bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Comparative analysis of single-cell RNA-sequencing methods

*Christoph Ziegenhain[1], Swati Parekh[1], Beate Vieth[1], Martha Smets[2], Heinrich Leonhardt[2], Ines Hellmann[1] and Wolfgang Enard[1]**

[1] Anthropology & Human Genomics, Department of Biology II, Ludwig-Maximilians University, Großhaderner Str. 2, 82152 Martinsried, Germany.

[2] Human Biology & Bioimaging, Department of Biology II, Ludwig-Maximilians University, Großhaderner Str. 2, 82152 Martinsried, Germany.

*** Corresponding author**

Wolfgang Enard

Anthropology and Human Genetics

Department of Biology II

Ludwig-Maximilians University

Großhaderner Str. 2

82152 Martinsried, Germany

Phone: +49 (0)89 / 2180 - 74 339

Fax: +49 (0)89 / 2180 - 74 331

E-Mail: enard@bio.lmu.de

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Single-cell mRNA sequencing (scRNA-seq) allows to profile heterogeneous cell populations, offering exciting possibilities to tackle a variety of biological and medical questions. A range of methods has been recently developed, making it necessary to systematically compare their sensitivity, accuracy, precision and cost-efficiency.

Here, we have generated and analyzed scRNA-seq data from 479 mouse ES cells and spike-in controls that were prepared with four different methods in two independent replicates each. We compare their sensitivity by the number of detected genes and by the efficiency with which they capture spiked-in mRNAs, their accuracy by correlating spiked-in mRNA concentrations with estimated expression levels, their precision by power simulations and variance decomposition and their efficiency by their costs to reach a given amount of power. While accuracy is similar for all methods, we find that Smart-seq on a microfluidic platform is the most sensitive method, CEL-seq is the most precise method and SCRB-seq and Drop-seq are the most efficient methods. Our analysis provides a solid basis to choose among four available scRNA-seq methods and to benchmark future method development.

Introduction

Genome-wide quantification of mRNA transcripts is a highly informative phenotype to characterize cellular states and understand regulatory circuits[1,2]. Ideally, this is collected with high temporal and spatial resolution and RNA-sequencing of single cells (scRNA-seq) is starting to reveal new biological and medical insights[3–5]. scRNA-seq requires to isolate intact single cells and turn their mRNA into cDNA libraries that can be quantified using high-throughput sequencing[4,6]. The sensitivity, accuracy, precision and throughput of this process determines how well single cell transcriptomes can be characterized. In order to choose among available scRNA-seq methods, it is important to estimate these parameters comparatively. It has previously been shown that scRNA-seq performed in the small volumes available via the automated microfluidic platform from Fluidigm (Smart-seq protocol on the C1-platform) performs better than Smart-seq or other commercially available kits in microliter volumes[7]. More recent scRNA-seq protocols have used unique molecular identifiers (UMIs) that tag mRNA molecules with a random barcode sequence during reverse transcription in order to identify sequence reads that originated during amplification[8–10]. This leads to a better quantification of mRNA molecules[11,12]. Another decisive recent development has been to add a barcode for each single cell during reverse

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transcription allowing to process hundreds or thousands of scRNA-seq libraries in one reaction, increasing the throughput of scRNA-seq library generation by one to three orders of magnitude[9,13–15]. However, a systematic comparison of sensitivity, accuracy and precision among such recently developed methods has not been performed yet. To this end, we have generated and analysed 479 scRNA-seq libraries from mouse embryonic stem (mES) cells cultured in two-inhibitor (2i/LIF) medium using four different methods run in two replicates each (Fig. 1).

Results

Generation and processing of 479 scRNA-seq libraries

We have used the Smart-seq protocol on the C1 platform from Fluidigm (Smart-seq/C1) that uses microfluidic chips to automatically separate up to 96 cells[7]. After lysis and the addition of artificial mRNAs of known concentrations (ERCCs[16]), cDNA is generated by oligo-dT priming, template switching and PCR amplification. These can then be turned into scRNA-seq libraries using 96 Nextera reactions. Advantages of this system are that the process of single cell isolation is automated, that captured cells can be imaged, that reaction volumes are small and that the protocol generates full length cDNAs. Disadvantages include that the protocol does not use UMIs and is expensive due to the use of the Smart-seq kit, the microfluidic chip and the separate Nextera reactions.

For the second approach we isolated single cells in a 96-well plate by sorting them into lysis buffer and used the SCRB-seq protocol to generate scRNA-seq libraries[13]. In this protocol, cDNA from cells and ERCCs is generated by oligo-dT priming, template switching and PCR amplification, similar to Smart-seq/C1. However, the oligo-dT primers contain well-specific barcodes and UMIs and the resulting cDNA is turned into scRNA-seq using a modified transposon-based fragmentation approach that enriches for 3' ends. The protocol is optimized for small volumes and few handling steps, but it does not generate full length cDNA profiles and its performance compared to the smaller reaction volumes in the C1 is unclear.

The third method we tested is Drop-seq, a recently developed microdroplet-based approach[15]. Similarly to SCRB-seq, each cDNA molecule is decorated with a cell-specific multiplexing barcode and a UMI to count abundance. In the case of Drop-seq, these barcoded oligo-dT primers are immobilized on beads and encapsulated with single-cells in nanoliter-sized emulsion droplets. In Drop-seq, simultaneous inclusion of ERCC spike-ins in the library preparation is not possible, as it would have to be included in the bead suspension. However, most beads will not carry a cell transcriptome, as cell concentration

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

has to be relatively low to avoid doublets. To address the issue of lacking ERCCs in our data, we used a published dataset[15], where ERCC spike-ins were sequenced by the Drop-seq method without single-cell transcriptomes. Finally, we used data[12] generated using CEL-seq[9] for which also two different batches of scRNA-seq libraries were generated from the same cell type and culture conditions. Similarly to Drop-seq and SCRB-seq, cDNA is tagged with multiplexing barcodes and UMIs. As opposed to the three PCR-based methods described before, CEL-seq relies on linear amplification by in-vitro transcription (IVT) for the initial pre-amplification of single-cell material. An overview of the methods and their features is given in Supplementary Table 1.

For Smart-seq/C1, SCRB-seq and Drop-seq we generated 437, 253 and 432 million reads from 192, 192 and ~200 cells in the two independent replicates. Data from CEL-seq consisted of 51 million reads from a total of 74 cells (Fig. 1). After demultiplexing and assignment of cell barcodes, we could map >50% of all reads to the mouse genome and 30-40% of all reads to exons (Fig. 2a), comparable to previous results[7,14]. As expected, the 3' counting methods showed a strong 3' bias while Smart-seq/C1 showed a moderate 3' bias (Supplementary Fig. 1a). We used the Drop-seq pipeline[15] to quantify UMIs and found that between 2.6% (Drop-seq, replicate B) and 59.5% (SCRB-seq, replicate A) of all exon reads were unique, i.e. that the scRNA-seq libraries had been sequenced on average 1.7- to 38-fold over (Fig. 2a). We filtered the scRNA-seq libraries for possible cell doublets either by their microscope images (Smart-seq/C1) or by discarding transcriptomes with more than twice the mean total UMI content. Furthermore, we disregarded scRNA-seq libraries with less than 100,000 reads mapped to exons, resulting in 479 single cell transcriptomes across the four methods (Fig. 1). As expected, these transcriptomes cluster according to method in a principal component analysis, whereas the biggest difference is between the 3' counting methods and Smart-seq/C1 (Supplementary Fig. 1b). In summary, we obtained high-quality scRNA-seq data for 479 cells distributed equally across methods and replicates, allowing to compare the four different scRNA-seq methods with respect to their sensitivity, accuracy and precision.

SCRB-seq and Smart-seq/C1 show the highest sensitivity

As a relative measure of sensitivity we compared the number of detected genes per method and replicate (Fig. 2b). Clearly, Drop-seq had the lowest sensitivity with a median of 3158 genes per cell, CEL-seq had 6384 and SCRB-seq and Smart-seq/C1 were very similar with a median of ~8800 genes detected per cell (Fig. 2b). This relative ranking remains unchanged if one subsamples 100.000 reads for each cell, but CEL-seq performs then

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

almost as good as SCRB-seq and Smart-seq/C1 (Supplementary Fig. 2a). While the number of detected genes per cell differs, the total number of detectable genes in all cell converges probably around 20,000 of the 39,000 annotated mouse genes for all 3' counting methods. In contrast, the Smart-seq/C1 method converges at more than 25,000 detected genes (Fig. 2c). Furthermore, while the majority of genes is detected by all four methods, Smart-seq/C1 detects 3974 genes not seen by the 3' counting methods (Supplementary Fig. 2b,c). This pattern remains when subsampling 50 cells à 100,000 reads per method (data not shown).

As an absolute measure of sensitivity, we compared the probabilities of detecting the 92 artificially spiked-in mRNAs (ERCCs) since their concentration per cell is known. We recorded the number of cells with an expression value of zero for each ERCC per method and applied a maximum likelihood method[17] to estimate the probability of successful detection of an ERCC mRNA with a given number of molecules per cell as a binomial process (Fig. 2d). This is more informative than non-zero detections of ERCC, as in that case it is only known that some copies were successfully captured but not how many exactly. Additionally, this method is also superior to other approaches as it does not try to relate the expression level to the capture efficiency. For Smart-seq/C1, a gene is detected in half of the cells when it has a concentration of ~8 molecules per cell. Drop-seq and SCRB-seq show similar estimates of 11 molecules per cell. However, CEL-seq needs 92 molecules to reach the 50% detection probability (Supplementary Fig. 2d). Similar estimates of 16, 17, 18 and 93 molecules per cell for Smart-seq/C1, SCRB-seq, Drop-Seq and CEL-seq are made when subsampling 10,000 ERCC reads for 50 cells per method. The drastically lower sensitivity of CEL-seq when using ERCCs contradicts the just slightly lower sensitivity when considering the number of detected genes. This discrepancy was also noted in the original CEL-seq publication and explained with a potential degradation of the ERCCs[12]. Hence, the ERCC reads in this CEL-seq data set might underestimate the sensitivity of this method considerably. The other discrepancy between the two sensitivity estimates is seen for Drop-seq. While it detects less than half the number of genes than Smart-seq/C1 and SCRB-seq, it shows a similar sensitivity for the ERCCs. However, it is important to realize that ERCCs can not be spiked in when using the Drop-seq approach, but have to be run separately from the cells[15]. It is unclear to what extend the lower complexity in an "ERCC-only" run might lead to higher sensitivities and/or to what extent this reflects a difference between data generated by Macosko et al and data generated by us. For now, we conservatively assume that our data is representative and hence that Drop-seq is less than half as sensitive as the other methods.

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In summary, we find that Smart-seq/C1 is the most sensitive method as it detects at least as many genes per cell as CEL-seq and SCRB-seq and detects in total ~ 4000 genes that are not detected by the 3' counting methods.

Accuracy is similar between Drop-seq, SCRB-seq and Smart-seq/C1

In order to quantify the accuracy of transcript level quantifications, we compared observed expression values with annotated molecule concentration of the 92 ERCC transcripts (Fig 3a). For each cell, we calculated the correlation coefficient (R^2) for a linear model fit (Fig 3b). Clearly, CEL-seq performed worst in this measure, while Drop-Seq, SCRB-seq and Smart-seq/C1 showed largely similar accuracy. This pattern remained with downsampled ERCC data (data not shown). As discussed above, this low performance of CEL-seq could be consistently explained with a potential degradation of the ERCCs. A previous study using CEL-seq from 10 pg of total RNA input and ERCC spike-in reported a mean correlation coefficient of R^2=0.87[9], similar to the ones reported for the other three methods. Assuming that this is representative for CEL-seq, we find that the accuracy is similarly high across the four methods.

Precision is highest for CEL-seq and strongly increased by UMIs

While a high accuracy is necessary to quantify absolute expression values, one most often analyses relative differences among cells or samples in order to identify differentially expressed genes or biological variation. Hence, the precision of the method, i.e. its reproducibility or amount of technical variation matters more.

We use the probabilistic model proposed by Kim et al[18] that estimates the proportion of technical variation from ERCCs. Since we use the same cells under the same culture conditions we assume that the remaining fraction of biological variation is the same across the data sets. We find that Smart-seq/C1 has the lowest proportion of technical variation, which is surprising given that it is the only method that lacks UMIs that are expected to increase the precision. Indeed, the total amount of variation is higher for Smart-seq/C1 than for the other three methods (Supplementary Figure 3a) suggesting that the technical variation is underestimated for Smart-seq/C1. Indeed, when we repeat the analysis without considering the UMI information, Drop-seq and SCRB-seq have now more total variation (Supplementary Fig. 3a) and less technical variation (Fig. 4a). Interestingly, the technical variation of CEL-seq remains almost unaffected and is the lowest among the four methods when considering the effect of UMIs (Fig. 4a). This pattern remains when filtering out lowly expressed genes and ERCCs (data not shown). When using another variance

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

decomposition method[19], we do not observe the difference between using and not using UMIs, probably due to filtering of genes whose variance composition can not be reliably estimated. Although this filtering makes it difficult to compare the fraction of technical variation, it is worth mentioning that CEL-seq has again the lowest amount of technical variation (Supplementary Fig. 3b). Hence, when using variance decomposition methods CEL-seq shows the highest precision. Maybe even more important, our analysis shows that some variance decomposition method can severely underestimate the technical variation from ERCCs for PCR-based amplification protocols.

In order to investigate precision in scRNA-seq methods independent of spike-in transcripts, we used power simulation. Based on the method by Wu et al[20], we used the mean-dispersion distributions of the data (Supplementary Fig. 4) to simulate the power and the false discovery rate for differential gene expression by DESeq2[21]. While the three UMI-based methods all had false discovery rates (FDRs) close to the expected 5%, Smart-seq/C1 resulted in progressive FDRs considerably above 5% (Fig. 4b). Furthermore, CEL-seq needed just 52 cells per group to detect small expression differences with 80% power, followed by SCRB-seq (112) and Drop-seq (170), while Smart-seq/C1 reached 80% power only for highly expressed genes (Fig. 4b and Supplementary Fig. 5a). Accordingly, when simulating the power to detect larger expression differences ($\log_2$ fold-change of 0.5-3) with a fixed sample size of 16 cells per group, CEL-seq performed best (Supplementary Fig. 6b), followed by SCRB-seq and Drop-seq, while Smart-seq/C1 had the lowest power. Notably, when analysing power without using UMIs, SCRB-seq and CEL-seq perform similar to Smart-seq/C1, while Drop-seq performs very poorly (Supplementary Fig. 5b). Hence, UMIs strongly increase the power to detect differentially expressed genes by scRNA-seq, especially for methods that use PCR amplification.

In summary, when using power simulations to compare precision among methods, CEL-seq performs best, followed by SCRB-Seq and Drop-Seq. Smart-seq/C1 performs worst, probably because it does not include UMIs.

Efficiency is highest for SCRB-seq and Drop-seq when considering costs and power

Practically, the costs of a method also matter when judging the performance of different scRNA-seq methodologies. Here, we estimate the cost-efficiency by calculating the costs for generating scRNA-seq data at a given amount of power. Given the number of single cells that are needed per group to reach 80% power (50% for Smart-seq/C1) as simulated above (Figure 4b), we calculate the costs to generate these libraries and to sequence

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

enough to obtain 100,000 exon-mapped reads (Figure 2). For example, SCRB-seq requires 112 cells per group and generating 224 SCRB-seq libraries costs ~450€. To generate 100,000 reads mapping to exons for 224 SCRB-seq libraries requires ~64 million paired-end reads generated with a 50 cycles single end kit that we assume cost 320€. When we do analogous calculations for the three other methods (Table 1), we find that Drop-Seq and SCRB-seq are equally cost-effective, closely followed by CEL-seq, while Smart-Seq/C1 is almost ten-fold less efficient due its high library costs that arise from the microfluidic chips and the costs for generating independent libraries. The estimate should be understood as a lower boundary for the real costs since many factors are not considered such as costs to set-up the methods, costs to isolate single cells, costs due to unequal pooling of libraries or costs due to practical constraints in generating a fixed number of scRNA-seq libraries and reads. Furthermore, power for differential gene expression is not the only factor that matters and in addition to the metrics investigated above, methods differ with respect to full length coverage of cDNAs (only possible for Smart-seq/C1), possibility of imaging cells (CEL-seq and Smart-seq/C1) and the possibility to have ERCCs for each library (not possible for Drop-seq).

Nevertheless, we think that our estimates allow a fair and transparent comparison of the four methods and reveal that Drop-Seq and SCRB-seq are most efficient and Smart-seq/C1 is almost ten-fold less cost-efficient.

Discussion

Single-cell RNA-sequencing (scRNA-seq) is a powerful technology to tackle biomedical questions. To facilitate choosing among the many approaches that were recently developed, we systematically compared four scRNA-seq methods and assessed their sensitivity, accuracy, precision and cost-efficiency. We chose a leading commercial platform (Smart-seq/C1), a method that uses in-vitro transcription for amplification from manually isolated cells (CEL-seq), a PCR-based method with a very high throughput (Drop-seq) and a PCR-based method that allows single cell isolation by FACS (SCRB-seq). All these methods can be set up by a molecular biology lab, as has been done by us for Smart-seq/C1, SCRB-seq and Drop-seq.

We find that SCRB-seq, Smart-seq and CEL-seq detect a similar number of genes per cell while Drop-seq - at least in our hands - detects over 50% less than the other methods (Fig. 2b and Supplementary Figure 2a). Whereas we detect 3,158 genes per ES cell, 5,663 genes per 3T3 cell are found in the original Drop-seq publication[15]. It is currently unclear whether

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

this discrepancy is due to biological or technical reasons. Importantly, Drop-seq does not detect generally fewer genes since the total number of detected genes converges around 20,000, similar as for SCRB-seq and CEL-seq (Fig. 2c). While these three 3' counting methods detect largely the same genes, Smart-seq/C1 detects almost 4000 additional genes (Supplementary Fig. 2b). So while ERCCs are measured with similar high accuracy by all four methods (Fig. 3), a considerable number of genes are missed by the 3' counting methods. Hence, Smart-seq/C1 is the most sensitive method because it detects considerably more genes in addition to its full length coverage of transcripts (Supplementary Fig. 1a). However, it is difficult to gauge how much information this adds for finding and defining cell types and characterizing their biological states and networks.

Probably more important than the sensitivity and accuracy for the performance of scRNA-seq methods is their precision, i.e. their reproducibility of measuring gene expression levels. We find that variance decomposition methods that use ERCCs to estimate precision can severely underestimate the technical variance when PCR is used to amplify cDNA without UMIs (Fig. 4a). Further work will be needed to investigate whether more ERCCs, other ERCCs and/or different estimation methods could solve this. Assuming that variance decomposition can accurately estimate precision when using UMIs, we find that CEL-seq performs better than the PCR-based Drop-seq and SCRB-seq. When using power simulations that use the entire data, we also find that CEL-seq is more precise since it detects differentially expressed genes with fewer cells (Fig. 4b). These simulations also show that UMIs increase the power considerably, especially for the PCR-based methods (Supplementary Figure 5b). Hence, in-vitro transcription (IVT)-based amplification is more precise than PCR-based amplification and - in contrast to bulk RNA-seq libraries[22] - UMIs decisively increase the power for differential gene expression from single cell RNA-seq libraries. However, the higher precision of CEL-seq comes with higher costs per library and in its current form also requires manual isolation of single cells (Table S1). Indeed, when we calculate the costs to reach a given power for differential gene expression, Drop-Seq and SCRB-seq are both more efficient than CEL-seq (Table 1). Smart-seq/C1 is by far the least efficient method due to its high costs and low precision. The efficiency could be improved for all methods, e.g. when implementing UMI-based protocols on the C1 platform[11], or using early barcoding to increase the number of cells per chip. The CEL-seq variant MARS-seq combines the higher precision of IVT-based amplification with higher throughput[14] and SCRB-seq is even more cost-efficient on a 384-well format[13].

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In summary, we find that Drop-seq is probably the most efficient method when analysing large numbers of cells and SCRB-seq might be preferable for analyzing smaller numbers, especially when considering that ERCCs can be spiked in, cells can be preselected by FACS and the same set-up can also be used for bulk RNA-seq.

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Published data

CEL-seq data for J1 mESC cultured in 2i/LIF condition[12] were obtained under accession GSE54695. Drop-seq ERCC[15] data were obtained under accession GSE66694. Raw fastq files were extracted using the SRA toolkit (2.3.5).

We trimmed cDNA reads to the same length and processed raw reads in the same way as data sequenced for this study (see below).

Cell culture of mESC

J1 mouse embryonic stem cells were maintained on gelatin-coated dishes in Dulbecco's modified Eagle's medium supplemented with 16% fetal bovine serum (FBS, Sigma-Aldrich), 0.1 mM β-mercaptoethanol (Invitrogen), 2 mM L-glutamine, 1x MEM non-essential amino acids, 100 U/ml penicillin, 100 µg/ml streptomycin (PAA Laboratories GmbH), 1000 U/ml recombinant mouse LIF (Millipore) and 2i (1µM PD032591 and 3µM CHIR99021 (Axon Medchem, Netherlands). J1 embryonic stem cells were obtained from E. Li and T. Chen and mycoplasma free determined by a PCR-based test. Cell line authentication was not recently performed.

Single cell RNA-seq library preparations

Drop-seq

Drop-seq experiments were performed as published[15] and successful establishment of the method in our lab was confirmed by a species-mixing experiment (data not shown). For this work, J1 mES cells (100/µl) and barcode-beads (120/µl, Chemgenes) were co-flown in Drop-seq PDMS devices (Nanoshift). Emulsions were broken by addition of perfluoroctanol (Sigma-Aldrich) and mRNA on beads reverse transcribed. Unused primers were degraded by addition of Exonuclease I (New England Biolabs). Washed beads were counted and aliquoted for pre-amplification (2000 beads / reaction). Nextera XT libraries were constructed from 1 ng of pre-amplified cDNA with a custom P5 primer (IDT).

SCRB-seq

RNA was stabilized by resuspending cells in RNAprotect Cell Reagent (Qiagen) and RNAse inhibitors (Promega). Prior to FACS sorting, cells were diluted in PBS (Invitrogen). Single cells were sorted into 5 µl lysis buffer consisting of a 1/500 dilution of Phusion HF buffer (New England Biolabs), spun down and frozen at -80 °C. Plates were thawed and libraries prepared as described previously[13]. Briefly, RNA was desiccated after protein digestion by

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Proteinase K (Ambion). RNA was reverse transcribed using barcoded oligo-dT primers (IDT) and products pooled and concentrated. Unincorporated barcode primers were digested using Exonuclease I (New England Biolabs). Pre-amplification of cDNA pools were done with the KAPA HiFi HotStart polymerase (KAPA Biosystems). Nextera XT libraries were constructed from 1 ng of pre-amplified cDNA with a custom P5 primer (IDT).

Smart-seq/C1

Smart-seq/C1 libraries were prepared on the Fluidigm C1 system according to the manufacturer's protocol. Cells were loaded on a 10-17 µm RNA-seq microfluidic IFC at a concentration of 200,000/ml. Capture site occupancy was surveyed using the Operetta (Perkin Elmer) automated imaging platform.

DNA sequencing

For each scRNA-seq method, final library pools were size-selected on 2% E-Gel Agarose EX Gels (Invitrogen) by excising a range of 300-800 bp and extracting DNA using the MinElute Kit (Qiagen) according to the manufacturer's protocol.
0.1 pmol of library pools were sequenced on a Illumina HiSeq1500 using High Output mode. Smart-seq/C1 libraries were sequenced 45 cycles single-end, whereas Drop-seq and SCRB-seq libraries were sequenced paired-end with 20 cycles to decode cell barcodes and UMI from read 1 and 45 cycles into the cDNA fragment.

Basic data processing and sequence alignment

Smart-seq/C1 libraries (i5 and i7) and Drop-seq/SCRB-seq pools (i7) were demultiplexed from the Nextera barcodes. All reads were mapped to the mouse genome (mm10) including mitochondrial genome sequences and unassigned scaffolds concatenated with the ERCC spike-in reference. Alignments were calculated using STAR 2.4.0[23] using all default parameters.
For libraries containing UMIs, cell- and gene-wise count/UMI tables were generated using the published Drop-seq pipeline[15]. We discarded the last 2 bases of the Drop-seq cell and molecular barcodes to account for bead synthesis errors.
For Smart-seq/C1, features were assigned and counted using the Rsubread package[24].

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Power Analysis

We utilized the R package PROPER [20] for statistical power evaluation. For each method, we estimated mean baseline expression and dispersions using DESeq2[21] and all genes with log2(dispersion) >-10 were considered for simulation. For each method, 100 RNA-seq experiments were simulated with 5 percent differentially expressed genes between two groups of sample size of 2^4, 2^5, 2^6, 2^7, 2^8 and 2^9 cells.

ERCC capture efficiency

To estimate the single molecule capture efficiency, we assume that the success or failure of detecting an ERCC is a binomial process, as described before[17]. Detections are independent from each other and are thus regarded as independent Bernoulli trials. We recorded the number of cells with nonzero and zero expression values for each ERCC per method and applied a maximum likelihood estimation to fit the probability of successful detection. The fit line was shaded with the 95% Wilson score confidence interval.

Variance decomposition

Variance decomposition was done as implemented by Kim et al[18]. We estimate the variation in capture efficiency (θ) and in sequencing efficiency (γ) separately for each replicate to account for batch effects. We integrated the published Drop-seq ERCC data by scaling it with size factors from our mESC data to make up 10% of counts in one cell.
We applied scLVM as described previously[19], both using the log-linear fit of mean and variance relationship as well as the ERCCs for estimation of technical variation.

Cost efficiency calculation

We based our cost efficiency extrapolation on the power simulations performed in this study (Fig. 4b). We determined the number of cells required per group for adequate power (80% in UMI-based methods and 50% for Smart-seq/C1) by an asymptotic fit to the median powers. Next, we determined the number of raw reads required to reach 100,000 exonic reads per cell. For this, we used the mean exon assignment rate from Fig. 2a (CEL-seq: 36%, Drop-seq: 40%, SCRB-seq: 44%, Smart-seq/C1: 44%). To account for reads that do not belong to desired cell-transcriptomes in early-barcoding methods, we assumed a loss of reads of 20% (CEL-seq), 50% (Drop-seq) and 20% (SCRB-seq). For the calculation of sequencing cost, we assumed 5€ per million raw reads, independent of method. Although UMI-based methods need paired-end sequencing, we assumed a 50 cycle sequencing kit is sufficient for all methods.

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data accession

The raw and analysed data files can be obtained in GEO under accession number GSE75790.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

CZ and WE conceived the experiments. CZ prepared scRNA-seq libraries and analyzed the data. SP helped in data processing and power simulations. BV estimated ERCC capture efficiencies. MS performed cell culture of mESC. WE and HL supervised the experimental work and IH provided guidance in data analysis. CZ and WE wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Christopher Mulholland for assistance with FACS sorting, Dominik Alterauge for help establishing the Drop-seq method and Stefan Krebs and Helmut Blum from the LAFUGA platform for sequencing. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through LMU excellence and the SFB1243 (Subproject A14) as well as a travel grant to CZ by the Boehringer Ingelheim Fonds.

References

1. Kim, H. D., Shay, T., O'Shea, E. K. & Regev, A. Transcriptional regulatory circuits: predicting numbers from alphabets. *Science* **325,** 429–432 (2009).
2. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
3. Sandberg, R. Entering the era of single-cell transcriptomics in biology and medicine. *Nat. Methods* **11,** 22–24 (2014).
4. 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).

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5. Eberwine, J., Sul, J.-Y., Bartfai, T. & Kim, J. The promise of single-cell sequencing. *Nat. Methods* **11,** 25–27 (2014).

6. Saliba, A.-E., Westermann, A. J., Gorski, S. A. & Vogel, J. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* **42,** 8845–8860 (2014).

7. Wu, A. R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11,** 41–46 (2014).

8. Fu, G. K., Hu, J., Wang, P.-H. & Fodor, S. P. A. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 9026–9031 (2011).

9. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2,** 666–673 (2012).

10. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9,** 72–74 (2012).

11. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11,** 163–166 (2014).

12. Grün, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nat. Methods* **11,** 637–640 (2014).

13. Soumillon *et al.* Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv* (2014). doi:10.1101/003236

14. 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).

15. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161,** 1202–1214 (2015).

16. Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21,** 1543–1551 (2011).

17. Marinov, G. K. *et al.* From single-cell to cell-pool transcriptomes: Stochasticity in gene

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression and RNA splicing. *Genome Res.* **24,** 496–510 (2014).

18. Kim, J. K., Kolodziejczyk, A. A., Illicic, T., Teichmann, S. A. & Marioni, J. C. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat. Commun.* **6,** 8687 (2015).
19. Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33,** 155–160 (2015).
20. Wu, H., Wang, C. & Wu, Z. PROPER: comprehensive power evaluation for differential expression using RNA-seq. *Bioinformatics* **31,** 233–241 (2015).
21. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
22. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. The impact of amplification on differential expression analyses by RNA-seq. *bioRxiv* 035493 (2015). doi:10.1101/035493
23. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
24. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41,** e108 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

Method	Power	n_{cells} (per group)	Library cost (€)	Sequencing cost (€)	Experiment cost (€)
CEL-seq	0.8[a]	52	~820[b]	~180[c]	~1000
Drop-seq	0.8[a]	170	~30[b]	~840[c]	~870
SCRB-seq	0.8[a]	112	~450[b]	~320[c]	~870
Smart-seq/C1	0.5[a]	129	~6560[b]	~290[c]	~6850

Table 1 | Cost efficiency extrapolation for single-cell RNA-seq experiments.

[a] Based on simulations (Fig. 4b) for detection of $\log_2$(fold-change) = 0.25

[b] assuming per cell library preparation costs of 8€, 0.1€, 2€, 25€ respectively

[c] Taking into account the mapping and exon assignment rates found in this study (see Fig. 2a), as well as the fraction of reads that falls into undesired cell barcodes in early-barcoding methods.

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

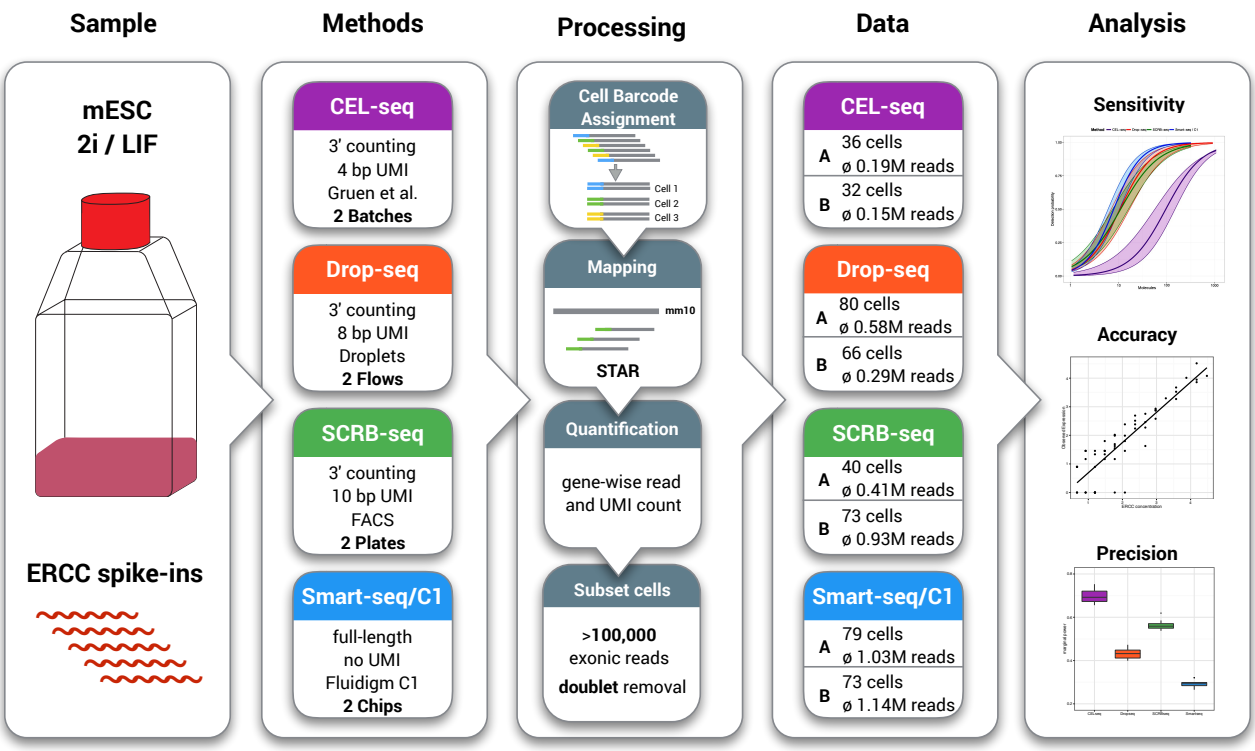

Figure 1 | Schematic of the experimental and computational pipeline. Mouse embryonic stem cells cultured in 2i/LIF and ERCC spike-in RNA were used to prepare single-cell RNA-seq libraries. The four methods differ by the presence and length of a unique molecular identifier sequence (UMI) allowing to identify reads generated during cDNA amplification. Data processing and subsetting of cells was done comparatively for all methods. Final cell numbers per method and replicate are shown with their mean sequencing depth. Colors represent the compared scRNA-seq methods: purple - CEL-seq, orange - Drop-seq, green SCRB-seq, blue - Smart-seq and are used throughout this study.

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

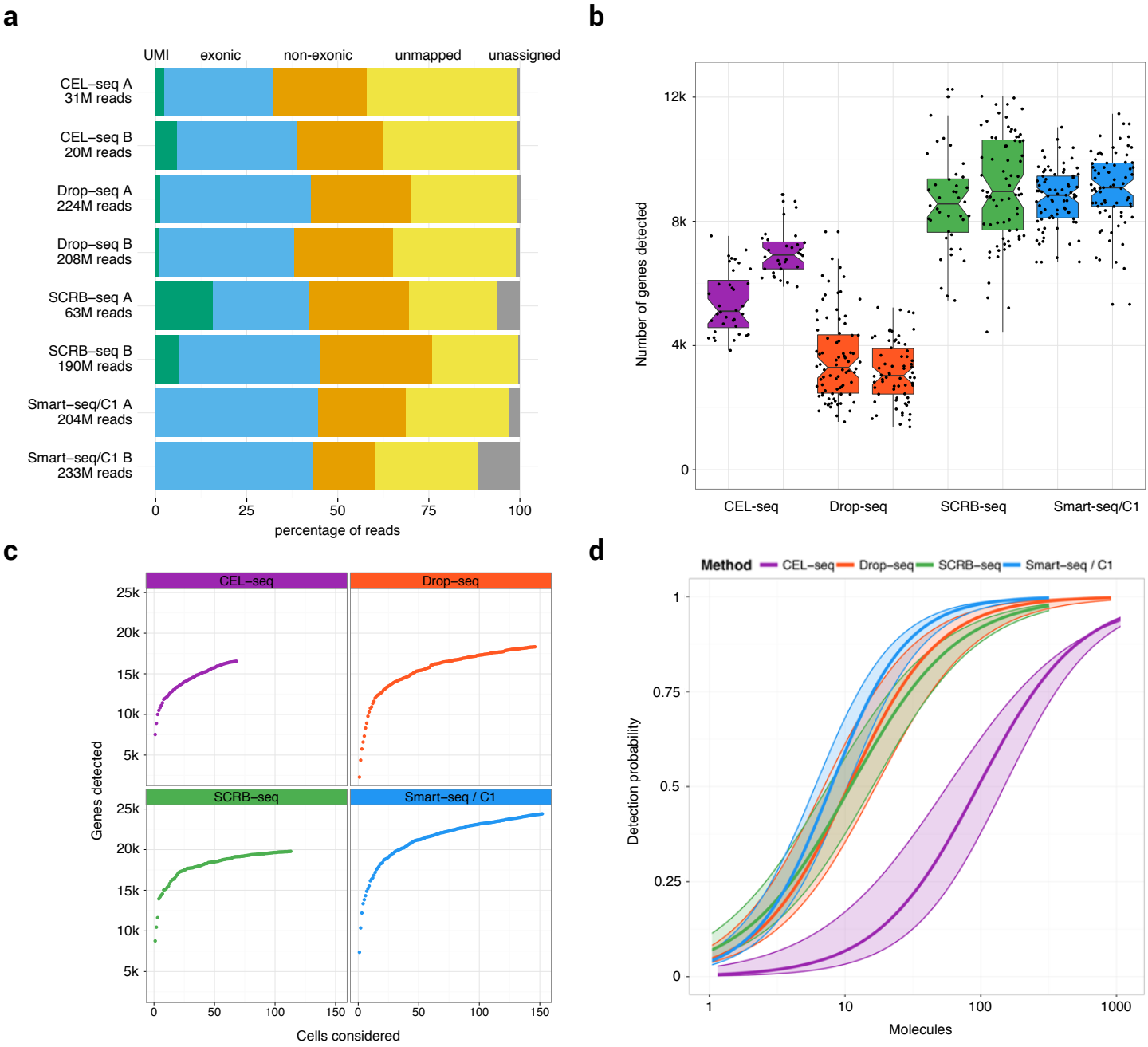

Figure 2 | Sensitivity of scRNA-seq methods.
(**a**) Percentage of the total reads that can not be assigned to a cellular barcode (grey), can not be mapped to the mouse genome (yellow), are mapped to regions outside exons (orange), inside exons (blue) and carry a unique UMI (green). (**b**) Number of genes with at least one read in all cells with more than 100,000 total reads in total. Each dot represents a cell and each boxplot represents the median, first and third quartile per replicate and method. (**c**) Cumulative number of genes detected as more cells are added. (**d**) Sensitivity per method estimated as the probability to detect ERCC transcripts dependent on their copy number per cell. The 95% confidence interval of an estimate is displayed as shaded area.

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

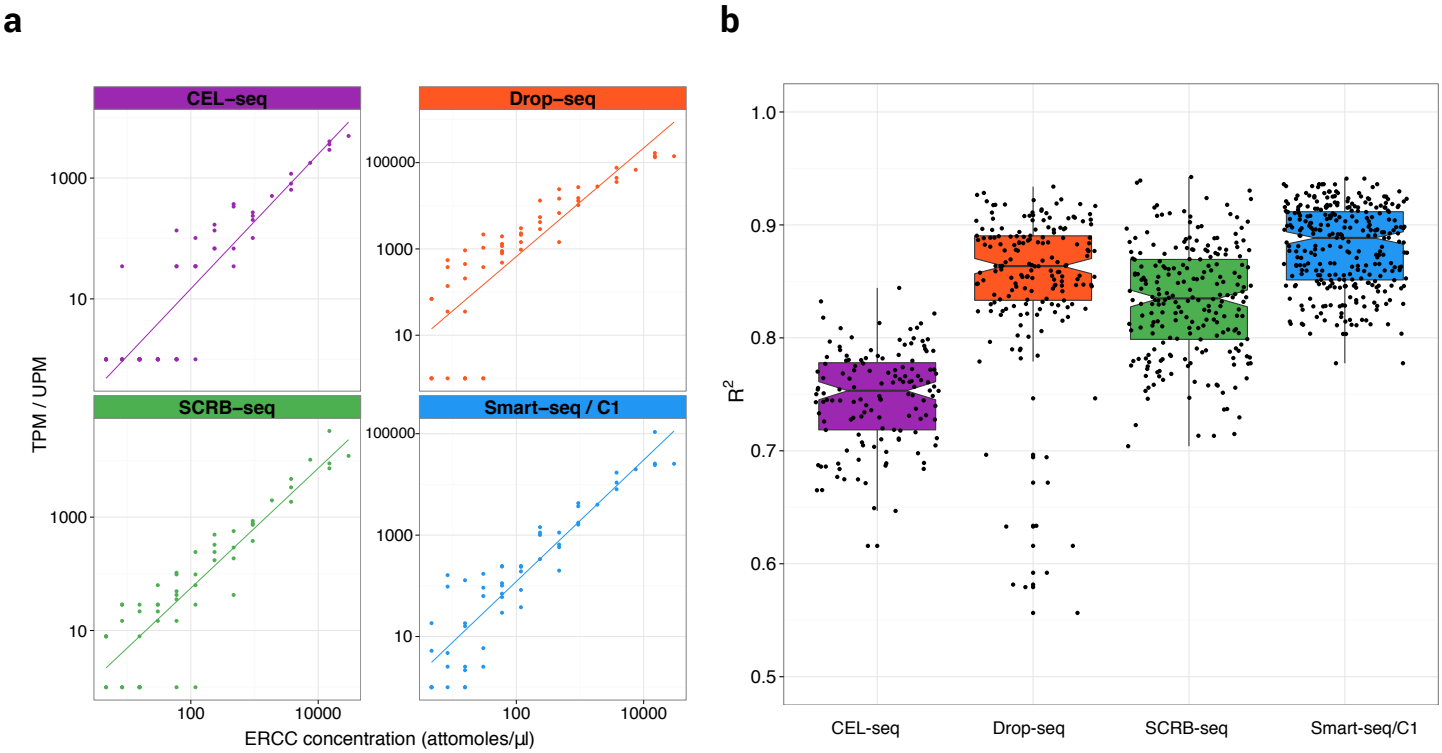

Figure 3 | Accuracy of scRNA-seq methods. (**a**) Representative correlations of single-cell / single-bead (Drop-seq) expression values for ERCC spike-ins and their annotated molarity. Expression is measured as transcripts per million reads (TPM, Smart-seq/C1) or UMIs per million reads (UPM). (**b**) Distribution of correlation coefficients (Pearson) across methods. Each dot represents a cell/bead and each boxplot represents the median, first and third quartile.

bioRxiv preprint doi: https://doi.org/10.1101/035758; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

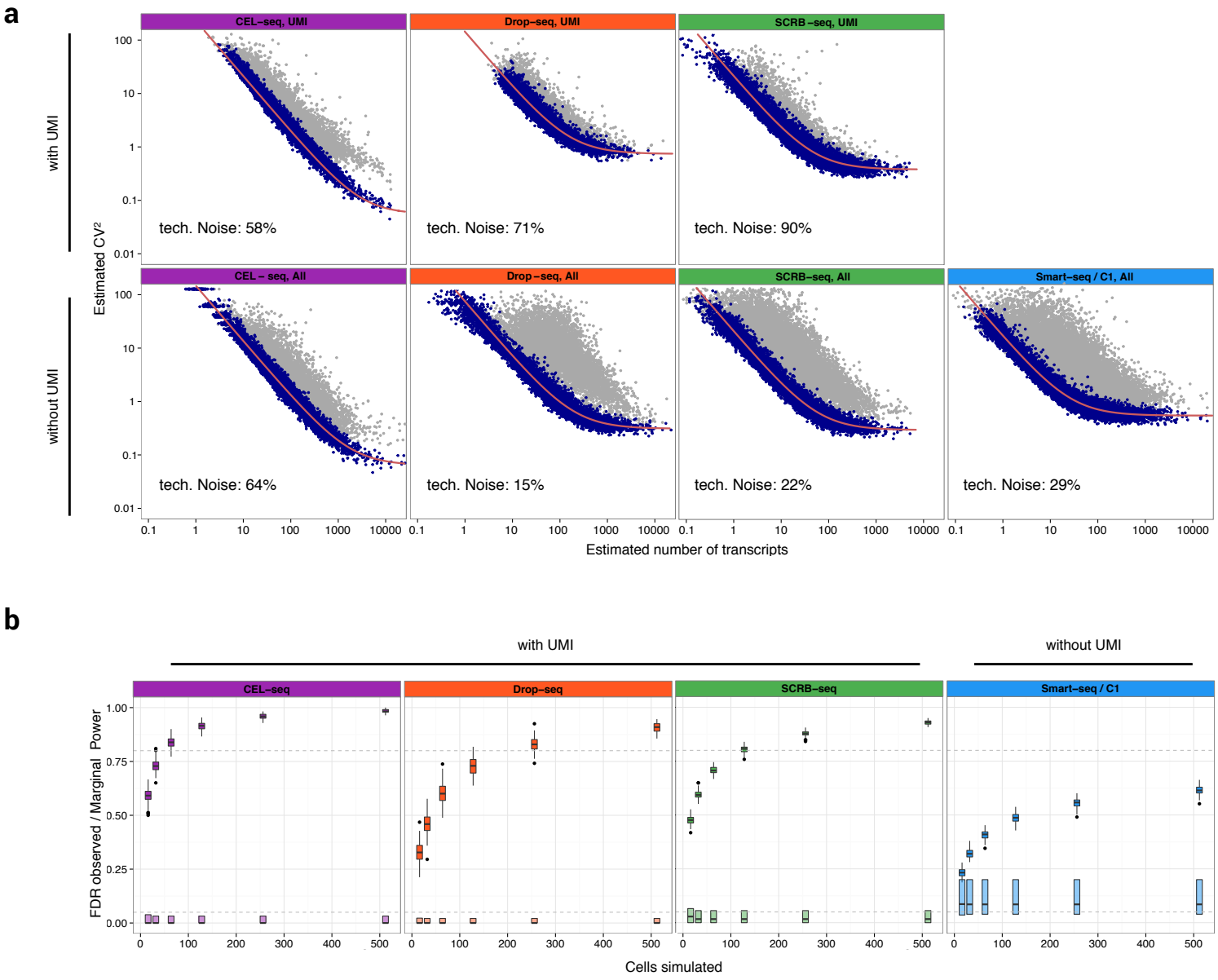

Figure 4 | Variance decomposition and power analysis. (**a**) Variance decomposition according to Kim et al. 2015. Technical and biological noise factors were estimated from complete read counts and UMI counts (where available) and further used to simulate mean / variance distributions per scRNA-seq method. Shown are simulated genes with technical and technical plus biological factor. Genes with technical factor only are shown in blue, with technical and biological factor in grey and the technical noise fit for ERCC spike-ins is shown in red. (**b**) Power simulations using empirical mean/ dispersion relationships. Boxplots represent the median, first and third quartile of 100 simulations. Power analysis of differentiation scenario with a small fixed effect size of 0.25 log2 fold-change and variable sample size n=16, n=32, n=64, n=128, n=256 and n=512. Shown are marginal power and observed FDR over 100 simulations with relevant cut-offs marked in dashed lines (marginal power = 0.8; FDR = 0.05).