Single-cell transcriptomics enters the age of mass production

Junker, J.P., van Oudenaarden, A.

Molecular Cell
2015 MAY 21 ; 58(4): 563-564
doi: 10.1016/j.molcel.2015.05.019
Publisher: Cell Press / Elsevier

Single cell transcriptomics enters the age of mass production

Jan Philipp Junker[1,2,*], Alexander van Oudenaarden[1,*]

[1] Hubrecht Institute–KNAW and University Medical Center Utrecht; 3584 CT Utrecht, the Netherlands

[2] Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine; 13092 Berlin-Buch, Germany

Correspondence: j.junker@hubrecht.eu, a.vanoudenaarden@hubrecht.eu

Two publications in the current issue of Cell introduce novel methods for high-throughput single cell transcriptomics by using droplet microfluidics and sophisticated barcoding schemes for transcriptional profiling of thousands of individual cells.

Single cell RNA-seq has recently gained enormous popularity and is currently being adopted by many laboratories around the world (Sandberg, 2014; Shapiro et al., 2013). Macosko et al., 2015 and Klein et al., 2015 have now developed new technologies that will allow researchers to massively increase the numbers of cells that can be sequenced. But why should we care about single cells, and why is it crucial to profile so many of them?

The tissues and organs of multicellular organisms are composed of different cell types whose transcriptional programs, spatial positions, and interactions with other cells are subject to intricate control mechanisms. Much of this complexity is lost in traditional ensemble-based methods such as bulk RNA-seq. Flow sorting can separate heterogeneous cell populations into their constituent cell types, but this approach requires a priori knowledge of all cell types present, and good markers for these cell types. Often, neither of these are available. However, this information can be obtained by single-cell RNA-seq. By sequencing cells from dissociated tissues such as the lung (Treutlein et al., 2014) and the brain (Zeisel et al., 2015),

researchers have been able to group cells by their gene expression profiles in a completely unbiased way, identifying many novel cell type markers along the way. Single-cell transcriptomics thus enables researchers to distinguish cell types without any prior knowledge, and allows systematic definition of cell types and cell states based on quantitative data.

In the last few years, several methods for preparation of single-cell RNA-seq libraries have been established and refined. More recently, the focus of attention has shifted towards innovative techniques for efficient handling of large numbers of cells. Increasing the number of cells to be sequenced is particularly important when analyzing rare cell types (Macosko et al., 2015). While previous publications have studied only a tiny subset of the cells in a tissue, the new methods by Klein et al. and Macosko et al. will soon enable researchers to profile entire organs at much greater depth. Combination of large single-cell sequencing datasets with known spatial expression patterns of marker genes will then allow reconstruction of the complex 3D architecture of entire embryos or organs (Achim et al., 2015; Junker et al., 2014; Satija et al., 2015).

But increasing the number of cells has important implications beyond detection of rare cells: Better statistics is also crucial since single-cell sequencing data is intrinsically noisy because of low detection efficiencies and the small numbers of molecules involved. With more cells, variation of gene expression profiles between cells of the same cell type, and subtle changes of expression profiles during development and aging will become accessible as well. Likewise, studies of regulatory interactions based on correlations of genes will benefit greatly from increased numbers of cells (Klein et al., 2015).

Early publications in single-cell RNA-seq relied heavily on cumbersome manual manipulation of cells and reagents. Automation by chip-based microfluidics (Treutlein et al., 2014; Zeisel et al., 2015) or by flow sorting and liquid handling robotics (Jaitin et al., 2014) has already allowed a considerable increase of cell throughput. However, the number of cells is mostly still limited to hundreds or a few thousand cells. Furthermore, serial processing of individual cells is relatively slow and typically requires multiple sequential preparations, introducing the risk of batch effects. The technique developed by Klein et al. and Macosko et al. is based on

droplet microfluidics, using nanoliter-sized water droplets in carrier oil as reaction chambers (Figure 1). While the fundamental principle is the same – single microparticles coated with barcoded polyT primers and single cells are co-encapsulated in an aqueous droplet – there are also considerable differences between the two experimental protocols. To ensure that single cells rather than doublets are loaded into droplets, both strategies rely on low loading rates, which entails that most droplets will not contain a cell (Figure 1A). Macosko et al. use the same strategy of stochastic mixing for coated microparticles, so that only a small fraction of droplets contains a cell and a microparticle. Klein et al., on the other hand, ensure that almost every droplet is loaded with a microparticle by making use of closely packed deformable hydrogel beads. In the next step, both strategies lyse the cells and hybridize the mRNA inside the droplets (Figure 1B). Macosko et al. then break the droplets and reverse transcribe the mRNA in bulk (Figure 1C). In the protocol by Klein et al., however, reverse transcription proceeds inside droplets after cleavage of barcoded primers (Figure 1C), and droplets are only broken after cDNA has been generated. Finally, the strategy for cDNA amplification is different in the two protocols (Figure 1D). It is important to note that droplet microfluidics is in no way a new technology. However, a lack of efficient barcoding strategies has so far prevented its use for single cell sequencing applications. Hence, the authors' strategies for ensuring that each microparticle is coated with a unique barcode sequence are crucial for the success of the experiment.

Droplet-based single-cell RNA-seq will be an attractive method for many laboratories because of its seemingly unlimited scalability. Since this technology accelerates cell handling while reducing the cost of library preparation, the cost of sequencing will possibly be the bottleneck in the near future. We anticipate that the impact of droplet-based single-cell RNA-seq will continue to grow as the cost of sequencing decreases. For this to happen, it will however be important that this technology is also accessible to researchers without microfluidics experience. There is clearly big potential for commercialization of the technique. The data by Klein et al. and Macosko et al. is of very high technical quality. However, it will be an important task for the field to systematically determine the influence of different cell handling techniques on data quality.

References

Achim, K., Pettit, J.-B., Saraiva, L.R., Gavriouchkina, D., Larsson, T., Arendt, D., and Marioni, J.C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nature Biotechnology 1–9.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., et al. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science *343*, 776–779.

Junker, J.P., Noël, E.S., Guryev, V., Peterson, K.A., Shah, G., Huisken, J., McMahon, A.P., Berezikov, E., Bakkers, J., and Van Oudenaarden, A. (2014). Genome-wide RNA Tomography in the Zebrafish Embryo. *159*, 662–675.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell *161*, 1187–1201.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Sandberg, R. (2014). Entering the era of single-cell transcriptomics in biology and medicine. Nature Publishing Group *11*, 22–24.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology 1–14.

Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. Nature Reviews Genetics *14*, 618–630.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 1–16.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Figures

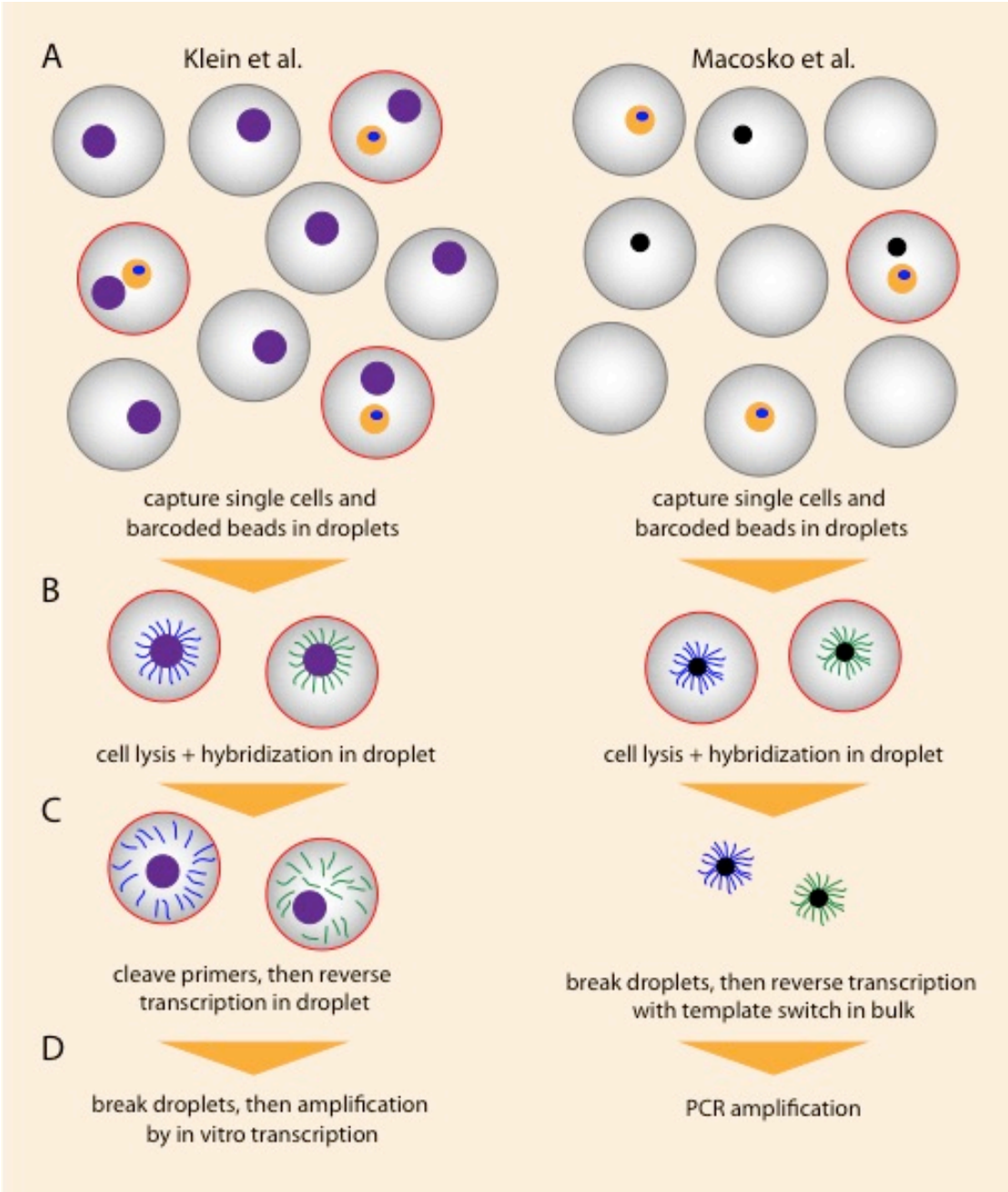

Figure 1. Schematic summary of experimental protocols for droplet-based single-cell RNA-seq using the methods by Klein et al. and Macosko et al. (see main text for details). Droplets are shown in gray, microparticles are purple (Klein et al., 2015) or black (Macosko et al., 2015). Cells are sketched in orange with blue nuclei. mRNA/cDNA molecules are drawn in different colors (blue and green) for different cells to illustrate cell-specific DNA barcodes. A. Mixing of cells and polyT-coated microparticles in nanoliter-scale droplets. Productive droplets containing a microparticle as well as a single cell are circled in red. B. Cell lysis and mRNA hybridization to microparticles. C. Reverse transcription. D. Amplification strategy.