Cell fixation and preservation for droplet-based single-cell transcriptomics

Jonathan Alles[1], Samantha D. Praktiknjo[1], Nikos Karaiskos[1], Stefanie Grosswendt[1], Salah Ayoub[1], Luisa Schreyer[1], Anastasiya Boltengagen[1], Christine Kocks[1,*], and Nikolaus Rajewsky[1,*]

[1] Systems Biology of Gene Regulatory Elements, Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany

* To whom correspondence should be addressed. Tel: +49-30-9406 2999; Fax: +49-30-9406 3068; Email: rajewsky@mdc-berlin.de

Correspondence may also be addressed to christine.kocks@mdc-berlin.de

The authors wish it to be known that, in their opinion, the first 3 authors should be regarded as joint First Authors.

ABSTRACT

'Drop-seq' is a droplet-based, microfluidic method that allows the transcriptional profiling of thousands of individual cells, in a highly parallel and cost-effective way. A critical and often limiting step is the preparation of cells in an unperturbed state, not compromised by stress or ageing. This can be particularly challenging during prolonged handling or transport, or when cells are rare and need to be collected over several days. Here, we tested whether chemical fixation can be used to overcome these problems. Methanol fixation allowed us to stabilize and preserve dissociated cells for several weeks. Fixed cells contained intact RNA and enabled high quality cDNA library generation. By using mixtures of fixed human and mouse cells, we showed that individual transcriptomes could be confidently assigned to one of the two species. Single-cell gene expression measurements from live and fixed samples were highly correlated with each other and with bulk mRNA-seq data. We expect that the availability of a simple cell fixation method will open up many new opportunities to quantitatively analyze transcriptional dynamics at single-cell resolution. As an additional resource, we provide 'dropbead', an R package for easy exploration and quantitative evaluation of informative parameters in Drop-seq data.

INTRODUCTION

A tissue is composed of many different cell types, each of which can have variable biological states. Rather than studying global gene expression of a tissue as a whole, it has been recognized that transcriptional profiling at a single-cell resolution (1-4) provides a much more complete and accurate description of its biological function (5,6). Recent advances in droplet-based microfluidic technologies have made it possible to capture, index and sequence the transcriptional profiles of thousands of individual cells in a highly parallel, ultrafast and affordable manner (7,8).

In the 'Drop-seq' method described by Macosko et al. (7), cells are separately encapsulated in nanoliter-sized droplets together with a single bead in a microfluidic device. One bead delivers barcoded primers, each harboring a polymerase chain reaction (PCR) handle, a cell barcode, and a multitude of different unique molecular identifiers (UMIs), followed by a polyT sequence. The beads are suspended in a lysis buffer, resulting in the cell to be lysed immediately upon droplet formation. Cellular mRNAs are released and can hybridize to the polyT sequences of the barcoded bead. After collection, the droplets are broken and the mRNA reverse-transcribed into cDNA, PCR-amplified, and sequenced in bulk. Computational analysis further allows to distinctly assign each mRNA transcript to its originating cell by means of the cell barcode. The UMIs are used to remove PCR duplicates and to digitally count distinct mRNA molecules.

Despite the rapid rise in high-throughput single-cell RNA-sequencing (RNA-seq) methods, including commercialized versions of automated platforms such as the Fluidigm C1 or 1CellBiO systems, comparatively little attention has been given to the limitations that need to be overcome in the preparation and handling of cellular input material (9). A major challenge in obtaining meaningful information is the use of a high-quality single-cell suspension, which appropriately reflects the transcriptional state of each cell within its natural or experimentally-intended environment. The steps between cell harvesting from culture or after tissue dissociation, isolation of single cells, and mRNA capture are particularly critical as they are prone to introduce transcriptional changes and degradation of RNA. Requirements such as the need to pool cells from several tissues or culture conditions, possibly combined with time course experiments, represent an additional restriction.

In principle, many of these problems could be addressed with the help of chemical fixation. Methanol and ethanol are coagulating fixatives that are superior to aldehydes in preserving nucleic acids, because they do not chemically modify them (10,11). Alcohols act by dehydration: in higher than 65% alcohol and in the presence of salts, nucleic acids occur in a collapsed state, that can be reverted to its original form by a simple rehydration. We have previously shown that fixation with 80% methanol is compatible with next-generation sequencing and library preparation for both, mRNAs and small RNAs (12). Fixation was critically required for successful genome-wide gene expression profiling of FACS-sorted, one to four-cell stage *C. elegans* embryos, a complex tissue undergoing rapid and dynamic transcriptional changes (12).

Here, we adapted the simple methanol-based fixation protocol from Stoeckius et al. (12) to preserve cells for subsequent profiling of single-cell transcriptomes by Drop-seq. We analyzed both live and fixed mixtures of cultured human (HEK) and mouse (3T3) cells. Systematic comparisons show high correlations between gene expression profiles of methanol-fixed and live cells, as well as bulk mRNA-seq data. Methanol fixation did not significantly alter the numbers of genes and transcripts (defined as the number UMIs) detected per cell. We show that RNA integrity is not compromised in cells preserved in methanol up to several weeks after fixation, and that cDNA yields after droplet-based mRNA capturing and further processing are comparable to those obtained from live cells.

Thus, our results demonstrate that Drop-seq profiling of single-cell transcriptomes with methanol-fixed cells can perform equally well as with live cells. Additionally, as a resource, we provide an R package which allows for flexible exploration and quantitative assessment of single-cell data generated by droplet-based microfluidics, and an R script for easy computational reproduction of our plots.

MATERIALS AND METHODS

Cells

Human Flp-In T-Rex 293 HEK cells were a gift from M. Landthaler (MDC, Berlin) originally obtained from Invitrogen (cat no. R78007); murine NIH-3T3 cells were from DSMZ (ACC 59). Cells were grown in DMEM (Invitrogen 61965-026) without antibiotics containing 10% fetal bovine serum, and confirmed to be mycoplasma-free (LookOut Mycoplasma PCR detection kit, Sigma Aldrich).

Preparation of cells for Drop-seq

Cells were grown to 30 to 60% confluence, dissociated with 0.05% bovine trypsin-EDTA (Invitrogen 25300062), quenched with growth medium, and further processed as described previously (Macosko et al. (7), Online-Dropseq-Protocol-v.3.1 http://mccarrolllab.com/dropseq/). Briefly, between 1 to 10 x 10^6 were handled always in the cold or kept on ice, pelleted at 300 x g for 5 minutes at 4˚C, washed with 1x PBS + 0.01% bovine serum albumin fraction V (BSA) (100 µg/ml; Biomol 01400), resuspended in PBS, filtered through a 40 or 35 µm cell strainer and counted. For Drop-seq, a [1 + 1] mix of [HEK + 3T3 cells] was prepared at a final (combined) input concentration of 100 cells/µl in 1x PBS + 0.01% BSA (corresponding to a final concentration of 50 cells/µl after mixing with lysis buffer in the co-flow device).

Fixation of cells for Drop-seq

Cells were trypsinized, and between 1 and 4 x 10^6 cells were processed as described above for Drop-seq (7). Cells were handled in regular (not: “low-binding”) microcentrifuge tubes to minimize cell loss, and kept cold at all times. After straining and counting, cells were pelleted at 300 x g for 5 minutes at 4˚C, the supernatant was removed manually and the cell pellet resuspended in 2 volumes (200 µl) of ice-cold PBS. Methanol-fixation was adapted from Stoeckius et al. (12). To avoid cell clumping, 8 volumes (800 µl) of methanol p. a. (pre-chilled to -20˚C) were added drop-wise, while gently mixing or

vortexing the cell suspension. Methanol-fixed cells were kept on ice for a minimum of 15 minutes and were then stored at -80˚C for up to several months, as indicated. For rehydration, cells were either kept on ice after fixation ("Fixed") or moved from -80˚C to 4˚C ("Fixed 2" samples) and kept in the cold throughout the procedure. Cells were pelleted at 1000 to 3000 x g, resuspended in PBS + 0.01% BSA, centrifuged again, resuspended in PBS + 0.01% BSA, passed through a 40 or 35 µm cell strainer and counted, and diluted for Drop-seq in PBS + 0.01% BSA as described above.

Drop-seq procedure, single-cell library generation and sequencing

Monodisperse droplets of about 1 nl in size were generated using microfluidic polydimethylsiloxane (PDMS) co-flow devices (Drop-seq chips, FlowJEM, Toronto, Canada; rendered hydrophobic by pre-treatment with Aquapel). Barcoded microparticles (Barcoded Beads SeqB; ChemGenes Corp., Wilmington, MA) were prepared and, using a self-built Drop-seq set up, flowed in closely following the previously described instrument set up and procedures by Macosko et al. (7) (Online-Dropseq-Protocol-v.3.1 http://mccarrolllab.com/dropseq/). For most microfluidic co-flow devices, emulsions were checked by microscopic inspection; typically less than 5% of bead-occupied droplets contained more than a single barcoded bead. Droplets were collected in one 50 ml Falcon tube for a run time of 12.5 minutes. Droplets were broken promptly after collection and barcoded beads with captured transcriptomes were reverse transcribed without delay, then exonuclease-treated and further processed as described (7). The 1st strand cDNA was amplified (after assuming loss of about 50% of input beads) by equally distributing beads from one run to 24 (Live; Fixed) or 48 (Fixed 2 samples) PCR reactions (50 µl per PCR reaction; 4 + 9 cycles). 20 or 10 µl fractions of each PCR reaction were pooled, then double-purified with 0.6x volume of Agencourt AMPure XP beads (Beckman Coulter, cat. no. A63881), and eluted in 12 µl. 1 µl of the amplified cDNA libraries were quantified on a BioAnalyzer High Sensitivity Chip (Agilent). 600 pg of each cDNA library was fragmented and amplified (12 cycles) for sequencing with the Nextera XT v2 DNA Library Preparation kit (Illumina) using custom primers that amplified only the 3' ends(7). The libraries were single- (Live; Fixed) or double-purified (Fixed 2 samples) with 0.6x volume of AMPure XP Beads, quantified and sequenced (paired end) on Illumina Nextseq500 sequencers (library concentration: 1.8 pM; Nextseq 500/550 High Output v2 kit (75 cycles); read1: 20 bp (bases 1-12 cell barcode, bases 13-20 UMI) (custom primer 1 Read1CustSeqB), index read: 8 bp, read 2 (paired end): 64 bp). Unique identifiers for the final libraries are: "Live": NR_CK_027 (ds013_50); "Fixed": NR_CK_028 (ds013_50fix); "Fixed 2 (1 week)": NR_CK_043 (ds019_50fix); "Fixed 2 (3 weeks)": NR_CK_048 (ds021_50fix).

Single-cell RNA-seq: data processing, alignment and gene quantification

We chose read 1 to be 20 bp long to avoid reading into the poly(A) tail, leaving 64bp for read 2. The sequencing quality was assessed by FastQC (version 0.11.2). The base qualities of read 1 were particularly inspected, since they contain the cell and molecular barcodes and their accuracy is critical for the downstream analysis. The last base of read 1 consistently showed an increase in T content, indicating possible errors during the beads' synthesis. We observed a similar trend when re-analyzing the original data from Macosko et al. (7). Part of these errors were handled and corrected as

described later. For read 2, we used the Drop-seq tools v. 1.12 (7) to tag the sequences with their corresponding cell and molecular barcodes, to trim poly(A) stretches and potential SMART adapter contaminants and to filter out barcodes with low quality bases. The reads were then aligned to a combined FASTA file of the hg38 and mm10 reference genomes, using STAR (13) v. 2.4.0j with default parameters. Typically, around 65% of the reads were found to uniquely map to either of the species genome; non-uniquely mapped reads were discarded. The Drop-seq toolkit (7) was further exploited to add gene annotation tags to the aligned reads (annotation used was Ensembl release 84) and to identify and correct some of the bead synthesis errors described above. The number of cells (cell barcodes associated with single-cell transcriptomes) was determined by extracting the number of reads per cell, then plotting the cumulative distribution of reads against the cell barcodes ordered by descending number of reads and selecting the inflection point ("knee") of the distribution. It was similar to the number of single-cell transcriptomes expected during the Drop-seq run (see Suppl. Figure S1 for details). Finally, the DigitalExpression tool (7) was used to obtain the digital gene expression matrix (DGE) for each sample.

We developed an R software package ('dropbead'; available for download at https://github.com/rajewsky-lab/dropbead, including a tutorial) which offers an easy and systematic exploration of Drop-seq data and quantitative assessment of basic statistics and important parameters. A threshold of 90% (90 out of 100 UMIs for one species) was selected to confidently declare a cell as being either of the species and not a human/mouse doublet. In order to assess whether fixation generates "low-quality cells" (14), we determined the proportion of non-mitochondrial reads: for every cell we computed the sum of UMIs corresponding to RNA encoded by the mitochondrial genome and then subtracted this number from the sum of all UMIs in that cell. We divided this number by the total number of UMIs in the cell to obtain the non-mitochondrial content as a percentage for every sample.

Single-cell RNA-seq: normalization and correlations of gene expression levels.

We discarded all cells which had less than 3,500 UMIs from all downstream analysis. The raw counts in the DGE were normalized to average transcripts per million (ATPM) as follows: the UMI counts for every gene in a given cell were divided by the sum of all UMIs in that cell. These counts were then multiplied by the sum of all UMIs of the cell with the highest number of UMIs in that library. Correlations of gene expression levels between single-cell samples were computed by first subsetting the DGEs of the two samples to the intersection of the genes captured in both libraries (typically ~10,000) and then computing the sum of gene counts across all cells in each library. Plotting of correlations is shown in log-space. For the correlation of our live and fixed single-cell libraries against mRNA-seq, we converted gene counts into RPKMs and used the mean value of all isoforms lengths for a given gene.

RESULTS

Methanol-fixation preserves single-cell transcriptomes for droplet-based sequencing

Drop-seq with methanol-fixed cells allows correct species assignments in species-mixing experiments. In order to assess whether methanol-fixation is compatible with Drop-seq, we adapted our previously developed methanol-fixation protocol (12) to adherent, mammalian cell lines (see Materials and Methods for details). Methanol-fixed cells remained visible under the microscope as single, intact round cells, which disappeared upon addition of Drop-seq lysis buffer due to complete cellular lysis (data not shown). Fixation did not induce an increase in cell doublets.

To assess the quality of single-cell transcriptomes generated by the Drop-seq procedure after methanol-fixation, we used a mixture of cultured human (HEK) and mouse (3T3) cells as performed previously (7). Both, live and fixed cell mixtures were used at a final concentration of 50 cells/µl for Drop-seq runs carried out on the same day, and cDNA libraries were processed in parallel. Figure 1 shows the experimental workflow and Figure 2 the results of this experiment. We counted the numbers of human and mouse transcripts (UMIs) that were associated with each cell barcode. Both, live (Fig. 2A) and fixed (Fig. 2B) cells, could be confidently assigned to their species of origin using a threshold of 90% species-specific transcripts, suggesting that methanol-fixation preserves cell integrity and the species-specificity of a cell's transcriptome. In addition, this experiment confirmed that fixation did not significantly increase the human/mouse cell doublet rate.

In Drop-seq, cell numbers are selected computationally from the inflection point ("knee") in a cumulative distribution of reads plotted against the cell barcodes ordered by descending number of reads. Cell barcodes beyond the inflection point are believed to represent ambient RNA, not cellular transcriptomes (7). As shown in Suppl. Figure S1A, our fixation protocol did not interfere with our ability to computationally select cells.

Transcript and gene numbers from live and fixed cells are similar. Figures 2C and D, respectively, show the number of genes and transcripts (UMIs) in live and fixed cells. Median transcript and gene numbers from fixed cells and their distributions were similar to live cells, indicating that methanol-fixation did not significantly change the sensitivity of Drop-seq results.

Gene expression levels correlate well between live and fixed cells. We treated single-cell transcriptomes as a bulk population and plotted transcript counts from fixed cells against transcript counts from live cells, to determine how well they correlate. Figure 3 shows that gene expression levels from live and fixed cells were highly correlated ($R \geq 0.94$; Fig. 3A). Furthermore, plotting transcripts from live and fixed cells against a previously published human (HEK) cell bulk mRNA-seq data set from our laboratory (15) revealed high correlations (Figure 4, upper panels).

Taken together, our data suggest that methanol-fixation faithfully preserves single-cell transcriptomes for the Drop-seq procedure.

Fixed cells can be stored for weeks to give reproducible Drop-seq results

We wanted to find out, whether fixed cells can be stored, and if so, for how long they can be used for Drop-seq experiments. In order to address this question, we fixed cells and stored them at -80˚C for one week "Fixed 2 (1 week)", or 3 weeks "Fixed 2 (3 weeks)". As shown in Suppl. Figures S1 and S2, single-cell transcriptomes from cells stored for 1 or 3 weeks performed well in experiments with mixed human and mouse cells. Our results were robust with respect to computational cell selection (Suppl. Fig. S1A), the ability to assign barcodes to an individual cells' organism of origin (Suppl. Fig. 2A, B), or the median number of genes and transcripts per cell (Suppl. Fig. 2C, D).

Our libraries varied with respect to the number of cells. We noticed a dependency of median transcript numbers (UMIs) on the number of cells processed in a Drop-seq run and during library preparation (Suppl. Figure S1B and C). This trend held true for other Drop-seq experiments in our lab (unpublished data) and was not merely a reflection of sequencing depth (more reads per cell), but represented a true gain of transcripts per cell.

Finally, gene expression profiles from fixed cells stored for 1 or 3 weeks correlated well with those of cells that were fixed immediately prior to Drop-seq (Fig. 3B and C) and with each other (Suppl. Fig, S2E). They also showed high correlation with bulk mRNA-seq data (Figure 4, lower panels). We conclude from these data, that fixed cells are stable for several weeks and can be used for Drop-seq experiments without loss in performance.

Methanol-Fixation preserves RNA integrity, cDNA library generation and cytoplasmic RNA content

High-quality RNA and cDNA libraries can be prepared from fixed cells. Figure 5 shows that high-quality, intact total RNA could be extracted from fixed cells after storage for 20 weeks (Fig. 5A; RNA was extracted from the same batch of fixed cells (Fixed 2) that were used to generate the results shown in Suppl. Fig. S1 and S2). Furthermore, Figure 5 shows BioAnalyzer traces corresponding to all four cDNA libraries analyzed in this study: cDNA libraries from methanol-fixed cells appeared indistinguishable from cDNA libraries obtained from live cells (Fig. 5B and unpublished data).

Mitochondrially encoded transcripts are not elevated in methanol-fixed cells. An increase in the proportion of transcripts from mitochondrial genes (37 mtDNA-encoded mRNAs) is believed to indicate low quality cells that are broken or damaged to varying degrees (14). This phenomenon is thought to be caused by leakage leading to relative loss of cytoplasmic mRNAs, compared to mitochondrially located mRNA transcripts that are protected by the two mitochondrial membranes. As shown in Suppl. Figure S3, we observed only a minor loss of cellular cytoplasmic mRNA content across all three fixed Drop-seq libraries. Thus, fixation does not seem to cause a significant increase in "low quality cells".

Taken together, our data therefore suggest that methanol-fixation is able to preserve RNA integrity, and subsequent cDNA library generation during the Drop-seq procedure. Our results also indicate that fixed cells can be stored for prolonged periods up to several weeks.

DISCUSSION

To our knowledge only few studies have explicitly dealt with cell preservation protocols for single-cell sequencing. One study uses cryopreservation followed by flow cytometry to sort single cells for subsequent processing (16). While cryopreservation is compatible with Drop-seq in principle, it remains to be determined how well Drop-seq will perform with recently thawed cells that may be fragile and prone to cell death. Another study describes fixation of cells by crosslinking with formaldehyde (17), followed by reverse crosslinking (breakage of methylene bridges between protein and RNA molecules with heat). Crosslinking induces chemical modifications that inhibit poly(dT) annealing, reverse transcriptase and cDNA synthesis (10), making crosslink reversal necessary. However, it seems unlikely that reverse crosslinking can be applied to droplet-based sequencing, because the emulsion droplets cannot withstand heat. Furthermore, reversal of crosslinking is inefficient and results in high losses of available RNA molecules.

By contrast, we have shown here that methanol-fixation of tissue culture cells does not lead to significant RNA loss or degradation, and is easily compatible with the Drop-seq single-cell sequencing protocol and workflow. It is possible to store single-cell suspensions for prolonged times at low temperatures and, therefore, to separate the sample preparation phase in time or location from the actual droplet-sequencing procedure. In addition to cell culture lines (human HEK and Hela cells, mouse NIH3T3 cells), we have successfully applied our methanol-fixation protocol to a variety of primary cell samples from diverse organisms such as *Drosophila* embryo cells, mouse brain tissues or *Planaria* (unpublished results).

However, methanol fixation may not work in all circumstances, for all tissues or cell types. Successful fixation may be challenging especially for tissues with a high content in proteases and RNAses such as pancreas, gall bladder, skin or lymphatic tissues. In support of this notion, we observed a failure to generate Drop-seq cDNA libraries and to isolate intact RNA from a mouse lymphoma (unpublished results). For these types of tissues, it will be important to determine at which step RNA degradation occurs, before or after fixation. Modifications to the fixation protocol or addition of RNAse inhibitor during the rehydration step might remedy these problems. It also remains to be determined whether methanol-fixation is compatible with the 'InDrop' protocol, another, recently developed droplet-based sequencing approach that involves a different type of chemistry and cDNA library construction, and a different detergent for cell lysis.

The availability of a simple cell fixation protocol will open up many previously inaccessible experimental avenues for droplet-based single-cell transcriptomics. Fixation and preservation of cells

at an early stage of preparation removes bias and technical variation, prevents cell stress, or unintended ageing during the experiment, and facilitates systematic assessment of experimental parameters. Cell fixation may also significantly ease the logistic coordination of large-scale experiments. In a variety of situations, fixation may be the only solution to being able to process and provide cellular input material: Examples are rare cells which cannot be obtained in one experimental session, clinical specimen which require transportation, or cells that are hard to isolate, and require extensive upstream processing such as tissue dissociation followed by flow cytometry.

In summary, we expect that the methanol-based cell fixation procedure presented here will greatly stimulate high-throughput single-cell sequencing studies in diverse areas. Our work is accompanied by an R software package we developed, 'dropbead', which offers an accessible framework for exploring Drop-seq generated data and quantitatively assessing single-cell RNA samples. To demonstrate dropbead's ease of use, we include a short R script with instructions to readily reproduce all figures in this study.

ACCESSION NUMBERS

The accession numbers for the raw and analyzed data in this study are: pending. Illumina sequencing data were uploaded to GEO (record GSE89164) and will be automatically transferred to NCBI's Sequence Read Archive.

SUPPLEMENTARY DATA are available as separate file.

FUNDING

This work was supported by Deutsche Forschungsgemeinschaft [RA838/5-1 to N.R.]; and the Berlin Institute of Health [CRG2aTP7 to N.R.]; and the Deutsches Zentrum für Herz-Kreislaufforschung e.V. [BER1.2VD to N.R.]. Funding for open access charge: Max Delbrück Center for Molecular Medicine in the Helmholtz association (MDC).

ACKNOWLEDGEMENTS

We thank Rahul Satija for advice with single-cell sequence analysis, Evan Macosko, Melissa Goldman and Steve McCarroll for help with setting up Drop-seq, Sara Formichetti for RNA extractions, Haiyue Liu for testing our R scripts, and our colleagues from the Rajewsky laboratory for manifold help and critical discussions.

Conflict of interest: none.

REFERENCES

1. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A. *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods*, **6**, 377-382.
2. Hashimshony, T., Wagner, F., Sher, N. and Yanai, I. (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep*, **2**, 666-673.
3. Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G. and Sandberg, R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature methods*, **10**, 1096-1098.
4. 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 (New York, N.Y.)*, **343**, 776-779.
5. Gawad, C., Koh, W. and Quake, S.R. (2016) Single-cell genome sequencing: current state of the science. *Nature reviews. Genetics*, **17**, 175-188.
6. Macaulay, I.C. and Voet, T. (2014) Single cell genomics: advances and future perspectives. *PLoS genetics*, **10**, e1004126.
7. 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.
8. 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.
9. Saliba, A.E., Westermann, A.J., Gorski, S.A. and Vogel, J. (2014) Single-cell RNA-seq: advances and future challenges. *Nucleic acids research*, **42**, 8845-8860.
10. Srinivasan, M., Sedmak, D. and Jewell, S. (2002) Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *The American journal of pathology*, **161**, 1961-1971.
11. Esser, C., Gottlinger, C., Kremer, J., Hundeiker, C. and Radbruch, A. (1995) Isolation of full-size mRNA from ethanol-fixed cells after cellular immunofluorescence staining and fluorescence-activated cell sorting (FACS). *Cytometry*, **21**, 382-386.
12. Stoeckius, M., Maaskola, J., Colombo, T., Rahn, H.P., Friedlander, M.R., Li, N., Chen, W., Piano, F. and Rajewsky, N. (2009) Large-scale sorting of C. elegans embryos reveals the dynamics of small RNA expression. *Nature methods*, **6**, 745-751.
13. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, **29**, 15-21.
14. Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C. and Teichmann, S.A. (2016) Classification of low quality cells from single-cell RNA-seq data. *Genome biology*, **17**, 29.
15. Rybak-Wolf, A., Jens, M., Murakawa, Y., Herzog, M., Landthaler, M. and Rajewsky, N. (2014) A variety of dicer substrates in human and C. elegans. *Cell*, **159**, 1153-1167.
16. Guillaumet-Adkins, A., Rodriguez-Esteban, G., Mereu, E., Villanueva, A., Vidal, A., Gut, M., Gut, I. and Heyn, H. (2016) Single Cell Transcriptome Conservation in Cryopreserved Cells and Tissues. *bioRxiv*.
17. Thomsen, E.R., Mich, J.K., Yao, Z., Hodge, R.D., Doyle, A.M., Jang, S., Shehata, S.I., Nelson, A.M., Shapovalova, N.V., Levi, B.P. *et al.* (2016) Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nature methods*, **13**, 87-93.

FIGURE LEGENDS

Figure 1. Experimental workflow of cell preparation for single-cell transcriptional profiling by Drop-seq.

Cultured human (HEK) and mouse (3T3) cells were dissociated, mixed and further processed to analyze the transcriptomes of either live or fixed cells by Drop-seq. Washed cells were gently resuspended in 2 volumes of PBS, then fixed by addition of 8 volumes of methanol. Methanol-fixed cells could be stored for up to several weeks at -80 °C. Prior to Drop-seq, cells were washed before passing them through a 35 – 40 μm cell strainer. Cells were then separately encapsulated in droplets together with a single bead in a microfluidic co-flow device and single-cell transcriptomes sequenced in a highly parallel manner. Downstream analysis and systematic quantitative comparisons were subsequently made from separate experiments using live or fixed cellular input material with an R package ('dropbead') that we developed and is freely available for download at https://github.com/rajewsky-lab/dropbead.

Figure 2. Drop-seq with fixed cells preserves transcriptome integrity and gene expression levels.

(A and B) Drop-seq analysis of human (HEK) and mouse (3T3) cell mixtures, analyzed at a final combined concentration of 50 cells/μl. Scatter plots show the number of human and mouse transcripts (UMIs) associating with each cell. Each dot represents a cell identified as human- or mouse-specific (blue or red, respectively), using a threshold of 90% reads mapping to the human or mouse genome, respectively, and a cut-off of more than 3500 UMIs expressed per cell (cells expressing less than 3500 UMIs are grey).

(A) Live cells contained only few human-mouse doublets (1%; violet dots).

(B) Fixed cells contained a similar number of human-mouse doublets compared to live cells (1.6%; violet dots) (see also Suppl. Figures S2A and S2B for two independent repeat experiments).

(C and D) Drop-seq sensitivity at non-saturating read depth: libraries were sequenced to a mean depth of 34,000 aligned reads per cell. Violin plots show the distribution and the median of the number of genes (C) or transcripts (UMIs) (D) detected per cell, for the top 100 cells (highest numbers of UMIs) of the libraries depicted in Figure 2A and 2B (Live: 27 human and 70 mouse cells; Fixed: 44 human and 50 mouse cells; human-mouse cell doublets were removed computationally).

Experiments using live and fixed cells were independently repeated with similar results (unpublished).

Figure 3. Gene expression levels from live and fixed cells correlate well.

(A to C) Correlation between gene expression measurements from Drop-seq experiments with live or fixed cells. Dots represent gene expression levels. Human (HEK) cells, light grey; mouse (3T3) cells, dark grey. (A) Correlation between live and fixed cells prepared and used for Drop-seq on the same day. (B) Correlation between fixed cells and cells fixed and stored at -80˚C, and used for Drop-seq independently on different days (biological repeat). (C) Correlation between the same batch of fixed cells, stored at -80˚C for one or three weeks and processed by Drop-seq independently. Drop-seq expression counts were converted to average transcripts per million (ATPM) and plotted as log2 (ATPM + 1).

Figure 4. Single-cell gene expression levels from Drop-seq experiments correlate well with bulk mRNA-seq data.

Correlations between gene expression measurements from Drop-seq and non-single cell bulk mRNA-seq. Dots represent gene expression levels of human (HEK) cells (Live: 288 cells; Fixed: 598 cells; Fixed 2 (1 week): 140 cells and Fixed 2 (3 weeks): 553 cells). Non single-cell bulk mRNA-seq data (15) were expressed as reads per kilobase per million (RPKM). Drop-seq expression counts were converted to average transcripts per million (ATPM) and plotted as log2 (ATPM + 1).

Figure 5. Fixation does not impair RNA integrity or cDNA library generation.

(A and B) Bioanalyzer traces of total RNA (A) and Drop-seq cDNA libraries (B).

(A) Total RNA from human (HEK) and mouse cells (3T3) extracted 20 weeks after cell fixation and storage at -80˚C is of highest quality. (RNA integrity factor (RIN): 10 out of 10).

(B) Fixation does not change the fragment distribution of Drop-seq cDNA libraries. Libraries correspond to the Drop-seq experiments shown in Figure 2 and Supplemental Figure S2.

Figure 1

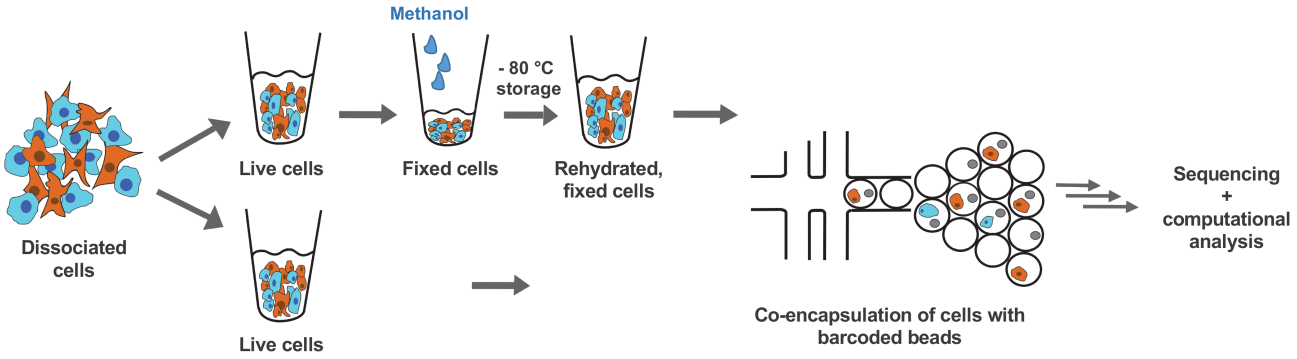

Methanol
- 80 °C
storage
Dissociated
cells
Live cells
Fixed cells
Rehydrated,
fixed cells
Live cells
Co-encapsulation of cells with
barcoded beads
Sequencing
+
computational
analysis

Figure 2

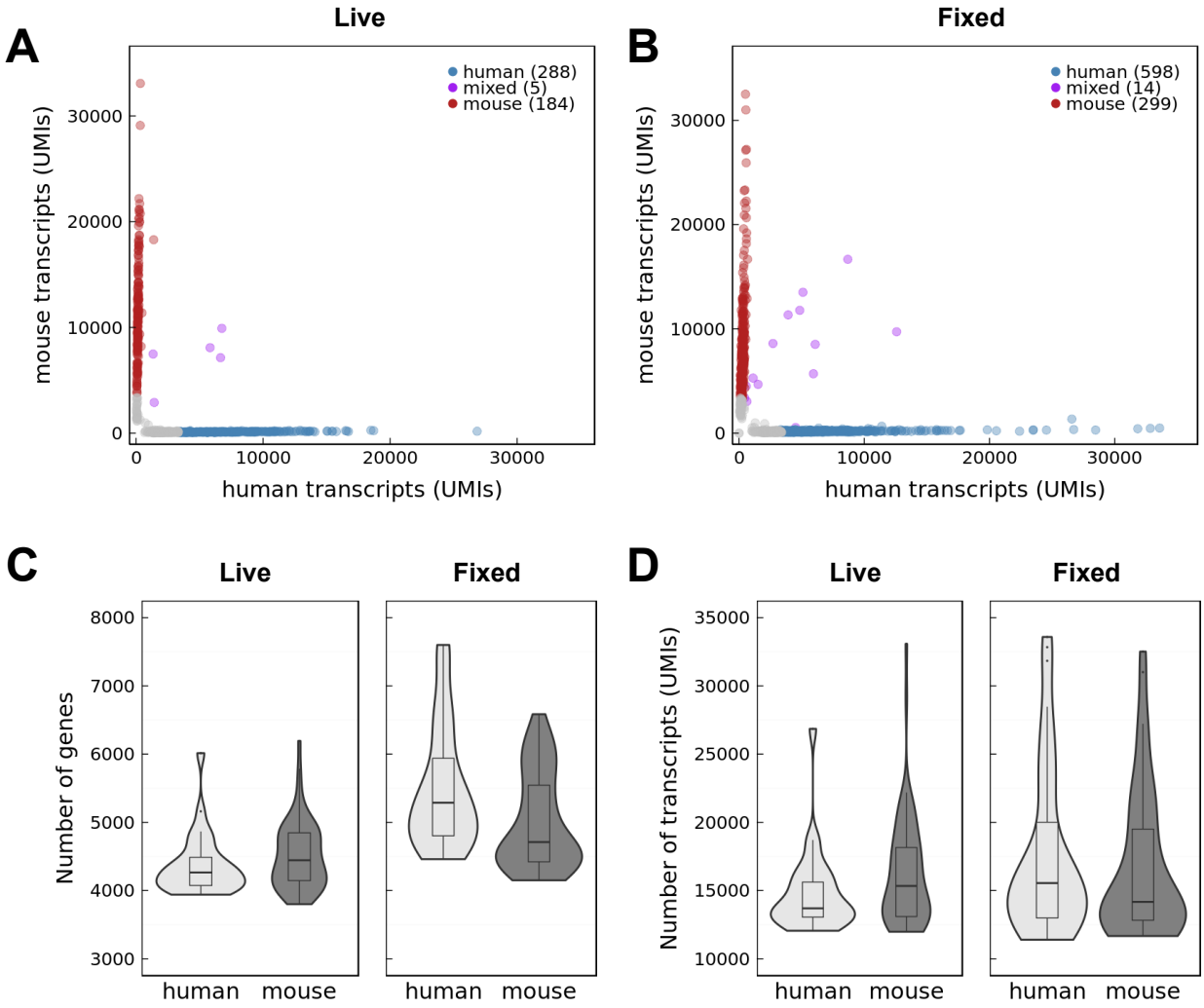

A
Live
human (288)
mixed (5)
mouse (184)
mouse transcripts (UMIs)
human transcripts (UMIs)
B
Fixed
human (598)
mixed (14)
mouse (299)
mouse transcripts (UMIs)
human transcripts (UMIs)
C
Live
Fixed
Number of genes
human
mouse
D
Live
Fixed
Number of transcripts (UMIs)
human
mouse

Figure 3

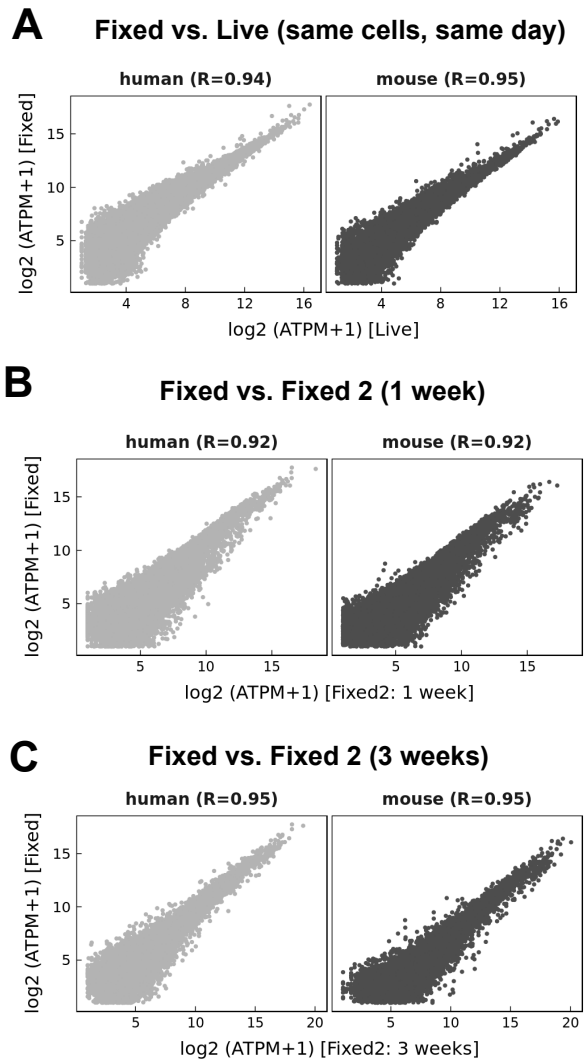

A
Fixed vs. Live (same cells, same day)
human (R=0.94)
mouse (R=0.95)
log2 (ATPM+1) [Fixed]
log2 (ATPM+1) [Live]
B
Fixed vs. Fixed 2 (1 week)
human (R=0.92)
mouse (R=0.92)
log2 (ATPM+1) [Fixed]
log2 (ATPM+1) [Fixed2: 1 week]
C
Fixed vs. Fixed 2 (3 weeks)
human (R=0.95)
mouse (R=0.95)
log2 (ATPM+1) [Fixed]
log2 (ATPM+1) [Fixed2: 3 weeks]

Figure 4

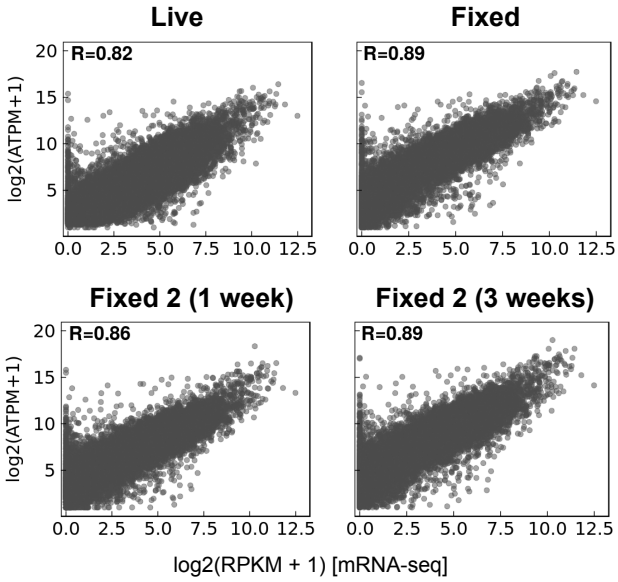
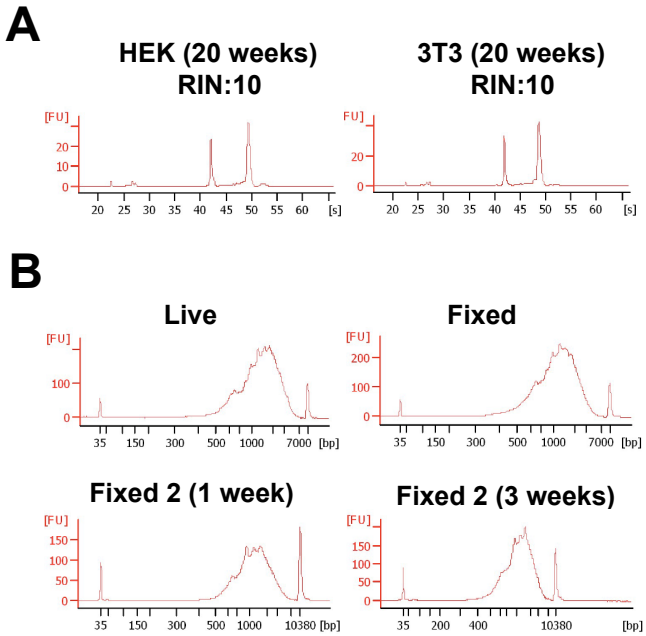

Live
Fixed
Fixed 2 (1 week)
Fixed 2 (3 weeks)
R=0.82
R=0.89
R=0.86
R=0.89
log2(ATPM+1)
log2(RPKM + 1) [mRNA-seq]
0.0 2.5 5.0 7.5 10.0 12.5
5 10 15 20

Figure 5


A
HEK (20 weeks)
RIN:10
3T3 (20 weeks)
RIN:10
[FU]
20
10
0
20 25 30 35 40 45 50 55 60 [s]
B
Live
Fixed
[FU]
100
0
35 150 300 500 1000 7000 [bp]
200
Fixed 2 (1 week)
Fixed 2 (3 weeks)
150
50
35 150 300 500 1000 10380 [bp]
35 200 400 10380 [bp]