry. In addition to assessing post digestion viability, it is also important to assess cell death under encapsulation conditions. This is particularly relevant when encapsulation times are extended or when cells are sensitive to low protein buffers such as those used in Drop-seq or inDrop workflows.

There are three parameters that determine sample quality for scRNA-seq experiments:

1) The percentage of dead cells: ideally any input into a droplet base scRNA-seq system should have cell viabilities at least 90-95%, although this is highly sample dependent. Cell viability can be generally measured using membrane impermanent DNA-intercalators (such as DAPI or PI) or vital agents (trypan blue).

2) The amount of debris in the sample: cell suspensions from primary samples, especially solid tissues, contain more debris than those from *in vitro* culture or those obtained from cells in circulation.

3) The percentage of doublets inherent to the cell preparation: formed either by incompletely digested cell aggregates or during encapsulation.

Although basic microscopy can assess the three parameters that define sample quality, additional characterization and sample preparation is offered by flow cytometry. Flow cytometry provides quantitative characterization of the sample with respect to viability (DAPI/PI incorporation and cell viability dyes such as CFSE or calcein), cell debris (low scatter signals), the percentage of cells that are not singlets (i.e. doublets and cell clumps via pulse processing – typically of the forward and side scatter signals) and multiparametric cell type composition using fluorescent-labelled monoclonal antibodies. In addition, the high-

View Article Online
DOI: 10.1039/C8LC01239C

throughput capacity of flow cytometry is able to provide a high-degree of sampling confidence and is less subjective than visual inspection of cells.

Initially droplet-based single cell RNAseq required fresh cells to produce high quality transcriptome data. This has limited its applicability on: 1) Samples where the RNA quality is highly time-dependent and are not located in the same laboratory, like clinical samples from multicentre collaborations; 2) Complex experimental designs where there are multiple samples that cannot be run on the same day or there are different timepoints and thus batch effects would profoundly affect the interpretation of the data. 3) Fragile primary cells, where the tissue dissociation process prior to droplet encapsulation highly stresses these cells and therefore it could affect their transcriptome. One possible solution lies in the cryopreservation of the material for subsequent analysis, although freezing could potentially result in RNA degradation. The degradation rate of each freeze-thaw cycle is estimated to account for about 20% RNA loss experimentally determined using synthetic RNA (22), however comparable results have been obtained in multiple studies and differences found between fresh and frozen samples might be due to sample heterogeneity or differential survival of specific cell types to the cryopreservation process (54). Therefore, in the last few years the addition of methanol-based fixation methods prior to tissue dissociation and/or droplet encapsulation have been suggested as an alternative solution (55-57). This method simply fixes the cells or tissues in 80% methanol, which allows sample storage at -20 or -80 °C for several months. Cells are then rehydrated and ready for droplet encapsulation and transcriptome profiling. Fixed primary cells from tissues and cell lines have shown high quality single-cell RNAseq data from Drop-seq (55).and 10X Chromium (56).

Methods to enrich viable target cells.

Cell separation methods are used to enrich viable target cells populations, particularly when samples have a high level of cell death and debris or to purify rare populations of interest. Two of the most common cell separation methods are FACS and MACS:

a. *Fluorescent Activated Cell Sorting (FACS):*

FACS or Fluorescent activated cell sorting is the process by which individual cells are separated from the bulk population based on multiparameter fluorescent and light scatter analysis of the cell population. The sort process can be achieved using electrostatic, physical, piezoelectric or magnetic fluidic sorting (58-60), however in practice the majority of FACS is performed using droplet based electrostatic FACS. The principle of electrostatic FACS relies on sorting droplets of a conductive solution (usually an isotonic Phosphate buffered saline) by charging the conductive stream just prior to droplet breakoff based on the 1) the presence of

View Article Online
DOI: 10.1039/C8LC01239C

a target cell and 2) (generally) the absence of a non-target cell. Under ideal conditions FACS results in the separation of a well-characterized population which is highly purified (>99%) for the cell(s) of interest.

FACS can be used to 1) remove dead cells and debris, 2) remove cell clumps and incompletely digested sample fractions, and 3) enrich or purify target populations defined by the expression of one or more cell surface receptor markers labelled with fluorescent labelled antibodies. There are several studies that used FACS and droplet-based scRNA-seq (61-63). An extended description of FACS sorting for scRNA-seq experiments is found in **Additional information BOX 1**.

b. Magnetic associated Cell Separation (MACS)™

MACS™ (64) is a non-sequential cell separation method where target cells are labelled with paramagnetic iron oxide nanoparticles conjugated to antibodies directed against cell surface receptors and exposed to a high gradient magnetic field. In our experience MACS™ is much gentler than FACS and is particularly useful for cell types, which we have shown to have low post-FACS viability (i.e. tumor samples taken from highly metastatic models of cancer). Although studies in some cell types have shown FACS to affect post-sort cell viability (65), and increase oxidative stress (66), to date there have been no definitive studies that have systematically compared the effect of FACS on scRNA-seq data in a cell type specific manner, thus viability still remains to be empirically determined in each experiment.

MACS™ is easily scaled and while the number of cells that can be purified by FACS is limited by a finite sorting speed (10^3-10^4 cells/second) MACS™ does not require cell to be sequentially analysed and provides highly parallel cell sorting. MACS™ however cannot be used to separate cells based on non-discrete marker expression or when cells are defined by both the presence and absence of markers and need to be sorted in the same run. Therefore, enrichment of rare cells that can be defined by the expression of one or more markers is more time effective using MACS™ compared to FACS.

Where very rare cells that have a complex phenotype are required we would recommend assessing the potential to use MACS™ and FACS in combination. In this workflow a pre-enrichment step would be performed using MACS™ to enrich for the target population (either by positive or negative enrichment) followed by FACS using a more complex panel of fluorescently labeled antibodies. These workflows give us the best of both worlds and allow separation of rare cells with complex phenotypes in the shortest possible time. A comparison of the effect of using FACS or MACS™ isolation in a complex, FACS-sensitive tissue with regards to the cDNA yield using Drop-seq is shown in **Figure 2**.

c. Microfluidic sorting

View Article Online
DOI: 10.1039/C8LC01239C

The downsizing of fluidic channel dimensions allows access to additional sorting mechanisms to the magnetic and electrical fields described above (67). Broadly, microfluidic sorting systems can be divided into passive and active, the former uses hydrodynamic features designed into the chip to achieve sorting, the latter uses an externally applied force field. A further distinction that can be made for active systems is that they can be either continuous or responsive systems, the former applies the same flow or force field to all cells (as does MACS™), and the latter measures a property of the cell and responds accordingly (like FACS).

Passive microfluidic sorting methods design fluid flows such that the particle trajectory through the chip is specific to their physical properties. Two prominent examples are deterministic lateral displacement systems (DLD) and pinch flow fractionation (PFF). In DLD a series of pillars are arranged on a skewed grid, such that particles of a size above a critical value will be bumped by the pillars into a new streamline. At each subsequent row of pillars, the skewed nature of the grid ensures that a further lateral bump occurs such that the streamlines between particle sizes become increasingly divergent, hence sorting is achieved (68-73). In PFF a constriction in the flow field, either a physical narrowing of the channel or caused by the addition of a buffer fluid, restricts the lateral locations in which cells can be located, typically pressing them against a wall. As the centre of a smaller particle can be closer to the wall than a larger particle, the streamline which each particle follows is size dependent. As such, beyond the constriction, when the flow is expanded, cells of different sizes follow distinct trajectories (74-76). Further passive examples include inertia microfluidics (77) in which sorting is achieved by the size dependent balance between lift and Dean drag forces (78), and grooved channels (79).

Active methods for microfluidic cell separation have been developed using a range of mechanisms, including magnetic (80, 81), optical (82), dielectrophoretic (DEP) (74, 83, 84) and acoustic forces (85). Continuous sorting typically distinguishes particles or cells based on their physical properties. The key requirement is a change in force amplitude based on the property of interest. For example, in an acoustic travelling wave, the force is approximately related to the sixth power of the radius, whilst the drag force opposing motion of the cell is related to the first power. Such a difference in the relationship allows for efficient separation as the migration speed is heavily size dependent (86). Similarly, standing acoustic waves (consisting of two counter propagating travelling waves) (87) and hybrid fields involving standing and travelling components (88) have been shown as effective in particle separation. The balance between the externally generated force field and the drag along the channel can also be used to generate a critical size which delineates between behaviour outcomes. The virtual DLD (vDLD) has been shown to laterally displace particles above a critical cut-off by trapping them in a force field angled across the channel (89). In addition, other physical

View Article Online
DOI: 10.1039/C8LC01239C

properties can be used for distinguishing cells, for example, Petersson *et al*. (85) have shown that cells stiffer and denser than the surrounding medium will move to the pressure nodes in an acoustic field, whilst those less stiff and dense will move to the antinode.

In contrast to passive methods, the external generation of a force field in active systems also allows for temporal variation. As such, these methods can be applied responsively, once a cell characteristic (usually fluorescence) has been identified a desired outcome can be imposed. Such on-chip FACS systems have been achieved using pulsed travelling acoustic waves to deflect the cells of interest into a different streamline, optimised by use of channel constrictions (90) or focused fields (91) and electrical methods (92). Whilst this section focuses on the sorting of cells and particles in bulk suspension for sample preparation, it is worth noting, that many of the techniques, have also been applied to the sorting of droplets containing cells, this includes DLDs for passive sorting (83) and a whole range of active-responsive methods which are detailed in a 2017 review (93).

2. Encapsulation

At the heart of all droplet based scRNA-seq systems is the ability to create individual water-in-oil reaction chambers compatible with a cellular scale. The adoption of droplet microfluidics systems such as the Drop-seq, inDrop and Chromium 10X by biologists has been driven by the need to minimize cost while at the same time increasing cellular throughput. Droplet based microfluidic devices fulfill that need by creating highly monodispersed reaction chambers in the nano/pico-litre range. The adoption of small volume reaction chambers also results in decreasing reagent cost and increasing sensitivity.

In the last few years, the field has seen a logarithmic increase in biomedical publications using droplet microfluidics indicating that the field is rapidly expanding, however the technical process of microfluidic droplet formation is still relatively poorly understood by most biologists. Several companies have streamlined the capture process creating devices that deliver standardized capture conditions (i.e. 10X Genomics' Chromium controller and Dolomite Bio's Nadia instrument). These commercial systems take a one-size-fits-all rule and there is little or no room for modification or optimization of the capture conditions beyond the cellular input described above. In contrast, open-source droplet based methods allow for continuous monitoring and the ability to fine-tune the capture conditions as and when required during the run. This is especially relevant when capturing unique samples (i.e. clinical biospecimens). In order to tailor the cell capture conditions to the needs of specific samples there are a number of practical considerations that need to be addressed, these are described in this section.

View Article Online
DOI: 10.1039/C8LC01239C

Practical overview of the main droplet-based encapsulation methods:

Drop-seq

Drop-seq is a low cost, low efficiency system that introduces a solid bead of between 30 and 60 um (ChemGenes Corporation, Wilmington MA) into a water-in-oil droplet together with a cell and lysis buffer. The solutions for the system published by Macosko and cols. are run at Oil: 15,000 uL per hour, Cells: 4,000 uL per hour, Beads: 4,000 uL per hour (13). As such highly monodispersed droplets with an approximate volume of around 1 nL are created at an approximate rate of 3,000 droplets per second (13). Expected cell capture rates are between 2 and 5% of total cell input. Recently, a commercial droplet generation system has been released by Dolomite Bio, this system uses a simple interface to control all three inputs (beads, cells and oil) via a single pressure source. Much like 10X Genomics' Chromium controller, this system takes advantage of a user friendly interface and in the standard format allows users to create water-in-oil droplets of a set size that is compatible with both the Drop-seq and DroNc-seq protocols. Importantly, unlike the Chromium controller, Dolomite Bio sell an additional control module (Nadia Innovate) that allows full control of all fluids. As noted above this is particularly useful for trained operators when capturing unique samples.

DroNc-seq is a modification of the Drop-seq platform aimed at providing a system to perform scRNA-seq on cell nuclei rather than whole cells. The ability to perform single nuclei RNAseq is useful for cell types and tissue samples that are either hard to digest to a live single cell suspension (i.e. neural cells) or frozen samples. The ability to perform DroNc-seq on nuclei extracted from frozen tissue samples is critically important in being able to utilize the vast arrays of archival frozen samples for single cell transcriptomic analysis.

In the original DroNc-seq paper, the authors showed that by making small changes to the original Drop-seq chip they were able to produce significantly smaller droplets (roughly 70 um diameter); and by altering the fluid flow rates: Oil: 16,000 uL per hour, Cells: 1,500 uL per hour, Beads: 1,500ul per hour, it is possible to create highly monodispersed droplets at rates around 4,500 droplets per second (94). The reduction in droplet volume to around 300 pL is intentionally designed to deal with the lower transcript amounts found in the nuclei as compared to the cell (94).

InDrop (indexing droplets)

The inDrop system is the most complex of the 4 systems listed here and as a result has the highest capture efficiency. As opposed to all other systems, the inDrop chip consists of 4 inlets. The 4th inlet here allows reverse transcription regents to be incorporated with the cell and bead at the time of droplet formation. When run at the following conditions Oil: 360 uL per hour, Cells: 250 uL per hour RT/ lysis: 250 uL per hour and Beads: 40 uL per hour the system is able to create droplets of around 5 nL at a rate of around 225 droplets per second (14).

View Article Online
DOI: 10.1039/C8LC01239C

Droplets contain barcoded hydrogel (polyacrylamide) spheres with cells and reverse transcription reagents together to form single water-in-oil droplets. The barcoded hydrogel microspheres (BHM) are designed in conjunction with the chip geometry to allow very high bead occupancy rates (>90% when fine-tuned) due to their high deformability, regular release and intimate packing. Cell are provided at a concentration such that there is, on average one cell in approximately every 5-10 droplets (~10% of droplets containing a cell). This cell capture is less biased to cell size due to a large cross-section of the microfluidics channel and the resultant larger sized droplet produced.

At the recommended run conditions each capture consists of 100 uL of emulsion containing 2,000-3,000 barcoded cells. Thus this system is ideal for low cell inputs - i.e. clinical samples or high-targeted populations, where very high cell recovery rates are required. As mentioned earlier, inDrop has the best bead occupancy control and it should be expected that between 80 and 95% of all cells passed through the system during the collection window will be co-encapsulated with a bead. It is important here to recognise that the time taken for system stabilization and system dead volume will reduce the total fraction of all cells recovered.

10X Genomics' Chromium

The Chromium system from 10X Genomics is a fully commercial solution for scRNA-seq and is therefore the most opaque in terms of chip design and run settings. It is clear that the Chromium system introduces both a cell, a bead and the molecular reagents to create barcoded cDNA in a droplet. The flow rates are likely controlled via pressure and settings are not widely available. Bead occupancy rates in the Chromium 10x system would indicate that between 50 and 65% of input cells should be co-encapsulated in a droplet. Again it is important here to recognise that the time taken for system stabilization and system dead volume will reduce the total fraction of all cells recovered.

General principles of droplet formation using microfluidic devices.

Despite some differences, all scRNA-seq systems described above are governed by the same microfluidic principles. Droplets are created by the introduction of two immiscible fluids into a microfluidic system resulting in a dispersed phase, typically of droplets contained within a continuous phase. The relative surface energies between each fluid and the channel walls dictates the roles taken (95). Importantly, in droplet based scRNA-seq systems the formation of droplets means that the dispersed phase consists of multiple isolated specimens (the equivalent of test tubes in the lab-on-a-chip concept), whilst the continuous phase offers a means to easily move these specimens around the microfluidic system. As such, two phase

microfluidic systems have become widespread finding applications including single cell encapsulation (96), on-chip chemistry (97) and production of microstructures (98).

In a microfluidics system, droplets can be actively or passively generated. Active systems are more sophisticated and offer many advantages such as enhanced control of droplet production, however the biggest disadvantage is their complexity. The use of active methods requires external actuation but allow highly tailored droplet formation in terms of size, timing and content. Continuous control of the size of droplets produced has been demonstrated using acoustic energy (99, 100) or micro valves (101). Acoustic actuation has also been used to dictate the time of production of individual (102) or a small number of droplets (103). Finally, the content of individual droplets has been varied by using channel deformations to control the input from multiple inlets (104).

In contrast, passive methods are simpler but offer a rapid and repeatable droplet production and lower cost. This is the mode that has been adopted in the scRNA-seq systems described above.

Passive droplet production occurs when the two phases are brought together at the junction between microfluidic channels (105, 106). There are three main junction designs: 1) a T-junction where one phase introduced by a side arm which joins the main channel perpendicularly (107); 2) a flow focusing junction, in which the second phase is introduced into the main channel from side arms located at both sides (as such, symmetry along the length of the main channel is added) (108, 109); and 3) a co-flowing arrangement in which two co-axial cylindrical channels are used, offering rotational symmetry (110). Remarkably these simple geometries, when combined with continuously imposed external flow sources, yield droplets in the nanolitre and even femtolitre (111) size at rates of up to 100 kHz (112).

The droplet systems discussed here use a flow-focusing junction and are designed to create droplets between 300 pL and around 5 nL depending on the system being used and the flow rates at which the device is run. The inDrop system has the largest droplet size at around 5 nL (8) while the modified version of the Drop-seq system designed to encapsulate nuclei instead whole cells (DroNc-seq), has the smallest droplet size at around 300 pL (15).

More detailed information about the physics of the droplet formation can be found in **Additional information BOX 2**, and the droplet formation in the open-sourced systems Drop-seq and inDrop can be visualised in action in the **Supp.Videos 1-3**. As the Chromium system employs a “black box” approach to droplet formation, the formation of droplets in this system is not shown but presumably the system is more similar to inDrop than to Drop-seq.

Distribution of cells and beads in droplets

View Article Online
DOI: 10.1039/C8LC01239C

Cell barcoding is achieved by combining single cells and barcoded beads inside droplets as they form inside the dispersed phase, which is usually oil with surfactant. The proportion of droplets that contain a single barcoded bead AND a single cell determines the useable yields for scRNA-seq. Ideally, each drop should contain both a single cell and a single bead, however none of these systems can control the arrival of both the bead and the cell such that we have 100% of droplets containing only one cell and one bead. A system that could guarantee this whilst still retaining current droplet creation rates would constitute a leap in the performance of these systems, reducing sample loss and help lower the cost barrier.

Without controlling the arrival time of cells or beads, encapsulation of particles is stochastic and follows a Poisson distribution of all possibilities: drops contain a) no bead or cell; b) one or more cells; c) one or more beads; or d) one or more cells and one or more beads. The proportion of these 4 possibilities is a direct function of the concentration of beads and cells. Whilst the molecular processes employed by these systems are designed to minimize the effect of scenarios A and B the formation of multiplets in scenario's C and D (i.e. drops that contain more than one cell or one bead) can dramatically affect the outcome of scRNA-seq experiments. As multiplets (generally referred as "doublets") of both beads and cells can occur, it is possible to have a situation in which multiple cells are barcoded with a single cell barcode and a situation where multiple barcodes are shared amongst a single cell. Although informatics and more recently wet lab tools (such as cell hashing antibodies) have been developed to deal with these situations, it is undoubtedly better to prevent or reduce the number of multiplets being generated by the system in the first place.

The simplest way to minimize doublets is by reducing the concentration of both cells and beads, however this results in poor bead and cell occupancy rates thereby increasing both waste and cost. In the Drop-seq system (and DroNC-seq) the arrival timing of both the beads and the cells is uncontrolled and encapsulation follows a Poisson model, i.e., bead and cell encapsulation are stochastic processes (13, 113). Thus, in this system reducing the formation of multiplets is only possible by reducing the concentration of both cells and beads. The general consensus in the field is to keep doublet rates <5%, in these conditions Drop-seq will only be able to barcode 2 – 5% of all input cells. A model of the distribution of multiplets and the effect of cell and bead concentrations in this system can be found in **Additional information BOX 3**.

One approach to minimize doublets whilst retaining relatively high bead occupancy rates involves the use of deformable beads. Abate and Weitz showed that Poisson encapsulation statistics can by beaten by using close-packed ordering of deformable particles, such as gel

beads, so that a single hydrogel bead can be loaded per drop (114). The inDrop system utilises closely packed barcoded hydrogel beads to achieve ~80% cell barcoding yield (14, 115). Deformable hydrogel beads are also used in the Chromium system and cell capture efficiencies have been quoted as being up to 65% (116). Unfortunately neither the inDrop or Chromium systems can currently control the arrival of the cells. In these systems, cell occupancy rate is still governed by a Poisson distribution.

Irrespective of the way the cells and beads are introduced into droplets, all systems discussed here are capable of generating an emulsion of thousands of droplets in a relatively short time frame (minutes not hours or days) and whilst it would be nice to have a single "best system" for all applications, it is important to note that each technology has its own unique strengths and weaknesses. For example, Drop-seq is currently the cheapest method and not commercially licensed, but it presents the lowest cell barcoding yield, due to the low cell barcode efficiency (and the practical initial volume set-up of pump-driven 3ml syringes), thus Drop-seq is not suitable for samples with low starting cell numbers. For this reason there have been some attempts to improve yields using inertial ordering. Using a spiral microchannel to produce uniform inter-bead distances, Moon *et al.* were able to reduce bead doublet rate between 10 and 20 fold, this deterministic encapsulation achieved 20% capture efficiency with little (1-2%) bead multiplet generation trade-off (117). In contrast higher cell barcode efficiencies can be achieved by the Chromium or inDrop systems and can handle low input volumes so they are best suited for scarce samples. The Chromium system has the lowest bar to entry and is therefore the go-to system for most biologists, there is however a trade off on cost and flexibility.

3. Barcoding and Molecular Amplification

The extraordinary development of single cell RNAseq technology in the last few years has been underpinned by the high-throughput encapsulation of cells in droplets, but also by the improvement of the molecular workflows. An increased molecular sensitivity (partially driven by the reduction in reaction volume) and the incorporation of the barcode into the transcript at an early stage (many cells can be pooled into batches without losing unique cellular identities) are two key molecular improvements that have dramatically simplified sample handling and minimized batch effects.

Droplet-based scRNA-seq systems are based on the encapsulation of barcoded particles (whether hard or soft) with cells (or nuclei) in nanoliter- and sub-nanoliter-sized water-in-oil droplets containing cell lysis buffer. These barcoded beads share the same basic structure 1) a linker region to attach the capture oligonucleotide to the bead; 2) a primer region to allow molecular amplification of the captured transcript; 3) the cell barcode, a unique series

View Article Online
DOI: 10.1039/C8LC01239C

of nucleotides that are identical for all oligonucleotides on the one bead, allowing all the amplified transcripts from the captured cell to be identified as belonging to the same cell; 4) a second series of nucleotides generally referred to as Unique Molecular Identifiers (UMI), designed to individually tag each polyadenylated sequence captured in each cell, and thus enable digital transcript counting and normalization of amplification artefacts; and 5) the poly-d(T) region which allows capture of polyadenylated RNA.

A common molecular workflow of droplet-based scRNA-seq starts after the co-encapsulation of bead and cell in a droplet, upon which the cell is lysed and its contents released. The polyadenylated RNA content of each cell is then able to bind the poly-d(T) region on the bead oligonucleotide. These capture oligonucleotides are either attached to a hard bead (Drop-seq & DroNc-seq) or otherwise incorporated into a deformable bead (inDrop & Chromium). First-strand synthesis (via reverse-transcription) incorporates the cell barcode and the UMI into the same cDNA molecule. The first-strand cDNA is then amplified by different methods depending on the scRNA-seq platform.

The differences in the molecular workflows of the various droplet-based scRNA-seq systems are hardcoded by the specific design of the components of the capture oligonucleotide detailed above and the subsequent amplification method. Critical to any of the methods is the quality and purity of these oligonucleotides, which greatly impacts barcode and UMI diversity and accuracy and the potential binding and specificity of bound polyadenylated sequences and subsequent cDNA yield production. The main molecular characteristics of the different scRNA-seq systems are described below (**Fig. 5**):

Drop-seq

Upon encapsulation of the beads and a cell in a droplet, the lysis buffer introduced with the beads lyses the cells and the poly-d(T) region of the bead-bound oligonucleotides captures the released polyadenylated (polyA) RNAs contained in the drop. Once the polyA-RNAs are annealed forming hybrids of RNA/DNA microparticles, the droplets are broken and pooled together for the subsequent molecular steps to generate a bulk barcoded cDNA library for RNA sequencing. The solid barcoded primer beads contain a linker, a constant template switching oligonucleotide (TSO) PCR handle (23 bases), 12 bases of a cell barcode (millions of the same cell barcode on each bead and 4^{12} different cell barcodes) and 8 bases of a Unique Molecular Identifier (UMI) different for each primer (4^8 different UMI per bead) and an anchored constant poly-d(T) primer (T 27) (**Fig. 5a**).

The first step after RNA capture is a first strand reverse transcription (RT) with template switching (118) under thermal conditions (42 ºC for 90 min). One RT reaction can generate cDNA from up to 90,000 beads. The template switching method uses the specific terminal

transferase activity of Moloney murine leukemia virus (MMLV) reverse transcriptase of adding three extra Cs at the end of the full cDNA first strand synthesis product that will serve as an anchor for a TSO primer, added together with the RT reagents, to extend and tag with a PCR anchor at the 3' end of the cDNA (corresponding to the 5' end of the original mRNA molecule). In that sense the resulting cDNA will be tagged at both ends with TSO PCR handles. The resulting RNA/cDNA micro particles are called STAMPs (**s**ingle-cell **t**ranscriptomes **a**ttached to **m**icro **p**articles) and are used as the template for the generation of a double stranded (ds) cDNA barcoded library upon PCR amplification (13 cycles) by using a single TSO primer. In order to obtain a good PCR enrichment and low bias, PCRs are performed in batches of 2000-4000 beads that would contain ~100 STAMPs, although this would depend on the capture efficiency (the standard is around 5%, which means 100 STAMPs in 2,000 beads). The number of PCR cycles and beads per PCR may be increased in samples with low transcriptional activity (i.e. resting T cells) or with experiments with lower than 5% capture efficiency. One unique characteristic of Drop-seq compared with inDrop and Chromium is that capture oligonucleotides are not released from the solid bead and thus mRNA capture relies on mRNA diffusion and binding to the poly-d(T) on the bead.

The next step requires fragmenting the cDNA, adding sequencing adaptors and multiplexing samples if needed. This is achieved by tagmentation of the cDNA (119) using the Illumina Nextera XT kit which takes the advantage of the activity of a transposase to enable the simultaneous and efficient fragmentation of the DNA and ligation of custom Drop-seq sequencing adaptors in a single reaction. Multiple Nextera reactions may be set up when sample multiplexing is required. The resulting tagmented DNA library will have an average size of 500-680 base pairs and will be ready for Illumina paired end sequencing.

InDrop

The barcoded oligonucleotides in inDrop are carried in the form of deformable beads or barcoded hydrogel microspheres (BHMs). These capture oligonucleotides consist of an acrylic phosphoroamidite moiety, a photo cleavable spacer, T7 RNA polymerase primer, a 5' PE1 sequencing adaptor, individual cell barcode combinations, a unique identifier molecule (UMI) and an anchored poly-T primer (**Fig. 5b**). The inDrop barcode is created by linking an oligo containing a photo cleavable spacer, T7 RNA polymerase primer and an adapter sequencer to the BHM. Following this, a split-and-pool approach combined with primer extension using 2 pools of 384 unique oligos is utilised. The first pool contains oligos that consist of an adapter region the first part of the cell barcode (8-12 bases) and an anchoring sequence. The second pool consists of a sequence complimentary to the anchor from pool one, an 8 base pair sequence for the second part of the cell barcode, a 6 base Unique Molecular Identifier (UMI)

View Article Online
DOI: 10.1039/C8LC01239C

region and an anchored-poly-d(T) sequence. After the barcode formation process this results in an on-bead barcode structure that consists of a linker, a T7 promoter region followed by a two part cell barcode (147, 456 unique possibilities), a 6 base pair UMI and a 19 base poly-d(T) region.

In the inDrop system, the droplets co-encapsulate cells and BHMs in a lysis/retrotranscription (RT) mix. Bead barcodes indexes are released exposing the emulsion to a high intensity UV light (365 nm at ~10 mW/cm^2, BlackRay Xenon Lamp). Upon release by UV, the poly-d(T) will anneal polyadenylated RNA sequences. The whole droplet emulsion is then subjected to thermal incubation (50 ºC for 2 h) for first strand retrotranscription to create barcoded RNA/DNA hybrids. At this point the droplet emulsion is demulsified and processed for amplification and DNA library preparation based on CEL-seq/MARS-seq protocols (6, 7).

In contrast to Drop-seq, the captured RNA material is not linked to a solid bead, however, prior to droplet breakage it is possible to subsample the library to adjust the number of cells to be assayed ensuring that the resultant library does not exceed the limits of barcode variety and the sequencing strategy can be calculated to achieve the desired sequencing depth per cell. Given the limited cell barcode diversity (about 147,000 unique barcodes) it is recommended to sequence no more than 3000 cells in any single inDrop sequencing library although multiple libraries can, of course be sequenced to increase cell numbers.

The inDrop protocol is the only one discussed here that is based on *in vitro RNA transcription (IVT)* resulting in linear amplification of the cDNA after second strand synthesis rather than the exponential PCR-based TSO approach utilised by Drop-seq, DroNc-seq and Chromium. IVT is performed by using the T7 RNA polymerase promoter sequence contained within the capture oligonucleotide as a priming site; the resulting amplified RNA libraries are then fragmented and retrotranscribed back to cDNA using a custom read 2 primer (PE2) to obtain the desired fragment size for Illumina sequencing along with the addition of a second read primer to each molecule, respectively. Finally, the cDNA is PCR enriched (10-13 cycles) using indexed custom read 1 and read 2 primers (PE1/PE2 mixes) to allow multiplexed sequencing of multiple captures. The expected average size of the DNA libraries is between 400-600 bp.

Chromium

The gel barcoded primer beads contain a linker, a constant TSO PCR handle, 14 bases of a cell barcode (around 734,000 unique barcodes) and 10 bases of a UMI and an anchored constant poly-d(T)30 primer (**Fig. 5c**). Chromium cDNA generation and amplification is very

View Article Online
DOI: 10.1039/C8LC01239C

similar to Drop-seq, since these two platforms are based on TSO PCR reactions with minor differences. However there are two important differences in the 10X molecular process, firstly the capture oligonucleotides are released into a liquid phase, in this case by SS-bond reduction with DTT. Secondly, the first strand synthesis with template switching RT occurs in each droplet prior to emulsion breakage.

Following droplet breakage, amplification of cDNA fragments is achieved via PCR followed by enzymatic fragmentation, A tailing, ligation and Sample Index (SI) PCR. The number of PCR cycles may be altered for low captured cell numbers and/or of samples expected to have low transcriptional activity. The final Illumina-compatible cDNA library is dual-sided size-selected to have an average bp size of 400-600, optimal for loading onto Illumina sequencing platforms.

4. Sequencing

Currently all common single cell transcriptomic platforms are designed around the Illumina sequencing instruments and a technique known as Sequencing-By-Synthesis (SBS). SBS requires that the genomic material to be sequenced is provided as a fragmented strand (generally in the region of between 200 and 1000 base pairs) which is bound to a solid, imageable substrate via complementary binding of the target strand to oligonucleotides pre-attached to the flow cell substrate. This is achieved through the incorporation of flow cell binding regions (referred to as the P5/P7 regions) into the target fragments during library preparation.

Once bound, the target strands undergo clustering. Clustering is designed to increase the number of target strands in a single location (the cluster) such that upon incorporation of a fluorescently labelled nucleotide base during the sequencing will result in a fluorescence signal in the cluster being sufficiently above background noise so as to be able to confidently determine which base is incorporated.

As there are 4 bases that may be read from any target strand, SBS must be able to confidently identify which base is being read at any point in time. To do this there must be 4 distinct fluorescent states. This can be achieved by reading distinct colours for or by using a combination of only 2 colours. Four colour chemistry requires that the imaging system used is capable of detecting all 4 fluorochromes whereas two colour chemistry relies on only 2 colours to create a binary system that can detect all combinations of these 2 colours, that is colour 1, colour 2, colour 1 and 2, no colour. While the detection system used may seem trivial, as we shall describe later, it is becoming clear that whilst 2 colour chemistry is cheaper, the older 4 colour chemistry seems to have some very distinct advantages.

View Article Online
DOI: 10.1039/C8LC01239C

Irrespective of the colour chemistry used, SBS is able to build high fidelity sequence reads by the progressive detection of fluorescently labelled base one base at a time. The process can be repeated up to 300 times in a single run and target fragments can be sequenced from both ends. This allows a current maximum length of 600 bases from any single target fragment. In most scRNA-seq the aim is to keep the number of sequence cycles required to a minimum as sequencing cost is directly proportional to cycle number and the lower the cost the more cells that can be analysed.

As each system has different barcode designs and incorporates the flow cell binding sequences at different points on the target fragment it is important to ensure that the sequencing conditions match the platform being used and the library produced. Unlike run configurations for bulk RNASeq, which generally utilise a read length distribution equal across paired-end reads, scRNA-seq run configurations usually do not need very many bases to be detected on each target fragment. To gain biologically interpretable data we only need to sequence far enough into the 3' region to ensure that we have read the full cell barcode and UMI region and far enough into the transcript (minimally about 18 bases) to be confident that the transcript detected cannot be mapped to any other transcript. Thus sequencing conditions should be customised to ensure these conditions are met with the smallest possible sequencing kit for the information desired.

Illumina kits are supplied in various "sizes" based on cycle number (equivalent to the numbers of bases able to be detected from each target fragment. The kits are nominally defined as say 75 cycles, 150 cycles, however these kits contains extra reagents sufficient for at least 2 x 8 bp index reads. We have found up to 91 cycles (software-limited) of sequencing can be obtained using a NextSeq500 75 cycle kit thus we are able to sequence inDrop, Drop-seq/DroNc-seq on the Illumina 75 cycle kits. 10X Genomics' Chromium Single-Cell solution however, recommends far longer reads for the transcript and therefore requires the use of a 150 cycle kit which essentially doubles the cost of sequencing samples generated on this platform. Subsampling of transcript reads from Chromium libraries suggests that shorter transcript reads can still identify the same number of transcripts and therefore may benefit from the reduced sequencing costs of a smaller sequencing kit size with no or minimal effect on data quality.

Below are some guidelines we use for sequencing the libraries from the various platforms:

Sequencing Specifications

View Article Online
DOI: 10.1039/C8LC01239C

Drop-seq/ DroNc-seq

Drop-seq is a 3'end scRNA-seq method using paired-end Illumina sequencing. Read 1 will read the cell barcode & UMI and should be a minimum of 20 cycles. It should be noted however that once read 1 exceeds 20 cycles it will enter the poly-A region of the transcript and clustering metrics will be negatively affected – these metrics however do not affect the overall quality of the data. Read 2 should be 50 cycles and reads the gene identity. When multiplexing it should include 8 cycles for index detection. Drop-seq requires the addition of a custom primer Read 1 in order to sequence the cell barcode and UMI regions.

InDrop

InDrop is also a 3'end scRNA-seq method using paired-end Illumina sequencing. InDrop libraries created using version 2 chemistry utilise Read 1 to read the gene sequence and is usually set to 37 cycles. Read 2 will read the cell barcode and UMI. Using 51 cycles allows the reading of cell barcode 1 (8-12 bases), the cell barcode anchor region (22 bases), cell barcode 2 (8 bases) and the UMI region (6 bases) as well as 3 bases into the poly T region. Six cycles are required on the index read when multiplexing for index detection. InDrop requires the addition of custom read primers. At the time of publication we have no experience with version 3 inDrop libraries as reagents are not commercially available but the sequencing approach removes the need to read the low diversity barcode anchor region and thus aims to give better data utility on the sequencing by reducing base call errors (120).

10X Genomics' Chromium

In 10X Genomics' Chromium 3' assays the recommended sequencing run uses a 150 cycle kit and the sequencing specifications are as follows. Read 1 reads the cell barcode and UMI and should be set a 26 cycles, as in Drop-seq libraries once the read exceeds this recommendation it hits the poly-A regions and sequencing metrics are adversely affected. Read 2 contains the gene sequence and is usually set to 98 cycles. When multiplexing, and Index read 1 of 8 cycles allows index detection. The Chromium system also comes as a 5' assay which allows gene sequences further from the poly A region to be detected and is important for many assays.

Special considerations for sequencing of scRNA libraries.

View Article Online
DOI: 10.1039/C8LC01239C

Illumina sequencers allow the use of custom read 1, read 2 and index read primers which provides flexibility in design of capture oligonucleotide sequences and adaptor/indexes, however care must be taken in designing custom sequencing primers to ensure compatibility with the fixed temperature profiles used for primer annealing during sequencing. Custom primers are able to be used completely independently of the default Illumina primers or alternatively can be spiked in alongside the default primers which then allows the use of the bacteriophage sequencing control library, PhiX. The presence of PhiX at a low concentration (~1%) allows a quality control during each sequencing run for cluster generation, sequencing and alignment. At high concentrations (>5%), PhiX can be used to offset low nucleotide diversity within library pools that can otherwise adversely affect sequencing quality. Diversity is especially important in the first 4 – 7 cycles of sequencing read 1 as the sequencing software uses images at these early cycles to identify the sequencing clusters during a process called template generation. Other sequencing metrics such as phasing/pre-phasing, colour matrix corrections and pass filter calculations occur during cycles 1 – 25, again highlighting the importance of nucleotide diversity. All these corrections and calculations are used in base calling and quality score calculations for all cycles in a run for those clusters that pass filter.

A brief comparison of different chemistries in Illumina sequencing platforms is described in **Additional information BOX 4**.

Technical challenges and future directions of Droplet genomics

Despite the justifiable uptake of the droplet-based scRNA-seq systems for use in understanding complex biological processes, they still suffer some limitations. These include 1) reliability, accessibility/simplicity, 2) efficiency, 3) the inability to analyse the full-length transcriptome, and 4) loss of the spatial information in the tissue architecture (physical relationships between cells).

1) *Reliability, accessibility/simplicity:*

As with all new technologies, early prototypes and non-commercial solutions there is a degree of risk and uncertainty that comes with interfacing with technology on the cutting edge of what is possible. In droplet microfluidics this is especially true as the technology brings together fields that are traditionally disparate. As we have discussed, the 3 droplet-based systems described here are conceived using essentially the same concepts, it is clear that the commercial platforms of Chromium single-cell solution (10X Genomics) and Nadia (Dolomite Bio) are the most accessible and simplest and come with a certain level of user support. Although an experienced operator can get the same reliability out of the inDrop, Drop-seq/DroNc-seq systems, the majority of users will not have access and experience with microfluidic droplet generators nor will invest the time required to make sure these platforms

View Article Online
DOI: 10.1039/C8LC01239C

perform at their optimal level, in which case high-density micro-well based solutions can be more suitable. For those users interested in developing their own assays, the flexibility offered by open access platforms would be preferable.

As the field matures many vendors are looking to produce commercial systems that look to rival the Chromium system for reliability, accessibility and simplicity. For those working in this space it is imperative that the systems; 1) can be used by anyone with basic biological wet lab experience; 2) the systems perform in an identical manner from one experiment to the next and does so for the duration of a standard project; 3) that instrument quality assurance metrics can be rapidly and easily obtained prior to running samples (something which is lacking in the systems discussed as part of this review); 4) such metrics are clearly defined and relevant to the experimental outcome 5) provides a complete ecosystem which allows data to be readily analysed and 6) the final dataset from point 5 incorporates quality controls metrics such that a user can identify the validity of the results prior to full scale analysis.

2) *Efficiency:*

The true efficiency of droplet based scRNA-seq systems is more than simply how many beads and cells are able to be co-encapsulated into the same droplet. Efficiency losses can occur at cell preparation, encapsulation and barcoding, library preparation, sequencing and informatics. In our experience we see losses at all levels that we believe can be overcome with future developments and a better developed understanding of entire workflows.

Losses at the cell preparation level are reasonably well understood in the field of flow cytometry and are generally a result of incomplete tissue digestion, processing (losses due to cell filtering and centrifugation), cell death and losses associated with any specific cell enrichment strategies (FACS, MACS™, microfluidics). We envision a day when these losses can be minimised through better cellular manipulation tools and we expect that microfluidics will play a major role here.

Encapsulation losses occurring when a droplet does not contain both a bead and a cell are the most commonly acknowledged losses in droplet scRNA-seq. As discussed above, systems like Drop-seq have very poor encapsulation efficiencies followed by the Chromium system and finally the inDrop system being the best performer (as it can be fine-tuned). Both Chromium and inDrop systems partially overcome the Poisson limit by utilising deformable beads. One of the most significant advancements in droplet scRNA-seq would be the ability to accurately control the arrival time of both cells and beads. In this way it would be possible to capture 100% of the cells that are placed into the system with zero or next to zero, multiplets.

View Article Online
DOI: 10.1039/C8LC01239C

Molecular losses are also a significant challenge. Current estimation suggests that only 10% to 20% of all transcripts from each cell are able to be efficiently captured and sequenced (12, 121). Thus, although droplet scRNA-seq systems provide a high-level interrogation of cellular heterogeneity and cell typing, generally they are not well suited to detecting lowly expressed genes (especially at low sequencing read depth). This limitation has been solved in the BD Rhapsody™ platform using molecular targeting of specific transcripts to improve detection and sensitivity of low abundant genes, with the trade-off of loss of whole transcript amplification information. An improvement in RNA to cDNA conversion efficiency or the ability to sequence directly from RNA at the single-cell level would also improve the total cellular characterisation power of these tools.

The efficiency of different scRNA-seq platforms has been evaluated previously using the same experimental system (122), classifying some plate- and droplet-based methods according to their sensitivity. This comparison concluded that Smart-seq2 was the most sensitive method, followed by SCRB-seq, and the Fluidigm C1 versions of Smart-seq and CEL-seq2, while MARS-seq and Drop-seq were the poorest performers with a reduction of 50% in their capacity to identify genes. In regards to technical evaluation of droplet-based scRNA-seq methods, a comparison of sensitivity using the probability of detection of ERCC (External RNA Controls Consortium) across different experimental cohorts (11, 22) showed that inDrop was the most sensitive of the methods followed by Drop-seq, with Chromium in third position. Importantly, when sequencing depth was normalized, inDrop maintained its first position over Chromium and Drop-seq. More recently, Zhang and cols.(123) directly compared these most common droplet based scRNA-seq methods using the same cell line as the experimental system. In this comparison, Chromium generally showed the least technical variation followed by Drop-seq, while inDrop showed the highest variation. Similarly, gene detection capacity was higher in Chromium closely followed by Drop-seq (especially at lower sequencing read depths, where Chromium and Drop-seq performed almost identically) and far higher than inDrop. Finally, Chromium and Drop-seq presented a similar noise level, which again was well below that of the inDrop platform.

It is important to highlight that droplet-based scRNA-seq methods are relatively new to the field and further molecular improvements are likely to occur as has happened with the plate-based methods. An example is the recent re-design of inDrop barcodes with more balanced GC content and increased Hamming distance, which allows correction of most of the single base errors (124). A simplified qualitative comparative of the three platforms can be found in **Figure 6**.

View Article Online
DOI: 10.1039/C8LC01239C

Even once libraries are created and are loaded onto the sequencing platform there continues to be additional losses. These occur due to limitations in the sequencing tools currently available as well as the accuracy of the chemistry employed during library sequencing. These losses are predominantly from an inability to identify the correct nucleotide base sequence, which can stem from two main sources: direct base mutation (i.e. where a base substitution/addition or deletion has occurred either during barcode synthesis or library generation) or by base call error (where the sequencing platform is unable to accurately call the base being read). In Illumina instruments (the most often used sequencing platform for scRNA-seq) the sequence-by-synthesis (SBS) process will generally contain base call errors at a rate of about 1 in 1000 but these rates dramatically increase when too much library is added to the sequence run (resulting in highly clustered flow cells) or with low complexity libraries.

Low complexity sequences are characterized by low nucleotide diversity in a given base position (i.e. biased proportion of T, C, G, or A nucleotides). A system with a limited number of barcodes will enrich for these low diverse regions, however this can be solved by a careful design of the barcodes ensuring equal proportion of each nucleotide in each position. However, all RNA captured using these technologies will encounter a region around the same area that will be rich for a polyA sequence. Low complexity libraries suffer high base call errors as the instruments simply cannot determine the fluorescent label on the nucleotide being read when the flow cell is saturated with the same base. This base call error not only affects that base being read but affects the quality of the base calls in the nucleotides surrounding the low complexity region. To deal with this, inDrop version 2 has a variable length cell barcode region in an attempt to offset the low complexity anchor region in different clusters and thus increase library complexity while inDrop version 3 simply removes the need to sequence through the low diversity barcode anchor region.

Another source of loss during sequencing comes from the short read nature of the SBS. SBS sequencing only allows relatively short reads (<300 bases) and thus much of the information in the captured transcript is lost. Whilst cellular subpopulation profiling only requires mapping of very short reads for accurate determination of transcriptome signatures, analysis of splice variants is not readily achievable using standard SBS approaches. To overcome this, long read sequencing, such as that offered by Oxford Nanopore Technologies (ONT) nanopore technology or Pacific Biosystems' Single Molecule, Real Time (SMRT) technology can be used instead. These systems however still suffer from reduced throughput, higher cost per read and a perceived base call accuracy issue. We predict the ONT Nanopore system will be increasingly embraced in scRNA-seq (125, 126) and will prove highly valuable

View Article Online
DOI: 10.1039/C8LC01239C

for looking at full length transcriptome reads or where specific targeting of sequences towards the 5' end of the RNA are required.

Sequencing efficiencies can also be increased by implementing a targeted approach to analyse the transcriptome. Essentially these approaches use *a priori* knowledge of the biological system being analysed to reduce the number of individual transcripts being probed. Instead of assaying tens of thousands of different transcripts in an experiment it is possible to focus on just a few hundred. Targeting dramatically reduces sequencing cost (less off-target reads) and may also improve transcript capture efficiencies. To our knowledge the only commercially available system that currently provides targeted capture workflows is the Rhapsody™ from Becton Dickinson.

3) Inability to analyse the full-length transcriptome.

Currently these droplet-based system also lack the ability to perform full length transcriptome analysis. While microfluidics was instrumental in the early adoption of full length single cell transcriptomic analysis (Fluidigm C1 system coupled with SMART-seq) these platforms quickly fell from favour due to the high running cost, relatively low throughput abilities, and inherent bias due to cells being physically captured in "wells" based on physical size and density as well as a major issue with doublet rates in the medium chips (13). Interestingly, Zhang *et al.* successfully performed a proof-of-concept experiment using the SMART-seq2 chemistry with the inDrop system (123). A full-scale adoption of this SMART2/inDrop system will produce another leap advance in the field of scRNA-seq.

4) Spatial information

The loss of spatial resolution in droplet-based scRNA-seq, produced by the inherent requirement of cells to be in suspension, is perhaps the biggest challenge for droplet microfluidics. Techniques such as Geo-seq (127), TIVA (128), NICHE-seq (129), FISSEQ (130), SeqFISH (131), MERFISH (132); and more recently Spatial Transcriptomics (133), HDST (134) and Slide-seq (135) have been designed to combine information of single cell transcriptomes with their physical cellular relationship. These technologies however are still 1) low throughput, 2) do not target the polyA mRNA (instead relying on a targeted approach with a limited number of probes), 3) require genetically modified organisms, 4) are inefficient or biased towards certain transcripts, 5) require specialist photo-activation or 6) are not performed at the level of single cell resolution (instead using discrete localization techniques allowing 100 um tissue resolution or requires specialist photo-activation) (136).

Concluding remarks

View Article Online
DOI: 10.1039/C8LC01239C

In the past few years droplet based scRNA-seq technologies such as Drop-seq, inDrop, and Chromium have enabled unprecedented investigation into both normal function and disease systems. The power of the scRNA-seq technology lies in the unbiased ability to characterise many thousands of cells from a tissue and has been driven by parallel developments in cytometry, microfluidics, genomics and informatics.

Multiple single-cell highly parallel molecular screening techniques have already benefited from microfluidic applications (137-143). Functional genomics is one of the latest (and with the deepest impact) fields to have adopted microfluidics as a platform to radically increase scRNA-seq throughput (144). As scRNA-seq becomes more dependent on microfluidic solutions, it converges closer with the wider field of Cytometry. With the meteoric development of scRNA-seq it becomes obvious that the ability to interrogate the transcriptomic profiles one-cell-at-the-time is just the tip of the iceberg, and we envision that other cellular genomic aspects will experience a similar expansion in the near future. Thus, the term Genomic Cytometry alludes to those technologies that use a genomic readout (notably nucleotide sequencing) to quantify cellular characteristics at the single cell level, including RNA, DNA, chromatin conformation and protein expression. As such a simultaneous characterization of individual cells from multiple 'omics perspectives is just around the corner. The systems are now being adapted to allow interrogation of the DNA (both nucleotide sequence as well as conformational states) and protein profiles of the cell. Although single-cell DNA sequencing is still not widely available and high-throughput single-cell epigenetic techniques are beginning to emerge, tools that allow for protein detection are readily available (145) using a DNA barcode linked to an antibody to allow detection of epitope expression via an antigen – epitope reaction and subsequent DNA sequencing readout. The combination of protein identification and whole transcriptome amplification (62) has proven to be a very successful method to interrogate complex biological systems. These tools have been commercialized in the totalSeq™ (Biolegend) and AbSeq™ (Becton Dickinson) range. Another example of a combined approach is G&T-seq, achieving genome and transcriptome sequencing of single cells simultaneosly (146).

A multiomics integration of information at single cell resolution will dramatically advance our ability to fully understand the complex machinery of the cell (147) paving the way to discoveries that will revolutionise our current understanding of biomedicine.

View Article Online
DOI: 10.1039/C8LC01239C

Additional Information

Additional information BOX 1

One particular aspect that can affect the experiment procedure is sort volume. For example, a 70 um nozzle operating at a pressure of 70 PSI and a droplet frequency of 90 KHz can be expected to generate a droplet of around 1.2 nL while a standard sorter using a 100 um nozzle operating at a pressure of 30 PSI and a droplet frequency of 30 KHz can be expected to generate a droplet of around 3.4 nL. The expected sort volume, and maximum event rates for standard electrostatic droplet based FACS is given below in table 1. This table can be used to estimate the concentration of cells post sort and will be critical in estimating scRNA-seq doublet rates (discussed in more detail below).

		Single cell (1 drop mask)		Purity Sort (1.5 drop mask)		Enrich/ Yield sort (2 drop mask)	
Nozzle size	Expected droplet frequency	Maximum sort rate (EPS)	Volume per 10, 000 cells	Maximum sort rate (EPS)	Volume per 10, 000 cells	Maximum sort rate (EPS)	Volume per 10, 000 cells
70um	90 kHZ	18000	12 ul	18000	17 ul	<90000	22 ul
85 um	50 kHZ	10000	20 ul	10000	28 ul	<50000	37 ul
100um	30 kHz	6000	34 ul	6000	50 ul	<30000	67 ul
130um	12 kHz	2400	70 ul	2400	103 ul	<12000	137 ul

It is also important to note that many factors other than the nozzle size can affect post sort viability of cells sorter design, i.e. cuvette vs. jet-in-air, sort/collection buffers used, accuracy of sort stream targeting and temperature.

Additional information BOX 2

The physics of passive droplet production has been studied extensively, with three production regimes identified; squeezing, dripping and jetting (**Fig. 3**) (45, 46). The transition between these regimes is best described using the capillary number $Ca = \mu v/\gamma$, in which v is the average flow velocity of the continuous phase, μ is the viscosity of the continuous phase, and γ is the surface tension between the two phases, as such, this dimensionless parameter, represents the relative importance of viscous forces over surface tension.

At capillary numbers below approximately 0.015 the squeezing regime occurs (45, 47, 48). Here, shear forces are small and the break-up is pressure driven. As the dispersed phase enters the continuous phase, the interface between the two fluids grows such that the flow of the continuous phase is briefly blocked. As this occurs, the surface tension which is acting to hold the interface together is overcome by the growing pressure on the upstream side of the interface and break-up occurs. The droplets which are formed fill the width of the channel, sometimes referred to as plugs. In this regime, the droplet size is largely independent of the ratio between the viscosities of the two phases (viscosity not being key to the break-up mechanism). However, a relationship is seen with the ratio of the flow velocities of the two phases, with a faster dispersed phase velocity leading to larger droplets as more fluid enters the channel before the pressure builds up enough to cause break-up (45).

In the dripping regime, the higher Capillary number means an increased significance associated with viscous effects. As the dispersed phase enters the continuous phase, the shear forces acting on the interface are sufficient to cause break-up. This can occur well before the channel is blocked; as such the droplets can be significantly smaller than the channel width. In the dripping regime the droplet size decreases rapidly with increasing Capillary number, this is in stark contrast to the squeezing regime in which little alteration is seen (48). Again in contrast to the squeezing regime, the size of the droplets created by dripping is independent of flow ratio (45). As such if the dispersed phase flow rate is kept fixed, an increase in continuous phase flow rate, and hence Capillary number, decreases droplet size and, due to mass conservation, increases production rate.

Finally, in the jetting regime the dispersed phase forms a filament within the flowing continuous phase, and break-up, due to Rayleigh-Plateau instabilities, only occurs some distance downstream of the junction (28).

Additional information BOX 3

If one assumes that the volume of the drop is much greater than the volume of barcoding beads or cells which are independently allocated to drops, the number of beads or cells encapsulated by drops is described by independent Poisson variables, B and C, respectively. The probability of i beads and j cells per droplet is given by the equation

$$Pr[B = i, C = j] = e^{-(\lambda_B + \lambda_C)} \frac{{\lambda_B}^i {\lambda_C}^j}{i!j!}$$

where λ_B and λ_C are the average number beads and cells per drop, respectively.

The probability that a cell is barcoded (at least one bead per drop) is

$$Pr[B > 0] = 1 - e^{-\lambda_B}$$

View Article Online
DOI: 10.1039/C8LC01239C

where $e^{-\lambda_B}$ is the probability that there are no beads per drop.

As discussed above however, scRNA-seq will be confounded by multiplets. The probability that there is a multiplet that includes barcoding beads is

$$Pr[multplet] = Pr[B > 0] - Pr[B = 1, C \leq 1]$$

where $Pr[B = 1, C \leq 1] = e^{-(\lambda_B + \lambda_C)}\lambda_B(1 + \lambda_C)$

Therefore, the probability of multiplets given cells are barcoded with selective amplification is

$$Pr[multiplet|\{B > 0\}] = 1 - \frac{e^{-(\lambda_B + \lambda_C)}\lambda_B(1 + \lambda_C)}{1 - e^{-\lambda_B}}$$

For systems where there is deterministic encapsulation of beads with exactly one bead per drop barcoding yield will be 100%. If cells are randomly assigned to droplets according to the Poisson distribution, the probability of cell multiplets (more than 1 cell per droplet) is

$$Pr[cell\ multiplets] = 1 - e^{-\lambda_C}(1 + \lambda_C)$$

The calculated cell yield and incidence of multiplets is shown in **Fig. 4A**. The number of cells that are barcoded is directly related to the number of beads per drop. While a 60% efficiency is possible at 1:1 bead/drop ratio (heavy blue and black lines), the incidence of multiplets is very high (>40 %), even when there are small numbers of cells per drop. The incidence of multiplets decreases dramatically as the concentration of beads and cells decrease; however the diminishing barcoding yield is a big trade-off. For example, to achieve a multiplet frequency of less than 1 per 1000 cells (dotted black and blue lines, **Fig. 4A**), both the bead and cell to drop ratio would be set at around 0.001 (0.1% barcoding yield).

The green line shows the incidence of cell multiplets for allocation of *exactly* 1 bead per droplet. Yields would be 100% with a dramatic reduction in multiplet formation.

Drop-seq/ DroNc-seq systems can still be used in the laboratory to good effect and are generally considered to allow between 2 – 5 % of all input cells to be captured with multiplet rates of around 5%. **Fig. 4B** shows the diminishing barcoding yield as the multiplet frequency is reduced for a range of cell/drop ratios. For cell barcoding yields less than 1%, reducing the cell/drop ratio lower than 0.01 has negligible effect on multiplet frequencies.

Additional information BOX 4

The latest Illumina sequencers utilise 2-colour chemistries which allow faster runtimes as only 2 images of the flow cell are taken at each cycle, rather than 4, however there are some indications that this chemistry, particularly on the NextSeq500 (other 2-colour platforms are untested) deal poorly with regions of low complexity, as is the case with the InDrop cell barcode/ UMI read whereby basecall quality and error rates are adversely affected due to a region of low complexity between barcodes 1 and 2 leading to poor usability of reads.

View Article Online
DOI: 10.1039/C8LC01239C

Sequences obtained after DNA sequencing from all three platforms consists of a paired read, in which one mate contains information on mRNA, while the other contains information on cell/barcode and UMIs. In theory, each bead should contain one barcode, but due to barcode synthesis and sequencing errors this is not necessarily the case. Even after computational corrections sequence errors can involve 15-20% loss of reads for each droplet platform, with Chromium outperforming Drop-seq, and inDrop version 2 chemistry retrieving less than 50% of reads with capture oligonucleotide quality in this case presumably plays a major role. Furthermore, unique molecular identifiers tend to be more diverse than expected (148), since sequencing errors introduce additional unique sequences. This overestimation of UMI counts is commonly present for short UMIs (<8), over sequenced libraries or sequencing technologies with larger error rates. In our experience, Illumina sequencing technologies based on 2-color chemistries are particularly susceptible to higher error rates, especially if reads have a common structure – a fixed sequence in the middle of all of them, as in the case of inDrop v2 (eliminated from v3) or even repeats of purines-pyrimidines as in the case of *ClonTracer*. These problems cannot be resolved through bioinformatics, but only by adding sequence diversity with PhiX spike-ins or sequencing with (usually more expensive) 4-color chemistries.

View Article Online
DOI: 10.1039/C8LC01239C

References

1. Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, et al. An estimation of the number of cells in the human body. Ann Hum Biol. 2013;40(6):463-71.
2. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6(5):377-82.
3. Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc. 2010;5(3):516-35.
4. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 2011;21(7):1160-7.
5. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012;30(8):777-82.
6. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. 2012;2(3):666-73.
7. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science. 2014;343(6172):776-9.
8. Soumillon M, Cacchiarelli D, Stefan S, Oudenaarden A, Mikkelsen TS. Characterisation of directed differentiation by high-throughput single-cell RNA-Seq. bioRxiv. 2014.
9. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods. 2013;10(11):1096-8.
10. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. 2016;17:77.
11. Bagnoli JW, Ziegenhain C, Janjic A, Wange LE, Vieth B, Parekh S, et al. Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. Nat Commun. 2018;9(1):2937.
12. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. Mol Cell. 2015;58(4):610-20.
13. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;161(5):1202-14.
14. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161(5):1187-201.
15. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049.
16. Rotem A, Ram O, Shoresh N, Sperling RA, Schnall-Levin M, Zhang H, et al. High-Throughput Single-Cell Labeling (Hi-SCL) for RNA-Seq Using Drop-Based Microfluidics. PLoS One. 2015;10(5):e0116328.
17. Yuan J, Sims PA. An Automated Microwell Platform for Large-Scale Single Cell RNA-Seq. Sci Rep. 2016;6:33883.
18. Gierahn TM, Wadsworth MH, 2nd, Hughes TK, Bryson BD, Butler A, Satija R, et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat Methods. 2017;14(4):395-8.
19. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. Cell. 2018;172(5):1091-107 e17.
20. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883-92.
21. Onuchic V, Hartmaier RJ, Boone DN, Samuels ML, Patel RY, White WM, et al. Epigenomic Deconvolution of Breast Tumors Reveals Metabolic Coupling between Constituent Cell Types. Cell Rep. 2016;17(8):2075-86.

22. Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, et al. Power analysis of single-cell RNA-sequencing experiments. Nat Methods. 2017;14(4):381-7.
23. Bach K, Pensa S, Grzelak M, Hadfield J, Adams DJ, Marioni JC, et al. Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nature Communications. 2017;8(1):2128.
24. Pal B, Chen Y, Vaillant F, Jamieson P, Gordon L, Rios AC, et al. Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. Nature Communications. 2017;8(1):1627.
25. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. Cell.172(5):1091-107.e17.
26. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature. 2014;510:363.
27. Navin NE. Cancer genomics: one cell at a time. Genome Biol. 2014;15(8):452.
28. Liu T, Wu H, Wu S, Wang C. Single-Cell Sequencing Technologies for Cardiac Stem Cell Studies. Stem Cells Dev. 2017;26(21):1540-51.
29. Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR, et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science. 2018;361(6402):594-9.
30. Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, et al. Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell. 2016;166(5):1308-23 e30.
31. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016;352(6282):189-96.
32. Jennifer Rood MC, Robert Majovski, Aviv Regev, and Orit Rozenblatt-Rosen. The Human Cell Atlas. The HCA consortium. 2017.
33. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell. 2016;167(7):1883-96 e15.
34. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol. 2018;18(1):35-45.
35. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun. 2017;8:15081.
36. Bach K, Pensa S, Grzelak M, Hadfield J, Adams DJ, Marioni JC, et al. Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nat Commun. 2017;8(1):2128.
37. Pal B, Chen Y, Vaillant F, Jamieson P, Gordon L, Rios AC, et al. Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. Nat Commun. 2017;8(1):1627.
38. Han X, Chen H, Huang D, Chen H, Fei L, Cheng C, et al. Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing. Genome Biol. 2018;19(1):47.
39. Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X, et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell. 2010;6(5):468-78.
40. Kolodziejczyk AA, Kim JK, Tsang JC, Ilicic T, Henriksson J, Natarajan KN, et al. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. Cell Stem Cell. 2015;17(4):471-85.
41. Olsson A, Venkatasubramanian M, Chaudhri VK, Aronow BJ, Salomonis N, Singh H, et al. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. Nature. 2016;537(7622):698-702.

View Article Online
DOI: 10.1039/C8LC01239C

42. Valdes-Mora F, Handler K, Law AMK, Salomon R, Oakes SR, Ormandy CJ, et al. Single-Cell Transcriptomics in Cancer Immunobiology: The Future of Precision Oncology. Front Immunol. 2018;9:2582.
43. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014;344(6190):1396-401.
44. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature. 2014;509(7500):371-5.
45. Venteicher AS, Tirosh I, Hebert C, Yizhak K, Neftel C, Filbin MG, et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. Science. 2017;355(6332).
46. Pavlicev M, Wagner GP, Chavan AR, Owens K, Maziarz J, Dunn-Fletcher C, et al. Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. Genome Res. 2017;27(3):349-61.
47. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell. 2018;173(4):879-93 e13.
48. Valdes-Mora F, Lee HJ. Single-Cell Genomics and Epigenomics. Springer, Berlin, Heidelberg. 2016:257-301.
49. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med. 2018;50(8):96.
50. Zappia L, Phipson B, Oshlack A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. PLOS Computational Biology. 2018;14(6):e1006245.
51. Nguyen A, Khoo WH, Moran I, Croucher PI, Phan TG. Single Cell RNA Sequencing of Rare Immune Cell Populations. Front Immunol. 2018;9:1553.
52. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet based single cell RNA sequencing data. bioRxiv. 2018.
53. Lun ATL, Riesenfeld S, T. A, Dao TP, Gomes T, Jamboree pitsHCA, et al. Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. bioRxiv. 2018.
54. Vieira Braga FA, Teichmann SA, Stubbington MJ. Are cells from a snowman realistic? Cryopreserved tissues as a source for single-cell RNA-sequencing experiments. Genome Biol. 2017;18(1):54.
55. Alles J, Karaiskos N, Praktiknjo SD, Grosswendt S, Wahle P, Ruffault PL, et al. Cell fixation and preservation for droplet-based single-cell transcriptomics. BMC Biol. 2017;15(1):44.
56. Chen J, Cheung F, Shi R, Zhou H, Lu W, Consortium CHI. PBMC fixation and processing for Chromium single-cell RNA sequencing. J Transl Med. 2018;16(1):198.
57. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science. 2017;357(6352):661-7.
58. Cheng Z, Wu X, Cheng J, Liu P. Microfluidic fluorescence-activated cell sorting (μFACS) chip with integrated piezoelectric actuators for low-cost mammalian cell enrichment. Microfluid Nanofluid. 2017;21(1):9.
59. Schmid L, Weitz DA, Franke T. Sorting drops and cells with acoustics: acoustic microfluidic fluorescence-activated cell sorter. Lab Chip. 2014;14(19):3710-8.
60. Franke T, Braunmuller S, Schmid L, Wixforth A, Weitz DA. Surface acoustic wave actuated cell sorting (SAWACS). Lab Chip. 2010;10(6):789-94.
61. Wilson Nicola K, Kent David G, Buettner F, Shehata M, Macaulay Iain C, Calero-Nieto Fernando J, et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell. 2015;16(6):712-24.
62. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017;14:865.

63. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell. 2015;163(7):1663-77.
64. Miltenyi S, Müller W, Weichel W, A. R. High gradient magnetic cell separation with MACS. Cytometry. 1990;11:231-8.
65. Suh TK, Schenk JL, Seidel GE, Jr. High pressure flow cytometric sorting damages sperm. Theriogenology. 2005;64(5):1035-48.
66. Llufrio EM, Wang L, Naser FJ, Patti GJ. Sorting cells alters their redox state and cellular metabolome. Redox Biol. 2018;16:381-7.
67. Reece A, Xia B, Jiang Z, Noren B, McBride R, Oakey J. Microfluidic techniques for high throughput single cell analysis. Curr Opin Biotechnol. 2016;40:90-6.
68. Morton KJ, Loutherback K, Inglis DW, Tsui OK, Sturm JC, Chou SY, et al. Hydrodynamic metamaterials: Microfabricated arrays to steer, refract, and focus streams of biomaterials. Proc Natl Acad Sci USA. 2008;105(21):7434-8.
69. Chou CF, Bakajin O, Turner SW, Duke TA, Chan SS, Cox EC, et al. Sorting by diffusion: an asymmetric obstacle course for continuous molecular separation. Proc Natl Acad Sci U S A. 1999;96(24):13762-5.
70. Inglis DW, Davis JA, Austin RH, Sturm JC. Critical particle size for fractionation by deterministic lateral displacement. Lab Chip. 2006;6(5):655-8.
71. Davis JA, Inglis DW, Morton KJ, Lawrence DA, Huang LR, Chou SY, et al. Deterministic hydrodynamics: taking blood apart. Proc Natl Acad Sci U S A. 2006;103(40):14779-84.
72. Huang LR, Cox EC, Austin RH, Sturm JC. Continuous particle separation through deterministic lateral displacement. Science. 2004;304(5673):987-90.
73. Loutherback K, Chou KS, Newman J, J. P, H. AR, C SJ. Improved performance of deterministic lateral displacement arrays with triangular posts. Microfluid Nanofluid. 2010;9(1143).
74. Park S, Zhang Y, Wang TH, Yang S. Continuous dielectrophoretic bacterial separation and concentration from physiological media of high conductivity. Lab Chip. 2011;11(17):2893-900.
75. Vig AL, Kristensen A. Separation enhancement in pinched flow fractionation. Appl Phys Lett. 2008;93(203507).
76. Yamada M, Nakashima M, Seki M. Pinched flow fractionation: continuous size separation of particles utilizing a laminar flow profile in a pinched microchannel. Anal Chem. 2004;76(18):5465-71.
77. Di Carlo D. Inertial microfluidics. Lab Chip. 2009;9(21):3038-46.
78. Kuntaegowdanahalli SS, Bhagat AA, Kumar G, Papautsky I. Inertial microfluidics for continuous particle separation in spiral microchannels. Lab Chip. 2009;9(20):2973-80.
79. Choi S, Park JK. Continuous hydrophoretic separation and sizing of microparticles using slanted obstacles in a microchannel. Lab Chip. 2007;7(7):890-7.
80. Xia N, Hunt TP, Mayers BT, Alsberg E, Whitesides GM, Westervelt RM, et al. Combined microfluidic-micromagnetic separation of living cells in continuous flow. Biomed Microdevices. 2006;8(4):299-308.
81. Wang Y, Zhao Y, Cho SK. Efficient in-droplet separation of magnetic particles for digital microfluidics. J Micromech Microeng. 2007;17(2148).
82. MacDonald MP, Spalding GC, Dholakia K. Microfluidic sorting in an optical lattice. Nature. 2003;426(6965):421-4.
83. Wiklund M, Gunther C, Lemor R, Jager M, Fuhr G, Hertz HM. Ultrasonic standing wave manipulation technology integrated into a dielectrophoretic chip. Lab Chip. 2006;6(12):1537-44.
84. Gascoyne PR, Vykoukal J. Particle separation by dielectrophoresis. Electrophoresis. 2002;23(13):1973-83.
85. Petersson F, Nilsson A, Holm C, Jonsson H, Laurell T. Continuous separation of lipid particles from erythrocytes by means of laminar flow and acoustic standing wave forces. Lab Chip. 2005;5(1):20-2.
86. Shi J, Huang H, Stratton Z, Huang Y, Huang TJ. Continuous particle separation in a microfluidic channel via standing surface acoustic waves (SSAW). Lab Chip. 2009;9(23):3354-9.

View Article Online
DOI: 10.1039/C8LC01239C

87. Johnson DA, Feke DL. Methodology for fractionating suspended particles using ultrasonic standing wave and divided flow fields. Separations Technology. 1995;5(4):251-8.
88. Ng JW, Collins DJ, Devendran C, Ai Y, Neild A. Flow-rate-insensitive deterministic particle sorting using a combination of travelling and standing surface acoustic waves. Microfluidics and Nanofluidics. 2016;20(151).
89. Collins DJ, Alan T, Neild A. Particle separation using virtual deterministic lateral displacement (vDLD). Lab Chip. 2014;14(9):1595-603.
90. Collins DJ, Neild A, Ai Y. Highly focused high-frequency travelling surface acoustic waves (SAW) for rapid single-particle sorting. Lab Chip. 2016;16(3):471-9.
91. Ma Z, Zhou Y, Collins DJ, Ai Y. Fluorescence activated cell sorting via a focused traveling surface acoustic beam. Lab Chip. 2017;17(18):3176-85.
92. Ding X, Lin SC, Lapsley MI, Li S, Guo X, Chan CY, et al. Standing surface acoustic wave (SSAW) based multichannel cell sorting. Lab Chip. 2012;12(21):4228-31.
93. Xi HD, Zheng H, Guo W, Ganan-Calvo AM, Ai Y, Tsao CW, et al. Active droplet sorting in microfluidics: a review. Lab Chip. 2017;17(5):751-71.
94. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods. 2017;14(10):955-8.
95. Chae SK, Lee CH, Lee SH, Kim TS, Kang JY. Oil droplet generation in PDMS microchannel using an amphiphilic continuous phase. Lab Chip. 2009;9(13):1957-61.
96. Collins DJ, Neild A, deMello A, Liu AQ, Ai Y. The Poisson distribution and beyond: methods for microfluidic droplet production and single cell encapsulation. Lab Chip. 2015;15(17):3439-59.
97. Sesen M, Alan T, Neild A. Droplet control technologies for microfluidic high throughput screening (muHTS). Lab Chip. 2017;17(14):2372-94.
98. Hong Z, Ethan T, A. SRM, C. WG, Eugenia K. Exploring Microfluidic Routes to Microgels of Biological Polymers. Macromolecular Rapid Communications. 2007;28(5):527-38.
99. Schmid L, Franke T. SAW-controlled drop size for flow focusing. Lab Chip. 2013;13(9):1691-4.
100. Schmid L, Franke T. Acoustic modulation of droplet size in a T-junction. Applied Physics Letters. 2014;104(13):133501.
101. Abate AR, Agresti JJ, Weitz DA. Microfluidic sorting with high-speed single-layer membrane valves. Applied Physics Letters. 2010;96(20):203509.
102. Collins DJ, Alan T, Helmerson K, Neild A. Surface acoustic waves for on-demand production of picoliter droplets and particle encapsulation. Lab Chip. 2013;13(16):3225-31.
103. Brenker J, Collins D, Van Phan H, Alan T, Neild A. On-chip droplet production regimes using surface acoustic waves2016.
104. Jin SH, Jeong HH, Lee B, Lee SS, Lee CS. A programmable microfluidic static droplet array for droplet generation, transportation, fusion, storage, and retrieval. Lab Chip. 2015;15(18):3677-86.
105. Zhao C-X, Middelberg APJ. Two-phase microfluidic flows. Chem Eng Sci. 2011;66(7):1394-411.
106. Baroud CN, Gallaire F, Dangla R. Dynamics of microfluidic droplets. Lab Chip. 2010;10(16):2032-45.
107. Thorsen T, Roberts RW, Arnold FH, Quake SR. Dynamic pattern formation in a vesicle-generating microfluidic device. Phys Rev Lett. 2001;86(18):4163-6.
108. Anna SL, Bontoux N, Stone HA. Formation of dispersions using “flow focusing” in microchannels. Applied Physics Letters. 2003;82(3):364-6.
109. Dreyfus R, Tabeling P, Willaime H. Ordered and disordered patterns in two-phase flows in microchannels. Phys Rev Lett. 2003;90(14):144505.
110. Adams LLA, Kodger TE, Kim S-H, Shum HC, Franke T, Weitz DA. Single step emulsification for the generation of multi-component double emulsions. Soft Matter. 2012;8(41):10719-24.
111. Leman M, Abouakil F, Griffiths AD, Tabeling P. Droplet-based microfluidics at the femtolitre scale. Lab Chip. 2015;15(3):753-65.

112. Shim JU, Ranasinghe RT, Smith CA, Ibrahim SM, Hollfelder F, Huck WT, et al. Ultrarapid generation of femtoliter microfluidic droplets for single-molecule-counting immunoassays. ACS Nano. 2013;7(7):5955-64.
113. Guo MT, Rotem A, Heyman JA, Weitz DA. Droplet microfluidics for high-throughput biological assays. Lab Chip. 2012;12(12):2146-55.
114. Abate AR, Chen CH, Agresti JJ, Weitz DA. Beating Poisson encapsulation statistics using close-packed ordering. Lab on a Chip. 2009;9(18):2628-31.
115. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, et al. Single-cell barcoding and sequencing using droplet microfluidics. Nature Protocols. 2017;12(1):44-73.
116. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nature Communications. 2017;8.
117. Moon HS, Je K, Min JW, Park D, Han KY, Shin SH, et al. Inertial-ordering-assisted droplet microfluidics for high-throughput single-cell RNA-sequencing. Lab Chip. 2018;18(5):775-84.
118. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques. 2001;30(4):892-7.
119. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. 2010;11(12):R119.
120. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, et al. Single-cell barcoding and sequencing using droplet microfluidics. Nat Protoc. 2017;12(1):44-73.
121. Huang M, Wang J, Torre E, Dueck H, Shaffer S, Bonasio R, et al. SAVER: gene expression recovery for single-cell RNA sequencing. Nat Methods. 2018;15(7):539-42.
122. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell. 2017;65(4):631-43 e4.
123. Zhang X, Li T, Liu F, Chen Y, Yao J, Li Z, et al. Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. Mol Cell. 2019;73(1):130-42 e5.
124. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 2018;174(5):1293-308 e36.
125. Singh M, Al-Eryani G, Carswell S, Ferguson JM, Blackburn J, Barton K, et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. bioRxiv. 2018.
126. Volden R, Palmer T, Byrne A, Cole C, Schmitz RJ, Green RE, et al. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. Proc Natl Acad Sci U S A. 2018;115(39):9726-31.
127. Chen J, Suo S, Tam PP, Han JJ, Peng G, Jing N. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. Nat Protoc. 2017;12(3):566-80.
128. Lovatt D, Ruble BK, Lee J, Dueck H, Kim TK, Fisher S, et al. Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. Nat Methods. 2014;11(2):190-6.
129. Medaglia C, Giladi A, Stoler-Barak L, De Giovanni M, Salame TM, Biram A, et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. Science. 2017;358(6370):1622-6.
130. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, et al. Highly multiplexed subcellular RNA sequencing in situ. Science. 2014;343(6177):1360-3.
131. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. Single-cell in situ RNA profiling by sequential hybridization. Nat Methods. 2014;11(4):360-1.
132. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015;348(6233):aaa6090.

View Article Online
DOI: 10.1039/C8LC01239C

133. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science. 2016;353(6294):78-82.
134. Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Äijö T, et al. High-density spatial transcriptomics arrays for in situ tissue profiling. BioRxiv. 2019.
135. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A Scalable Technology for Measuring Genome-Wide Expression at High Spatial Resolution. bioRxiv. 2019.
136. Salmen F, Stahl PL, Mollbrink A, Navarro JF, Vickovic S, Frisen J, et al. Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. Nat Protoc. 2018;13(11):2501-34.
137. Wootton RC, Demello AJ. Microfluidics: Analog-to-digital drug screening. Nature. 2012;483(7387):43-4.
138. Chen CH, Miller MA, Sarkar A, Beste MT, Isaacson KB, Lauffenburger DA, et al. Multiplexed protease activity assay for low-volume clinical samples using droplet-based microfluidics and its application to endometriosis. J Am Chem Soc. 2013;135(5):1645-8.
139. Miller OJ, El Harrak A, Mangeat T, Baret JC, Frenz L, El Debs B, et al. High-resolution dose-response screening using droplet-based microfluidics. Proc Natl Acad Sci U S A. 2012;109(2):378-83.
140. Clausell-Tormos J, Lieber D, Baret JC, El-Harrak A, Miller OJ, Frenz L, et al. Droplet-based microfluidic platforms for the encapsulation and screening of Mammalian cells and multicellular organisms. Chem Biol. 2008;15(5):427-37.
141. Brouzes E, Medkova M, Savenelli N, Marran D, Twardowski M, Hutchison JB, et al. Droplet microfluidic technology for single-cell high-throughput screening. Proc Natl Acad Sci U S A. 2009;106(34):14195-200.
142. Mazutis L, Gilbert J, Ung WL, Weitz DA, Griffiths AD, Heyman JA. Single-cell analysis and sorting using droplet-based microfluidics. Nat Protoc. 2013;8(5):870-91.
143. Piyasena ME, Graves SW. The intersection of flow cytometry with microfluidics and microfabrication. Lab Chip. 2014;14(6):1044-59.
144. Griffiths JA, Scialdone A, Marioni JC. Using single-cell genomics to understand developmental processes and cell fate decisions. Mol Syst Biol. 2018;14(4):e8046.
145. Shahi P, Kim SC, Haliburton JR, Gartner ZJ, Abate AR. Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. Sci Rep. 2017;7:44447.
146. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods. 2015;12(6):519-22.
147. Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet. 2019.
148. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. 2017;27(3):491-9.

View Article Online
DOI: 10.1039/C8LC01239C

Legends to figures and supplementary materials:

Fig. 1. Single cell RNAseq sample journey. Schematic representation of a typical workflow for an scRNA-seq experiment.

Fig. 2. Effects of sorting on Drop-seq. Reverse transcriptase cDNA product form equivalent numbers of barcoded beads after different capture conditions. Two types of samples were tested, mouse mammary tumours **(A)** and human PBMCs **(B).**

Fig. 3. Droplet formation at T-junction. Schematic representation of the modalities for droplet formation using microfluidic systems.

Fig. 4. Theoretical modelling of particles occupancy in droplet-based microfluidic systems. A) Simulation of cell yield and multiplet frequency assuming Poisson encapsulation statistics. The x and left y axes (black) show the average number of cells per drop and the percentage of cells that are barcoded (yield), respectively. The percentage of cells that are barcoded depends on the number of beads per drop (see legend) and is independent of the number of cells per drop. The right y axis (blue) shows the percentage of droplets containing cells that contain multiplets, which depends on both the number of beads and cells per drop. If there is deterministic bead sorting into droplets at a 1:1 ratio (green line) the frequency of multiplets is dramatically reduce because there are only cell multiplets. **B)** Barcoding yield versus the incidence of multiplexes. The x and y axes are the percentage yield of cells that are barcoded, and the proportion of barcoded cells that form multiplets, respectively, assuming both beads and cells are randomly allocated to droplets according to the Poisson model.

Fig. 5. Molecular design and workflow of the different droplet-based scRNA-seq modalities. A) Drop-seq/ DroNc-seq. **B)** Chromium **C)** InDrop.

Fig. 6. scRNA-seq platform performance comparison. Qualitative analysis of key features characteristic of the droplet-based scRNA-seq platforms.

View Article Online
DOI: 10.1039/C8LC01239C

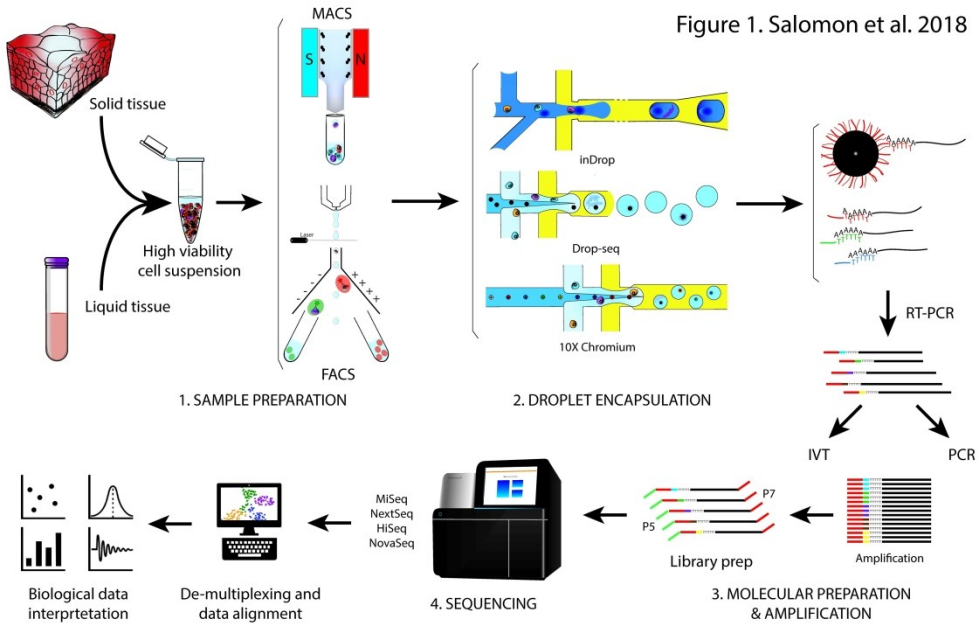

Fig. 1. Single cell RNAseq sample journey. Schematic representation of a typical workflow for an scRNAseq experiment.

345x225mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C8LC01239C

Figure 2. Salomon et al. 2018

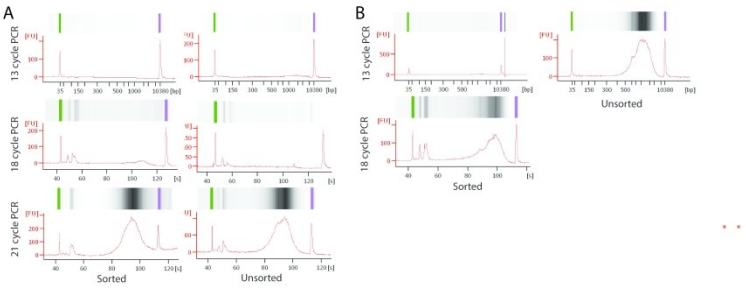

Fig. 2. Effects of sorting on DropSeq (BOX 1). Reverse transcriptase cDNA product form equivalent numbers of barcoded beads after different capture conditions. Two types of samples were tested, mouse mammary tumours (A) and human PBMCs (B).

View Article Online
DOI: 10.1039/C8LC01239C

Figure 3. Salomon et al. 2018

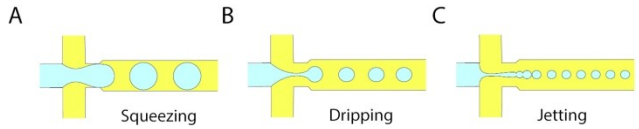

Fig. 3. Droplet formation at T-junction (BOX 2). Schematic representation of the modalities for droplet formation using microfluidic systems.

View Article Online
DOI: 10.1039/C8LC01239C

Figure 4. Salomon et al. 2018

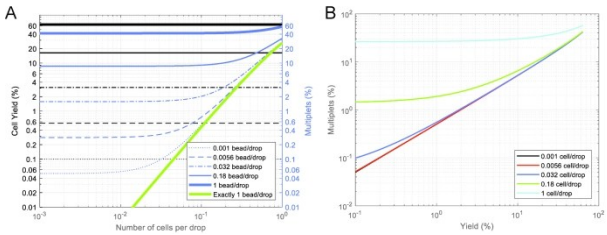

Fig. 4. Theoretical modelling of particles occupancy in droplet-based microfluidic systems. A) Simulation of cell yield and multiplet frequency assuming Poisson encapsulation statistics. The x and left y axes (black) show the average number of cells per drop and the percentage of cells that are barcoded (yield), respectively. There percentage of cells that are barcoded depends on the number of beads per drop (see legend) and is independent of the number of cells per drop. The right y axis (blue) shows the percentage of droplets containing cells that contain multiplets, which depends on both the number of beads and cells per drop. If there is deterministic bead sorting into droplets at a 1:1 ratio (green line) the frequency of multiplets is dramatically reduce because there are only cell multiplets. B) Barcoding yield versus the incidence of multiplexes. The x and y axes are the percentage yield of cells that are barcoded, and the proportion of barcoded cells that form multiplets, respectively, assuming both beads and cells are randomly allocated to droplets according to the Poisson model.

View Article Online
DOI: 10.1039/C8LC01239C

Figure 5. Salomon et al. 2018

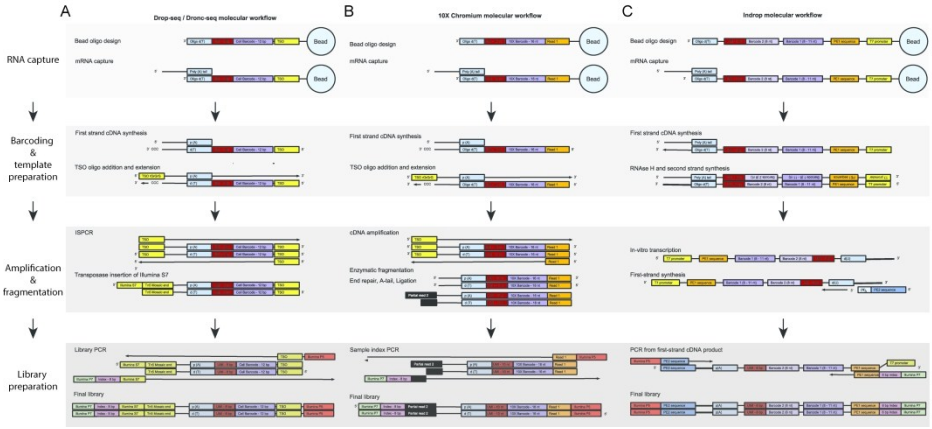

Fig. 5. Molecular design and workflow of the different droplet-based scRNAseq modalities. A) Drop-seq/ DroNc-seq. B) Chromium 10x. C) InDrop.

View Article Online
DOI: 10.1039/C8LC01239C

Figure 7. Salomon et al. 2018

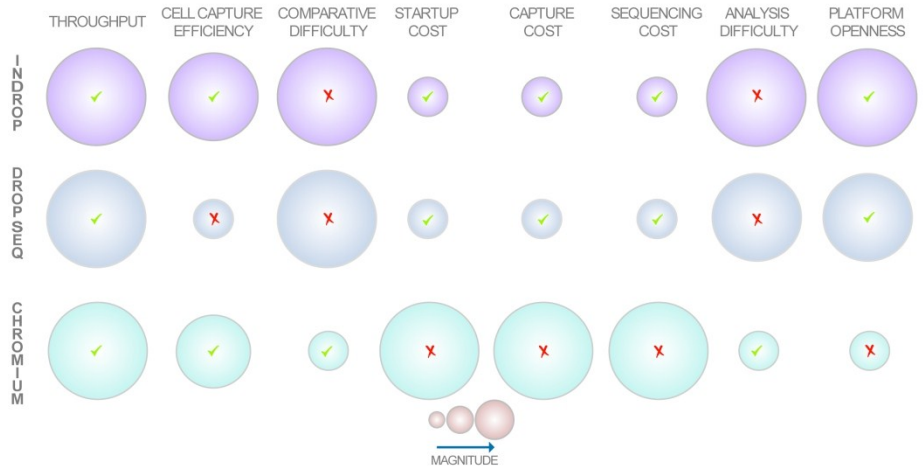

Fig. 6. scRNAseq platform performance comparison. Qualitative analysis of key features characteristic of the droplet-based scRNAseq platforms.

View Article Online
DOI: 10.1039/C8LC01239C

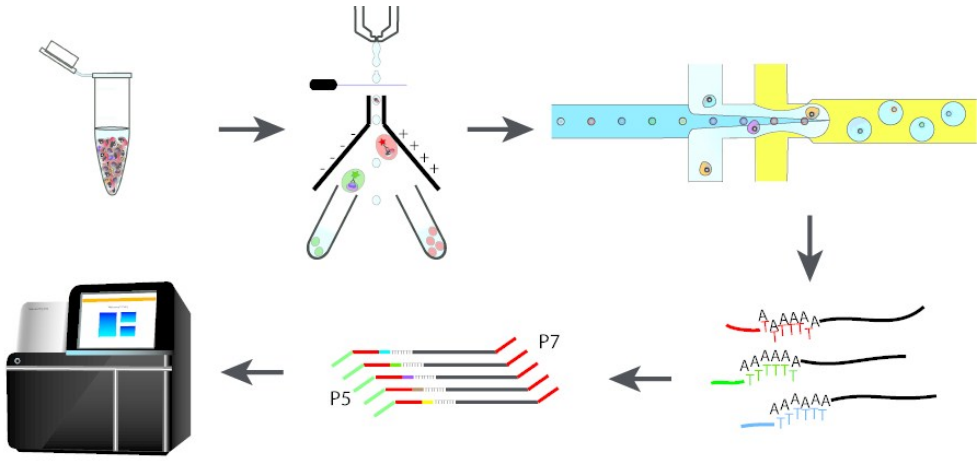

Table of contents figure