Exponential scaling of single-cell RNA-seq in the past decade

Valentine Svensson[1,2], Roser Vento-Tormo[2] & Sarah A Teichmann[1,2]

Measurement of the transcriptomes of single cells has been feasible for only a few years, but it has become an extremely popular assay. While many types of analysis can be carried out and various questions can be answered by single-cell RNA-seq, a central focus is the ability to survey the diversity of cell types in a sample. Unbiased and reproducible cataloging of gene expression patterns in distinct cell types requires large numbers of cells. Technological developments and protocol improvements have fueled consistent and exponential increases in the number of cells that can be studied in single-cell RNA-seq analyses. In this Perspective, we highlight the key technological developments that have enabled this growth in the data obtained from single-cell RNA-seq experiments.

Many biological discoveries have been made possible by the development of new methods and technological progress. For example, the improvements made to microscopes in the 17th century allowed Robert Hooke and Anton van Leeuwenhoek to describe the cell as the structural unit of life for the very first time[1]. Since then, many studies have focused on cell characterization, redefining the cell as not only the structural but also the functional unit of life[2].

Cells have traditionally been classified according to their morphology or the expression patterns of certain proteins in functionally distinct settings[3–5]. Changes in cellular activity and identity are reflections of distinct gene expression programs[6,7]. Proteomic technologies have the advantage of assaying the final functional product of gene expression, but at the single-cell level, proteomics assays are constrained to a limited, pre-selected repertoire of molecules, which precludes an unbiased, comprehensive analysis of cell phenotypes. Assay of the transcriptome (the abundance of all transcribed RNAs) in a single cell with methodologies such as RNA-seq provides an alternative means to classify and characterize cells at the molecular level[8,9].

All single-cell RNA-seq protocols share a common initial step in which transcribed RNA from cells is converted to cDNA. The next step is an amplification by molecular biology methods such as polymerase chain reaction (PCR) and *in vitro* transcription (IVT). The subsequent steps, culminating in sequencing, allow the expression level of gene products to be quantified.

[1]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, UK. [2]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. Correspondence should be addressed to V.S. (vale@ebi.ac.uk) or S.A.T. (st9@sanger.ac.uk).

Received 21 March 2017; accepted 29 November 2017; published online 1 March 2018; corrected before print 22 March 2018; doi:10.1038/nprot.2017.149

Eberwine *et al.*[10] measured the expression of a handful of individual genes from single cells for the first time in 1992, using a sophisticated approach based on *in vivo* reverse transcription (RT) followed by amplification through IVT. Later, simpler, PCR-based methods emerged[11] that led to scaling up of the number of cells and genes assayed over the years[12,13]. Eventually, untargeted single-cell mRNA (or cDNA) amplification techniques were developed, which allowed researchers to perform transcriptome-wide studies using microarrays[14–17]. Building upon this, Tang *et al.* adapted the technologies to make them compatible with high-throughput DNA sequencing, thus allowing completely unbiased transcriptome-wide investigation of the mRNA in a single cell for the first time[18].

Early single-cell experiments were motivated by the prospect of an in-depth analysis of gene expression in a few precious cells[18–22]. A shift in the field came when Guo *et al.* demonstrated that distinct cell types could be identified without pre-sorting: the authors carried out RT-qPCR of 48 genes in parallel on more than 500 cells[23], demonstrating the utility of measuring larger numbers of cells. This inspired the Linnarsson group[24] to develop methods explicitly targeted at assaying many cells in parallel via multiplexed unbiased RNA-seq, with the long-term goal of eventually cataloging all neuronal cell types[25,26].

In the past few years, many sensitive and accurate single-cell RNA-sequencing (scRNA-seq) protocols have been introduced[27]. Here we focus on the increase in the number of cells profiled per study, a key factor in the ability to catalog cell types. We review the exponential growth of scale in single-cell transcriptomics experiments, from tens of cells up to hundreds of thousands of single cells per study in less than a decade, and discuss the technological advances that have enabled this (**Fig. 1**, **Supplementary Table 1**).

Many different incremental improvements have contributed to increases in scale, such as, for example, decreases in required

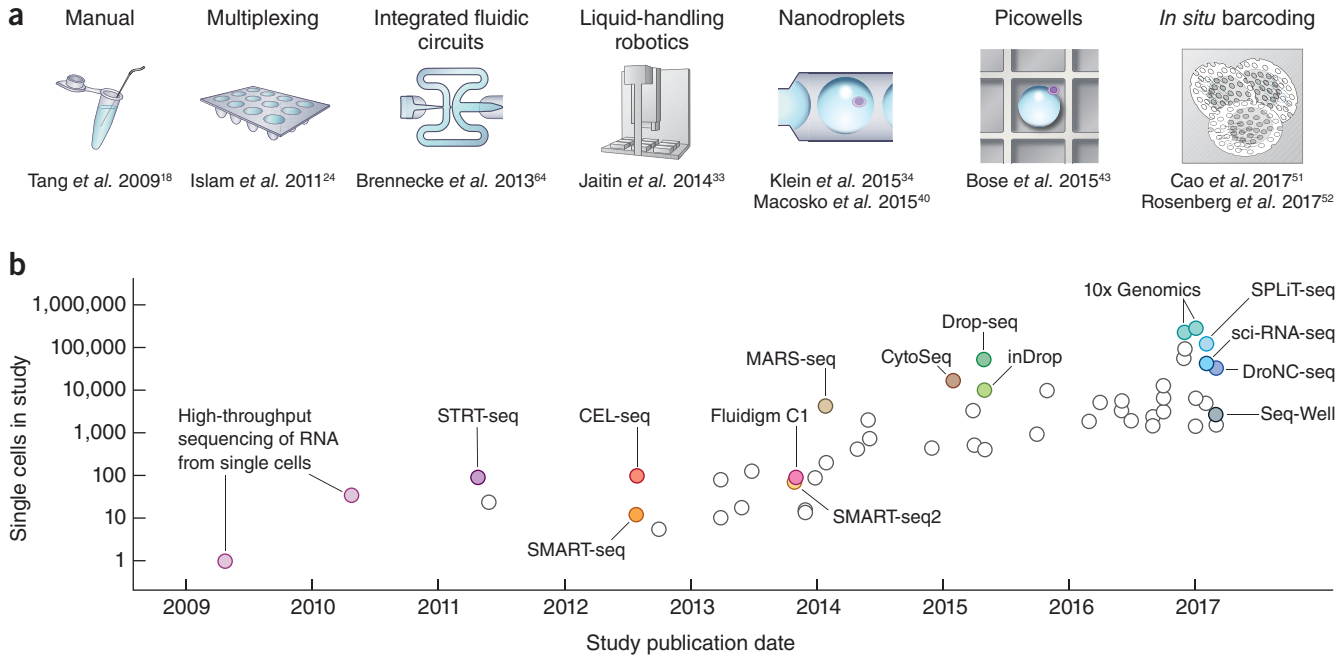

Figure 1 | Scaling of scRNA-seq experiments. (**a**) Key technologies that have allowed jumps in experimental scale. A jump to ~100 cells was enabled by sample multiplexing, and then a jump to ~1,000 cells was achieved by large-scale studies using integrated fluidic circuits, followed by a jump to several thousands of cells with liquid-handling robotics. Further orders-of-magnitude increases bringing the number of cells assayed into the tens of thousands were enabled by random capture technologies using nanodroplets and picowell technologies. Recent studies have used *in situ* barcoding to inexpensively reach the next order of magnitude of hundreds of thousands of cells. (**b**) Cell numbers reported in representative publications by publication date. Key technologies are indicated. A full table with corresponding numbers is available as **Supplementary Table 1**.

reagent volumes and consumable costs[9,28] facilitated by the introduction of microfluidic technologies, random capture methods and *in situ* barcoding (**Fig. 1a**). In this Perspective we focus on solutions for three key challenges: untargeted amplification of whole transcriptomes from single cells, automatic isolation of cells and the ability to process many cells in parallel.

Untargeted amplification of whole transcriptomes from single cells

For successful signal detection with RNA-seq, on the order of 0.1–1.0 µg of total RNA is needed (see, for example, information from the Truseq and NEBNext kits: https://support.illumina.com/sequencing/sequencing_kits/truseq_rna_sample_prep_kit/input_req.html; https://www.neb.com/faqs/2012/11/19/what-is-the-starting-material-i-need-to-use-when-preparing-libraries-using-the-nebnext-ultra-dire). The amount of RNA present in a single cell is limited, and ranges from 1–50 pg depending on cell type[29]. One way to overcome this problem is to convert RNA from a single cell into cDNA and amplify it before a sequencing library is created. (DNA-sequencing kits typically require on the order of 1 ng of DNA; see, for example, https://support.illumina.com/sequencing/sequencing_kits/nextera_xt_dna_kit/input_req.html). For this to be achieved, adaptor sequences need to be added to all mRNA transcripts. Delivery of these adaptors without pre-specification of target sequences of particular genes was the main technical challenge in early single-cell RNA studies.

Amplification of cDNA can be exponential, by PCR amplification, or linear, via multiple rounds of IVT, with each requiring different strategies for adaptor addition. PCR amplification requires the addition of adaptor sequences at both ends of the double-stranded cDNA. Most techniques use the poly(A) tail of the mRNA to generate the first strand of cDNA, by initiating RT using a poly(T)-oligonucleotide that also contains a universal adaptor sequence. The second adaptor can then be incorporated during the cDNA-amplification step via several different strategies. In the homopolymer tailing approach, a transferase adds a poly(A) tail to the 3′ end of the first-strand cDNA. Subsequently, a poly(T) primer with a different universal anchor is incorporated into the second strand. The two different adaptor sequences are then used for PCR amplification. This protocol was first applied to single-cell analysis in 2002 (ref. 30) and eventually was adapted for use with microarrays[15,16] and scRNA-seq[18]. Although long cDNAs have been transcribed by this approach[18], a drawback of the method is the reduction of 5′ transcript coverage owing to premature termination of the RT. Template-switching PCR is an alternative strategy that ensures the full transcription of the RNA. This method is based on the intrinsic ability of the transcriptases of the Moloney murine leukemia virus to add a small number of nucleotides (mostly cytosines) when the RT reaches the end of the mRNA[31]. The addition of a helper oligonucleotide containing complementary nucleotides and the second adaptor allows the polymerase to automatically initiate second-strand synthesis without requiring homopolymer tailing. In the single-cell tagged reverse-transcription sequencing (STRT-seq) method[24], full-length cDNA is amplified by template switching,

but only the 5′-end fragment is captured and sequenced. By contrast, fragments from the full-length cDNA are sequenced in the SMART-seq method[21].

The second option, IVT, is the approach that was originally used to pioneer single-cell analysis of the transcriptome[10]. The CEL-seq (cell expression by linear amplification and sequencing) protocol leveraged this for linear mRNA amplification from single cells. Here the RT adaptor also contains a T7 promoter sequence, which allows the final double-stranded DNA to be transcribed into multiple copies of antisense RNA[32]. Once enough amplified antisense RNA has been produced, it can be fragmented and again reverse-transcribed to cDNA. This method was also used for MARS-seq (massively parallel scRNA-seq)[33], and later for InDrop (droplet sequencing)[34].

Simplified, automated cell isolation

The earliest single-cell experiments analyzed only a handful of cells that were each manually isolated in single tubes before RNA was extracted via cell lysis. With improvements in the methodology, larger numbers of cells could be analyzed in parallel, and microtiter plates were used for the new methods. Although some devices such as custom-built semiautomatic cell pickers[24] were designed, an easy-to-implement alternative was to use fluorescence-activated cell sorting (FACS) to isolate cells into microwell plates.

FACS has frequently been used to increase the scale of single-cell sequencing. Some methods have used specialized adaptations of FACS: the MASC-seq method makes use of a spatially barcoded poly(dT) array which cells are sorted onto[35]. The recently published nanoliter-volume microwell array platform used for the STRT-seq-2i protocol also takes advantage of FACS to load cells into miniscule wells in custom capture plates.

A further increase in scale was achieved through robotic automation after cell capture[33,36]. However, specialized flow cytometers are expensive and require dedicated staff. Different forms of passive and random cell-capture technologies were developed to reduce the need for such instrumentation.

The first commercial tool for passive cell capture for scRNA-seq was the microfluidic C1 system released by Fluidigm. The company had expertise in producing systems for parallel RT-qPCR (https://www.fluidigm.com/about/aboutfluidigm) in single cells, and adapted the technology for sequencing. The key feature of the Fluidigm technology is the design of microfluidics devices (or chips) that allow the sequential delivery of very small and precise volumes into tiny reaction chambers. Cells are loaded onto the chip and are passively captured in (up to) 96 isolated chambers in about half an hour. Several steps of the SMART-seq protocol can be carried out within the chambers of the chip before cDNA is extracted from the chip and deposited in microwell plates for the generation of a sequencing library. Complex single-use integrated fluidic circuit (IFC) chips are fundamentally limited to a set number of chambers for capturing cells (96 in the current version), and in many cases only a fraction of the chambers successfully capture cells, depending on the cell type[25,37]. Some large-scale studies made use of a large number of IFCs to create big data sets[25,38].

More recently, methods emerged to randomly capture and manipulate individual cells in nanoliter droplet emulsions[39]. The inDrop and Drop-seq protocols both present strategies to isolate cells in droplets and carry out barcoded cDNA preparation within each droplet[34,40]. Here, two flows of liquid—one containing reagents (including lysis buffer and reverse transcriptase) and beads with poly(T) RT primers, and the other containing cells in buffer—are merged into a combined flow. This flow is separated into droplets by the addition of oil at set intervals. By calibrating the relative rate of the two flows and controlling the creation of droplets, a user can ensure that in most cases only single cells will be isolated in droplets by Poisson statistics. Because of the random nature of this process, it requires a large number of cells and thus is not suited to samples with limited availability of cells.

In an alternative strategy, cells can be deposited at random into picoliter wells that contain barcoded beads and reagents[41–43]. The same beads as for droplet protocols can be used to deliver RT primers into individual wells, but no pumps or other microfluidic equipment is needed. Instead, cells are isolated into the picoliter wells by gravity. Each plate has an upper limit of cells determined by the number of wells, and Poisson statistics are used to calculate the limiting dilution required to avoid doublets, resulting in only a small fraction of the wells containing a single cell.

Barcoding strategies for efficient multiplexing

"Multiplexing" refers to the practice of processing and analyzing multiple samples at once. In sequencing assays this is achieved through the addition of molecular barcodes to cDNA fragments. This way, material from many cells can be pooled, which allows subsequent steps to be performed in a single tube. When the multiplexing step is performed early in the workflow, the amount of labor and material needed to create a sequencing library is dramatically reduced. It also enables the use of technologies that require a minimal input level of material, such as IVT[44]. Furthermore, libraries from many cells can be sequenced together to spread the sequencing read depth over many cells.

Islam *et al.* published the first single-cell RNA-seq protocol for multiplexing cells from a single 96-well plate by using a unique template-switching oligo (TSO) in each well via their STRT-seq method[24] (a later version of STRT-seq removed early multiplexing in favor of passive cell capture with the Fluidigm C1 platform). In early multiplexing strategies, a unique barcode for each isolated cell was incorporated into the cDNA fragments before library generation, in either the TSO[24] or the RT adaptor[40,44,45], or during full-length PCR[46].

Furthermore, different multiplexing approaches can be combined through the use of multiple barcoding steps. In MARS-seq, a variant of the CEL-seq protocol, an additional barcode is added during library preparation to increase the number of cells that can be sequenced in parallel. A similar strategy was introduced in the STRT-Seq-2i protocol[46]—here up to 100 barcodes are added during full-length PCR, followed by 96 additional barcodes for different regions of their array during library preparation, allowing a total of 9,600 cells to be multiplexed and sequenced together.

Early multiplexing is critical for random cell-isolation methods because no reagents can be added to the individual cell lysates after the cells have been captured. To allow labeling of cDNA from cells isolated in individual droplets (or picoliter wells), the beads with RT primers also contain a barcode. RT and barcoding can therefore happen in each individual isolate. After this, material can be treated as in other protocols. For example, InDrop uses the same IVT protocol as CEL-seq on the barcoded material[34], whereas, for example, Drop-seq and Seq-Well use PCR-based methods for multiplexed amplification of barcoded cDNA[40,41].

To increase the number of cells that can be multiplexed, a greater number of barcodes is required, which necessitates the use of longer oligos at prohibitive synthesis costs. This can be avoided by two different strategies: the combination of multiple shorter designed barcodes (e.g., 8–10 bases) into longer barcodes (e.g., 8 bases + 22-base linker + 10 bases = 40 bases), or the synthesis of very long (e.g., 12 bases) random barcodes.

The first, combinatorial approach was first described with the targeted CytoSeq method. The authors used initial pools of 96 barcodes into which beads were split-pooled three times to generate a set of 96^3 = 884,736 barcoded beads[42]. A similar method called 'mix & expand' was later published by which the authors created 96^2 = 9,216 barcoded beads[43]. Similarly, the InDrop method was used to barcode hydrogel beads through combinatorial ligation of two sets of 384 barcodes, creating 384^2 = 147,456 unique barcoded beads[34]. This approach is used by the commercial Illumina SureCell system, with 96^3 = 884,736 unique beads (https://support.illumina.com/sequencing/sequencing_kits/surecell-wta-3-kit.html).

The second approach was taken in the Drop-seq and Seq-Well protocols, where the authors used 12 rounds of split-pool single-base DNA synthesis on the beads to generate 4^{12} = 16.7 million barcodes[40,41]. This procedure is simpler than the first approach, and does not require any synthesized oligos for the barcodes. However, in the first approach barcodes can be designed to avoid biases[47] and ensure that barcode sequences will be distinct. In order for potential barcode collisions to be avoided, randomly synthesized barcodes need to be longer, which decreases the probability that similar barcodes will get assigned to droplets with cells in them.

The Poisson statistics of cell capture needed to ensure that mostly single cells are isolated during random cell-isolation methods means there will always be large inefficiencies in terms of cell isolation, and the pool of barcodes will always have to be substantially larger than the number of cells captured to avoid barcode duplication. A remarkable strategy that allows a user to avoid both of these pitfalls is combinatorial *in situ* barcoding. This technique was initially devised for single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq)[48] and later was adapted for whole-genome sequencing[49] and single-cell Hi-C[50]. Recently, these concepts were adapted to scRNA-seq in the single-cell combinatorial-indexing RNA-seq[51] and split-pool ligation-based transcriptome sequencing[52] methods. Here single cells are never individually isolated and lysed; instead, cells are fixed and the mRNA is manipulated *in situ* inside each cell. Cells are split into mini-pools of ~10–100 cells by FACS, distributed over multiwell plates with unique barcodes in each well. First-strand synthesis labels all cells in the well with a first barcode. Cells are then pooled and again randomly split into mini-pools in plate wells by FACS, and a second well-specific barcode is then added. This procedure can be repeated *ad infinitum* before final pooling of the cells to amplify the material and create a sequencing library. This results in an arbitrarily low probability that any two cells will co-locate in the same sequence of wells, and so RNA from each cell is uniquely labeled.

Unlike in isolation-based methods, the number of potential labeled cells in an experiment can be exponentially scaled through increasing numbers of barcoding rounds. This technology was also reported to be compatible with single-nucleus sequencing[52]. Developments such as this hold the promise of enabling very large studies in the future, such as the cataloging of *Caenorhabditis elegans* cells described in ref. 51.

Current limitations and future directions

The isolation and handling of individual cells have become much simpler in recent years. However, a remaining limitation is the requirement for large volumes of cell suspensions. Using the earlier, labor-intensive manual isolation methods, one can ensure that as much cellular material as possible is used, but when cells are automatically and randomly isolated in droplets or wells, only a fraction of the cells in the suspension will be captured. (In principle one can overcome this by massively expanding the pool of barcoded beads and decreasing the cell capture speed, but this entails significant increases in cost.)

For certain tissues it can also be difficult to make single-cell suspensions, and storage of cell suspensions or tissues can sometimes prove challenging. Recent work has proposed moving to sequencing of individual nuclei to simplify the creation of suspensions, as nuclei can be extracted from tissues (including frozen or fixed samples) by Dounce homogenization followed by purification to create a clean suspension, with positive results[46,53–55].

The cost of sequencing is also still prohibitive for studies that involve hundreds of thousands of cells, even at low sequencing depths. The new Illumina NovaSeq 6000 sequencing instrument promises a 45% reduction in sequencing cost for large studies (https://www.illumina.com/systems/sequencing-platforms/novaseq.html), but a radical change in sequencing technology might be needed to bring costs down further.

An important aspect of the cellular heterogeneity of tissues is the spatial context of cells[56]. When a suspension of cells is created, information about this is lost. Recently, a method based on CEL-seq was created to retain the spatial location of mRNA expression in thin tissue slices[57], although not with single-cell resolution. Other methods based on single-molecule fluorescent *in situ* hybridization have been scaled up to measure the expression of up to 1,000 predefined target genes in tens of thousands of single cells by imaging[58–60]. Recent promising work has shown untargeted sequencing of cDNA from mRNA inside cells via the fluorescent *in situ* sequencing method[61,62]. The ability to also investigate the spatial context of gene expression with single-cell RNA-seq will

enable scientists to understand cellular heterogeneity, identify genes important for tissue structure[63] and form a technology complementary to highly multiplexed sequencing of cell suspensions.

Conclusion

In this Perspective we have discussed the scaling up of RNA-seq experiments over the past decade, which has permitted a shift from the study of individual or small numbers of cells to the study of whole organisms and their cell-type compositions. It is now standard practice for published papers to investigate tens of thousands of cells from a number of experimental systems or tissues, with the largest studies consisting of hundreds of thousands of cells. Thanks to these advances in scRNA-seq technology, researchers are taking steps toward more fully understanding what people are made of, by reproducibly cataloging the diversity of cell types and gene expression patterns within them.

ACKNOWLEDGMENTS We thank Z.-S. Chong, K.N. Natarajan, T. Gomes, N. Edner and K. Meyer for reading the manuscript and giving helpful feedback. R.V.-T. is supported by an EMBO Long-Term Fellowship and a Human Frontier Science Program Long-Term Fellowship.

AUTHOR CONTRIBUTIONS V.S., R.V.-T. and S.A.T. wrote the manuscript. V.S. prepared **Figure 1**.

COMPETING INTERESTS The authors declare no competing interests.

1. Gest, H. The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society. *Notes Rec. R. Soc. Lond.* **58**, 187–201 (2004).
2. Arendt, D. *et al.* The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
3. Poulin, J.-F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J.M. & Awatramani, R. Disentangling neural cell diversity using single-cell transcriptomics. *Nat. Neurosci.* **19**, 1131–1141 (2016).
4. Mosmann, T.R., Cherwinski, H., Bond, M.W., Giedlin, M.A. & Coffman, R.L. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J. Immunol.* **136**, 2348–2357 (1986).
5. Orkin, S.H. Diversification of haematopoietic stem cells to specific lineages. *Nat. Rev. Genet.* **1**, 57–64 (2000).
6. Zhu, J. Transcriptional regulation of Th2 cell differentiation. *Immunol. Cell Biol.* **88**, 244–249 (2010).
7. Ivanov, I.I., Zhou, L. & Littman, D.R. Transcriptional regulation of Th17 cell differentiation. *Semin. Immunol.* **19**, 409–417 (2007).
8. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res.* **25**, 1491–1498 (2015).
9. Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C. & Teichmann, S.A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
10. Eberwine, J. *et al.* Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. USA* **89**, 3010–3014 (1992).
11. Lambolez, B., Audinat, E., Bochet, P., Crépel, F. & Rossier, J. AMPA receptor subunits expressed by single Purkinje cells. *Neuron* **9**, 247–258 (1992).
12. Peixoto, A., Monteiro, M., Rocha, B. & Veiga-Fernandes, H. Quantification of multiple gene expression in individual cells. *Genome Res.* **14**, 1938–1947 (2004).
13. Sheng, H.Z., Lin, P.X. & Nelson, P.G. Analysis of multiple heterogeneous mRNAs in single cells. *Anal. Biochem.* **222**, 123–130 (1994).
14. Tietjen, I. *et al.* Single-cell transcriptional analysis of neuronal progenitors. *Neuron* **38**, 161–175 (2003).
15. Kurimoto, K. *et al.* An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.* **34**, e42 (2006).
16. Kurimoto, K., Yabuta, Y., Ohinata, Y. & Saitou, M. Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nat. Protoc.* **2**, 739–752 (2007).
17. Esumi, S. *et al.* Method for single-cell microarray analysis and application to gene-expression profiling of GABAergic neuron progenitors. *Neurosci. Res.* **60**, 439–451 (2008).
18. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
19. Tang, F. *et al.* Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* **6**, 468–478 (2010).
20. Tang, F. *et al.* Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS One* **6**, e21208 (2011).
21. Ramsköld, D. *et al.* Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
22. Brouilette, S. *et al.* A simple and novel method for RNA-seq library preparation of single cell cDNA analysis by hyperactive Tn5 transposase. *Dev. Dyn.* **241**, 1584–1590 (2012).
23. Guo, G. *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* **18**, 675–685 (2010).
24. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167 (2011).
25. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
26. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).
27. Svensson, V. *et al.* Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14**, 381–387 (2017).
28. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
29. Boon, W.C. *et al.* Increasing cDNA yields from single-cell quantities of mRNA in standard laboratory reverse transcriptase reactions using acoustic microstreaming. *J. Vis. Exp.* **3144**, e3144 (2011).
30. Klein, C.A. *et al.* Combined transcriptome and genome analysis of single micrometastatic cells. *Nat. Biotechnol.* **20**, 387–392 (2002).
31. Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R. & Siebert, P.D. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* **30**, 892–897 (2001).
32. Baugh, L.R., Hill, A.A., Brown, E.L. & Hunter, C.P. Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.* **29**, E29 (2001).
33. Jaitin, D.A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
34. Klein, A.M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
35. Vickovic, S. *et al.* Massive and parallel expression profiling using microarrayed single-cell sequencing. *Nat. Commun.* **7**, 13182 (2016).
36. Muraro, M.J. *et al.* A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* **3**, 385–394 (2016).
37. Lönnberg, T. *et al.* Single-cell RNA-seq and computational analysis using temporal mixture modeling resolves TH1/TFH fate bifurcation in malaria. *Sci. Immunol.* **2,** eaal2192 (2017).
38. Shalek, A.K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
39. Mazutis, L. *et al.* Single-cell analysis and sorting using droplet-based microfluidics. *Nat. Protoc.* **8**, 870–891 (2013).
40. Macosko, E.Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
41. Gierahn, T.M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
42. Fan, H.C., Fu, G.K. & Fodor, S.P.A. Combinatorial labeling of single cells for gene expression cytometry. *Science* **347**, 1258367 (2015).
43. Bose, S. *et al.* Scalable microfluidics for single-cell RNA printing and sequencing. *Genome Biol.* **16**, 120 (2015).
44. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
45. Zheng, G.X.Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
46. Hochgerner, H. *et al.* STRT-seq-2i: dual-index 5′ single cell and nucleus RNA-seq on an addressable microwell array. Preprint at https://www.biorxiv.org/content/early/2017/04/20/126268 (2017).

47. Costea, P.I., Lundeberg, J. & Akan, P. TagGD: fast and accurate software for DNA Tag generation and demultiplexing. *PLoS One* **8**, e57521 (2013).
48. Cusanovich, D.A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
49. Vitak, S.A. *et al.* Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat. Methods* **14**, 302–308 (2017).
50. Ramani, V. *et al.* Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
51. Cao, J. *et al.* Comprehensive single cell transcriptional profiling of a multicellular organism by combinatorial indexing. Preprint at https://www.biorxiv.org/content/early/2017/02/02/104844 (2017).
52. Rosenberg, A.B. *et al.* Scaling single cell transcriptomics through split pool barcoding. Preprint at https://www.biorxiv.org/content/early/2017/02/02/105163 (2017).
53. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
54. Lake, B.B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
55. Lake, B.B. *et al.* A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **7**, 6031 (2017).
56. Lee, J.H. Quantitative approaches for investigating the spatial context of gene expression. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **9**, e1369 (2017).
57. Ståhl, P.L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
58. Moffitt, J.R. *et al.* High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proc. Natl. Acad. Sci. USA* **113**, 14456–14461 (2016).
59. Moffitt, J.R. *et al.* High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. USA* **113**, 11046–11051 (2016).
60. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
61. Lee, J.H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).
62. Lee, J.H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
63. Svensson, V., Teichmann, S.A., & Stegle, O. Spatial DE: identification of spatially variable genes. Preprint at https://www.biorxiv.org/content/early/2017/11/08/143321 (2017).
64. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).

Erratum: Exponential scaling of single-cell RNA-seq in the past decade

Valentine Svensson, Roser Vento-Tormo & Sarah A Teichmann

Nat. Protoc. **13**, 599–604 (2018); published online 01 March 2018; corrected before print 22 March 2018

In the version of this article initially published, the sentence: "This method was also used for MARS-seq (massively parallel scRNA-seq)[33], and later for InDrop (droplet sequencing)[34] and Seq-Well (a well-array-based variation on MARS-seq)[35]" was incorrect. The words "and Seq-Well (a well-array-based variation on MARS-seq)[35]" should have been deleted, and as a result, refs. 36–41 should have been renumbered as 35–40 and ref. 35 should have been renumbered as ref. 41 throughout the text and in **Figure 1**. These errors have been corrected in the HTML and PDF versions of the article.