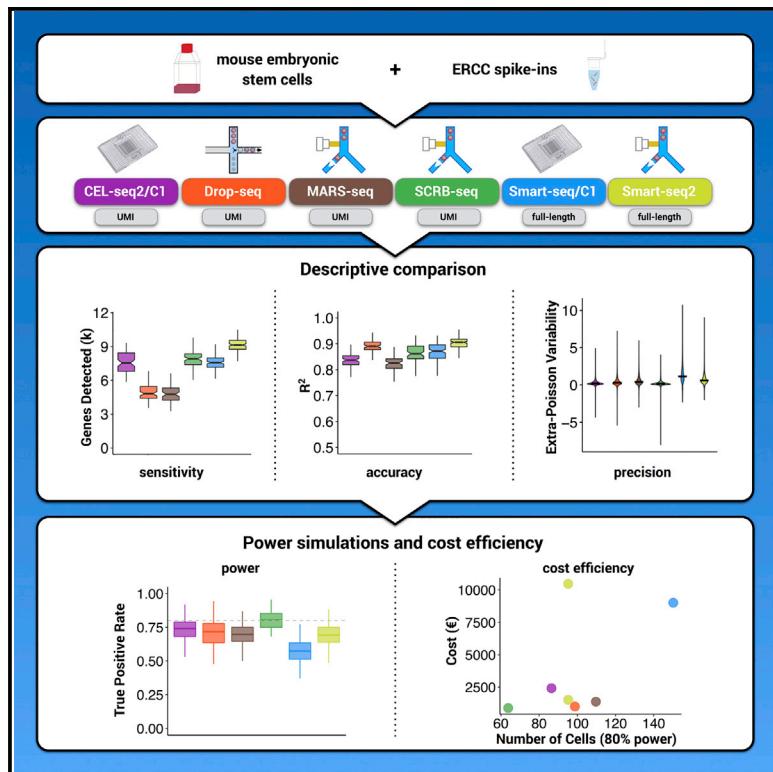


# Molecular Cell

## Comparative Analysis of Single-Cell RNA Sequencing Methods

### Graphical Abstract



### Authors

Christoph Ziegenhain, Beate Vieth,  
Swati Parekh, ..., Holger Heyn,  
Ines Hellmann, Wolfgang Enard

### Correspondence

enard@bio.lmu.de

### In Brief

Ziegenhain et al. generated data from mouse ESCs to systematically evaluate six prominent scRNA-seq methods. They used power simulations to compare cost efficiencies, allowing for informed choice among existing protocols and providing a framework for future comparisons.

### Highlights

- The study represents the most comprehensive comparison of scRNA-seq protocols
- Power simulations quantify the effect of sensitivity and precision on cost efficiency
- The study offers an informed choice among six prominent scRNA-seq methods
- The study provides a framework for benchmarking future protocol improvements

# Comparative Analysis of Single-Cell RNA Sequencing Methods

Christoph Ziegenhain,<sup>1</sup> Beate Vieth,<sup>1</sup> Swati Parekh,<sup>1</sup> Björn Reinius,<sup>2,3</sup> Amy Guillaumet-Adkins,<sup>4,5</sup> Martha Smets,<sup>6</sup> Heinrich Leonhardt,<sup>6</sup> Holger Heyn,<sup>4,5</sup> Ines Hellmann,<sup>1</sup> and Wolfgang Enard<sup>1,7,\*</sup>

<sup>1</sup>Anthropology & Human Genomics, Department of Biology II, Ludwig-Maximilians University, Großhaderner Straße 2, 82152 Martinsried, Germany

<sup>2</sup>Ludwig Institute for Cancer Research, Box 240, 171 77 Stockholm, Sweden

<sup>3</sup>Department of Cell and Molecular Biology, Karolinska Institutet, 171 77 Stockholm, Sweden

<sup>4</sup>CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain

<sup>5</sup>Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain

<sup>6</sup>Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians University, Großhaderner Straße 2, 82152 Martinsried, Germany

<sup>7</sup>Lead Contact

\*Correspondence: enard@bio.lmu.de

<http://dx.doi.org/10.1016/j.molcel.2017.01.023>

## SUMMARY

Single-cell RNA sequencing (scRNA-seq) offers new possibilities to address biological and medical questions. However, systematic comparisons of the performance of diverse scRNA-seq protocols are lacking. We generated data from 583 mouse embryonic stem cells to evaluate six prominent scRNA-seq methods: CEL-seq2, Drop-seq, MARS-seq, SCRБ-seq, Smart-seq, and Smart-seq2. While Smart-seq2 detected the most genes per cell and across cells, CEL-seq2, Drop-seq, MARS-seq, and SCRБ-seq quantified mRNA levels with less amplification noise due to the use of unique molecular identifiers (UMIs). Power simulations at different sequencing depths showed that Drop-seq is more cost-efficient for transcriptome quantification of large numbers of cells, while MARS-seq, SCRБ-seq, and Smart-seq2 are more efficient when analyzing fewer cells. Our quantitative comparison offers the basis for an informed choice among six prominent scRNA-seq methods, and it provides a framework for benchmarking further improvements of scRNA-seq protocols.

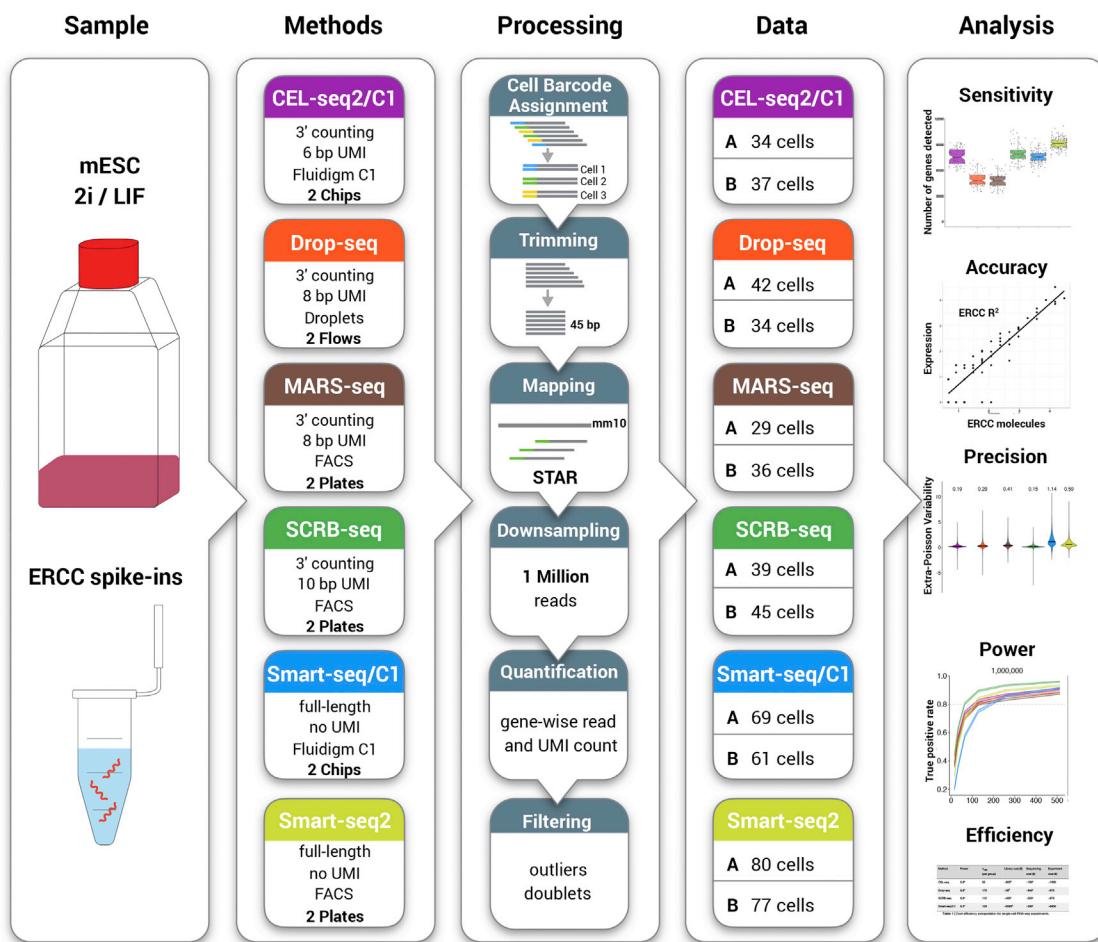
## INTRODUCTION

Genome-wide quantification of mRNA transcripts is highly informative for characterizing cellular states and molecular circuitries (ENCODE Project Consortium, 2012). Ideally, such data are collected with high spatial resolution, and single-cell RNA sequencing (scRNA-seq) now allows for transcriptome-wide analyses of individual cells, revealing exciting biological and medical insights (Kolodziejczyk et al., 2015a; Wagner et al., 2016). scRNA-seq requires the isolation and lysis of single cells, the conversion of their RNA into cDNA, and the amplification of cDNA to generate high-throughput sequencing libraries. As the

amount of starting material is so small, this process results in substantial technical variation (Kolodziejczyk et al., 2015a; Wagner et al., 2016).

One type of technical variable is the sensitivity of a scRNA-seq method (i.e., the probability to capture and convert a particular mRNA transcript present in a single cell into a cDNA molecule present in the library). Another variable of interest is the accuracy (i.e., how well the read quantification corresponds to the actual concentration of mRNAs), and a third type is the precision with which this amplification occurs (i.e., the technical variation of the quantification). The combination of sensitivity, precision, and number of cells analyzed determines the power to detect relative differences in expression levels. Finally, the monetary cost to reach a desired level of power is of high practical relevance. To make a well-informed choice among available scRNA-seq methods, it is important to quantify these parameters comparably. Some strengths and weaknesses of different methods are already known. For example, it has previously been shown that scRNA-seq conducted in the small volumes available in the automated microfluidic platform from Fluidigm (C1 platform) outperforms CEL-seq2, Smart-seq, or other commercially available kits in microliter volumes (Hashimshony et al., 2016; Wu et al., 2014). Furthermore, the Smart-seq protocol has been optimized for sensitivity, more even full-length coverage, accuracy, and cost (Picelli et al., 2013), and this improved Smart-seq2 protocol (Picelli et al., 2014b) has also become widely used (Gokce et al., 2016; Reinius et al., 2016; Tirosh et al., 2016).

Other protocols have sacrificed full-length coverage in order to sequence part of the primer used for cDNA generation. This enables early barcoding of libraries (i.e., the incorporation of cell-specific barcodes), allowing for multiplexing the cDNA amplification and thereby increasing the throughput of scRNA-seq library generation by one to three orders of magnitude (Hashimshony et al., 2012; Jaitin et al., 2014; Klein et al., 2015; Macosko et al., 2015; Soumillon et al., 2014). Additionally, this approach allows the incorporation of unique molecular identifiers (UMIs), random nucleotide sequences that tag individual



**Figure 1. Schematic Overview of the Experimental and Computational Workflow**

Mouse embryonic stem cells (mESCs) cultured in 2i/LIF and ERCC spike-in RNAs were used to generate single-cell RNA-seq data with six different library preparation methods (CEL-seq2/C1, Drop-seq, MARS-seq, SCRB-seq, Smart-seq/C1, and Smart-seq2). The methods differ in the usage of unique molecular identifier (UMI) sequences, which allow the discrimination between reads derived from original mRNA molecules and duplicates generated during cDNA amplification. Data processing was identical across methods, and the given cell numbers per method and replicate were used to compare sensitivity, accuracy, precision, power, and cost efficiency. The six scRNA-seq methods are denoted by color throughout the figures of this study as follows: purple, CEL-seq2/C1; orange, Drop-seq; brown, MARS-seq; green, SCRB-seq; blue, Smart-seq; and yellow, Smart-seq2. See also Figures S1 and S2.

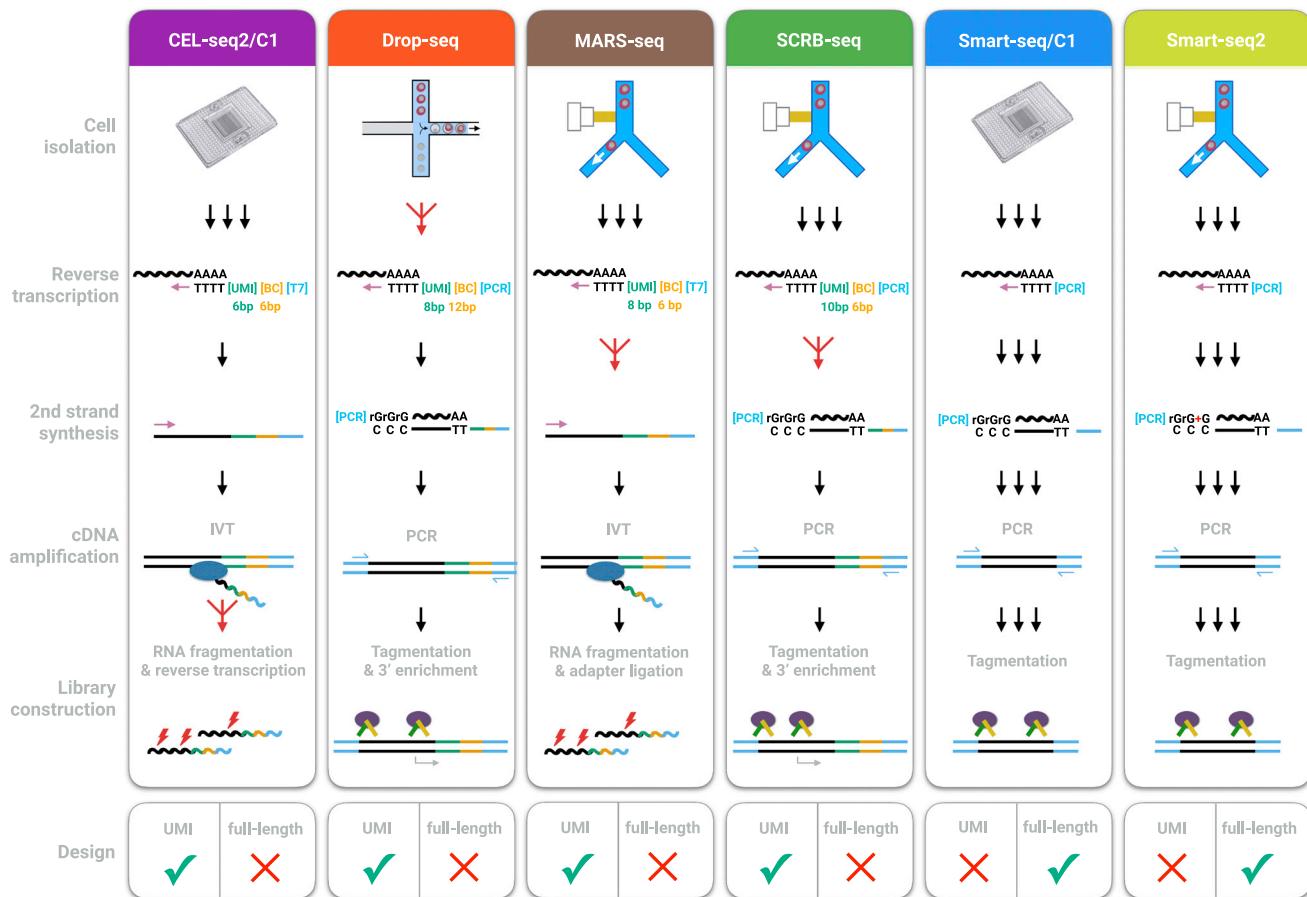
mRNA molecules and, hence, allow for the distinction between original molecules and amplification duplicates that derive from the cDNA or library amplification (Kivioja et al., 2011). Utilization of UMI information improves quantification of mRNA molecules (Grün et al., 2014; Islam et al., 2014), and it has been implemented in several scRNA-seq protocols, such as STRT (Islam et al., 2014), CEL-seq (Grün et al., 2014; Hashimshony et al., 2016), CEL-seq2 (Hashimshony et al., 2016), Drop-seq (Mácosko et al., 2015), inDrop (Klein et al., 2015), MARS-seq (Jaitin et al., 2014), and SCRB-seq (Soumillon et al., 2014).

However, a thorough and systematic comparison of relevant parameters across scRNA-seq methods is still lacking. To address this issue, we generated 583 scRNA-seq libraries from mouse embryonic stem cells (mESCs), using six different methods in two replicates, and we compared their sensitivity, accuracy, precision, power, and efficiency (Figure 1).

## RESULTS

### Generation of scRNA-Seq Libraries

Variation in gene expression as observed among single cells is caused by biological and technical variation (Kolodziejczyk et al., 2015a; Wagner et al., 2016). We used mESCs cultured under two inhibitor/leukemia inhibitory factor (2i/LIF) conditions to obtain a relatively homogeneous cell population (Grün et al., 2014; Kolodziejczyk et al., 2015b), so that biological variation was similar among experiments and, hence, we mainly compared technical variation. In addition, we spiked in 92 poly-adenylated synthetic RNA transcripts of known concentration designed by the External RNA Control Consortium (ERCCs) (Jiang et al., 2011). For all six tested scRNA-seq methods (Figure 2), we generated libraries in two independent replicates.



**Figure 2. Schematic Overview of Library Preparation Steps**

For details, see the text. See also Table S1.

For each replicate of the Smart-seq protocol, we performed one run on the C1 platform from Fluidigm (Smart-seq/C1) using microfluidic chips that automatically capture up to 96 cells (Wu et al., 2014). We imaged captured cells, added lysis buffer together with the ERCCs, and we used the commercially available Smart-seq kit (Clontech) to generate full-length double-stranded cDNA that we converted into 96 sequencing libraries by tagmentation (Nextera, Illumina).

For each replicate of the Smart-seq2 protocol, we sorted mESCs by fluorescence activated cell sorting (FACS) into 96-well PCR plates containing lysis buffer and the ERCCs. We generated cDNA as described (Picelli et al., 2013, 2014b), and we used an in-house-produced Tn5 transposase (Picelli et al., 2014a) to generate 96 libraries by tagmentation. While SmartSeq/C1 and Smart-seq2 are very similar protocols that generate full-length libraries, they differ in how cells are isolated, their reaction volume, and in that the Smart-seq2 chemistry has been systematically optimized (Picelli et al., 2013, 2014b). The main disadvantage of both Smart-seq protocols is that the generation of full-length cDNA libraries precludes an early barcoding step and the incorporation of UMIs.

For each replicate of the SCRB-seq protocol (Soumillon et al., 2014), we also sorted mESCs by FACS into 96-well PCR plates

containing lysis buffer and the ERCCs. Similar to the Smart-seq protocols, cDNA was generated by oligo-dT priming, template switching, and PCR amplification of full-length cDNA. However, the oligo-dT primers contained well-specific (i.e., cell-specific) barcodes and UMIs. Hence, cDNA from one plate could be pooled and then converted into sequencing libraries, using a modified tagmentation approach that enriches for the 3' ends. SCRB-seq is optimized for small volumes and few handling steps.

The fourth method evaluated was Drop-seq, a recently developed microdroplet-based approach (Macosko et al., 2015). Here a flow of beads suspended in lysis buffer and a flow of a single-cell suspension were brought together in a microfluidic chip that generated nanoliter-sized emulsion droplets. On each bead, oligo-dT primers carrying a UMI and a unique, bead-specific barcode were covalently bound. Cells were lysed within these droplets, their mRNA bound to the oligo-dT-carrying beads, and, after breaking the droplets, cDNA and library generation was performed for all cells in parallel in one single tube. The ratio of beads to cells (20:1) ensured that the vast majority of beads had either no cell or one cell in its droplet. Hence, similar to SCRB-seq, each cDNA molecule was labeled with a bead-specific (i.e., cell-specific) barcode and a UMI. We confirmed that

the Drop-seq protocol worked well in our setup by mixing mouse and human T cells, as recommended by Macosko et al. (2015) (Figure S1A). The main advantage of the protocol is that a high number of scRNA-seq libraries can be generated at low cost. One disadvantage of Drop-seq is that the simultaneous inclusion of ERCC spike-ins is quite expensive, as their addition would generate cDNA from ERCCs also in beads that have zero cells and thus would double the sequencing costs. As a proxy for the missing ERCC data, we used a published dataset (Macosko et al., 2015), where ERCC spike-ins were sequenced using the Drop-seq method without single-cell transcriptomes.

As a fifth method we chose CEL-seq2 (Hashimshony et al., 2016), an improved version of the original CEL-seq (Hashimshony et al., 2012) protocol, as implemented for microfluidic chips on Fluidigm's C1 (Hashimshony et al., 2016). As for Smart-seq/C1, this allowed us to capture 96 cells in two independent replicates and to include ERCCs in the cell lysis step. Similar to Drop-seq and SCRB-seq, cDNA was tagged with barcodes and UMIs; but, in contrast to the four PCR-based methods described above, CEL-seq2 relies on linear amplification by *in vitro* transcription after the initial reverse transcription. The amplified, barcoded RNAs were harvested from the chip, pooled, fragmented, and reverse transcribed to obtain sequencing libraries.

MARS-seq, the sixth method evaluated, is a high-throughput implementation of the original CEL-seq method (Jaitin et al., 2014). In this protocol, cells were sorted by FACS in 384-well plates containing lysis buffer and the ERCCs. As in CEL-seq and CEL-seq2, amplified RNA with barcodes and UMIs were generated by *in vitro* transcription, but libraries were prepared on a liquid-handling platform. An overview of the methods and their workflows is provided in Figure 2 and in Table S1.

### Processing of scRNA-Seq Data

For each method, we generated at least 48 libraries per replicate and sequenced between 241 and 866 million reads (Figure 1; Figure S1B). All data were processed identically, with cDNA reads clipped to 45 bp and mapped using Spliced Transcripts Alignment to a Reference (STAR) (Dobin et al., 2013) and UMIs quantified using the Drop-seq pipeline (Macosko et al., 2015). To adjust for differences in sequencing depths, we selected all libraries with at least one million reads, and we downsampled them to one million reads each. This resulted in 96, 79, 73, 93, 162, and 187 libraries for CEL-seq2/C1, Drop-seq, MARS-seq, SCRB-seq, Smart-seq/C1, and Smart-seq2, respectively.

To exclude doublets (libraries generated from two or more cells) in the Smart-seq/C1 data, we analyzed microscope images and identified 16 reaction chambers with multiple cells. For the four UMI methods, we calculated the number of UMIs per library, and we found that libraries that have more than twice the mean total UMI count can be readily identified (Figure S1C). It is unclear whether these libraries were generated from two separate cells (doublets) or, for example, from one large cell before mitosis. However, for the purpose of this method comparison, we removed these three to nine libraries. To filter out low-quality libraries, we used a method that exploits the fact that transcript detection and abundance in low-quality libraries correlate poorly with high-quality libraries as well as with other low-quality libraries (Petropoulos et al., 2016). Therefore, we determined

the maximum Spearman correlation coefficient for each cell in all-to-all comparisons that allowed us to identify low-quality libraries as outliers of the distributions of correlation coefficients by visual inspection (Figure S1D). This filtering led to the removal of 21, 0, 4, 0, 16, and 30 cells for CEL-seq2/C1, Drop-seq, MARS-seq, SCRB-seq, Smart-seq/C1, and Smart-seq2, respectively.

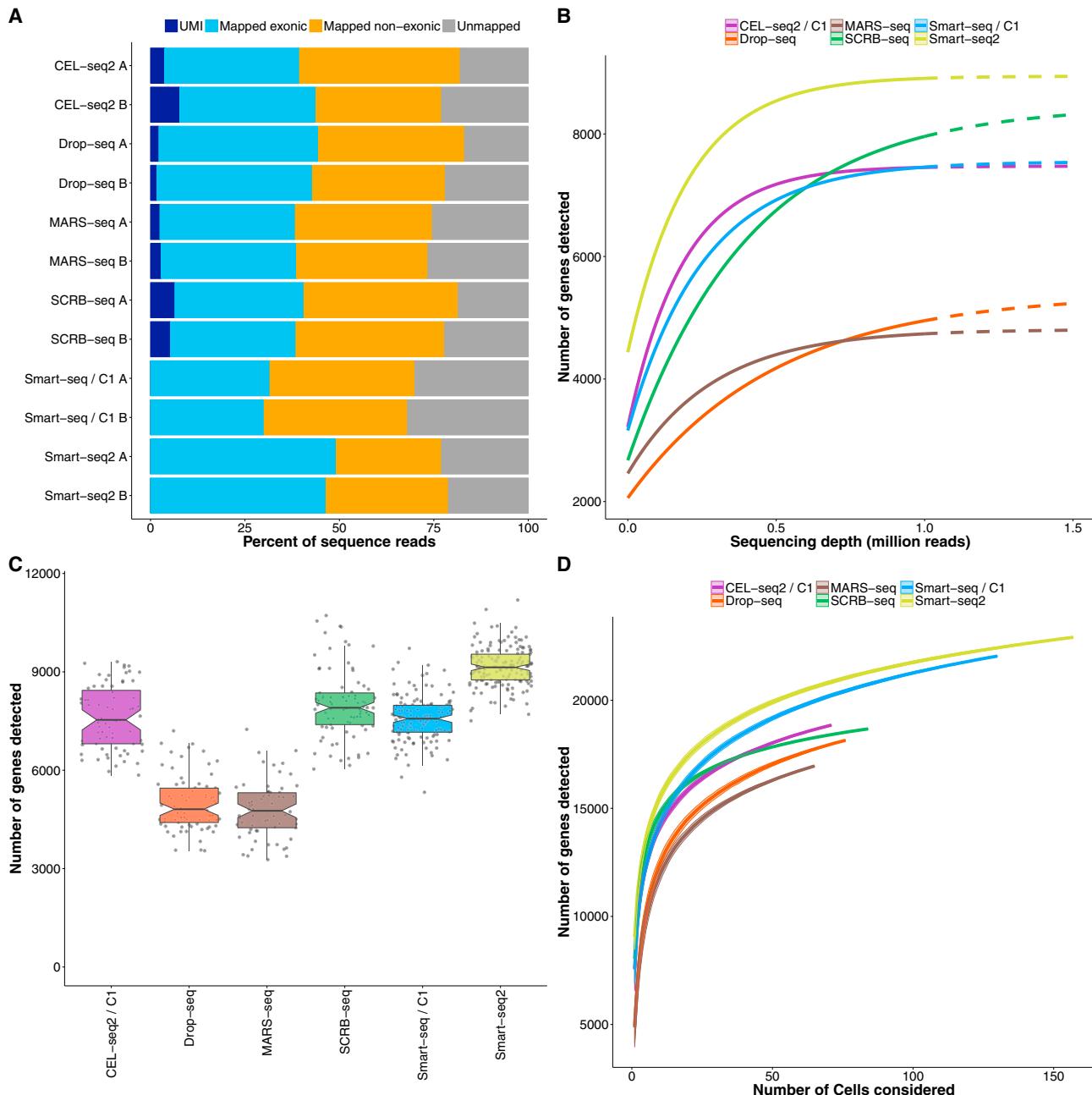
In summary, we processed and filtered our data so that we ended up with a total of 583 high-quality scRNA-seq libraries that could be used for a fair comparison of the sensitivity, accuracy, precision, power, and efficiency of the methods.

### Single-Cell Libraries Are Sequenced to a Reasonable Level of Saturation at One Million Reads

For all six methods, >50% of the reads could be unambiguously mapped to the mouse genome (Figure 3A), which is comparable to previous results (Jaitin et al., 2014; Wu et al., 2014). Overall, between 48% (Smart-seq2) and 30% (Smart-seq/C1) of all reads were exonic and, thus, were used to quantify gene expression levels. However, the UMI data showed that only 14%, 5%, 7%, and 15% of the exonic reads were derived from independent mRNA molecules for CEL-seq2/C1, Drop-seq, MARS-seq, and SCRB-seq, respectively (Figure 3A). To quantify the relationship between the number of detected genes or mRNA molecules and the number of reads in more detail, we downsampled reads to varying depths, and we estimated to what extent libraries were sequenced to saturation (Figure S2). The number of unique mRNA molecules plateaued at 56,760 UMIs per library for CEL-seq2/C1 and 26,210 UMIs per library for MARS-seq, was still marginally increasing at 17,210 UMIs per library for Drop-seq, and was considerably increasing at 49,980 UMIs per library for SCRB-seq (Figure S2C). Notably, CEL-seq2/C1 and MARS-seq showed a steeper slope at low sequencing depths than both Drop-seq and SCRB-seq, potentially due to a less biased amplification by *in vitro* transcription. Hence, among the UMI methods, CEL-seq2/C1 and SCRB-seq libraries had the highest complexity of mRNA molecules, and this complexity was sequenced to a reasonable level of saturation with one million reads.

To investigate saturation also for non-UMI-based methods, we applied a similar approach at the gene level by counting the number of genes detected by at least one read. By fitting an asymptote to the downsampled data, we estimated that ~90% (Drop-seq and SCRB-seq) to 100% (CEL-seq2/C1, MARS-seq, Smart-Seq/C1, and Smart-seq2) of all genes present in a library were detected at one million reads (Figure 3B; Figure S2A). In particular, the deep sequencing of Smart-seq2 libraries showed clearly that the number of detected genes did not change when increasing the sequencing depth from one million to five million reads per cell (Figure S2B).

All in all, these analyses show that scRNA-seq libraries were sequenced to a reasonable level of saturation at one million reads, a cutoff that also has been suggested previously for scRNA-seq datasets (Wu et al., 2014). While it can be more efficient to invest in more cells at lower coverage (see our power analyses below), one million reads per cell is a reasonable sequencing depth for our purpose of comparing scRNA-seq methods.



**Figure 3. Sensitivity of scRNA-Seq Methods**

(A) Percentage of reads (downsampled to one million per cell) that cannot be mapped to the mouse genome (gray) are mapped to regions outside exons (orange) or inside exons (blue). For UMI methods, dark blue denotes the exonic reads with unique UMIs.

(B) Median number of genes detected per cell (counts  $\geq 1$ ) when downsampling total read counts to the indicated depths. Dashed lines above one million reads represent extrapolated asymptotic fits.

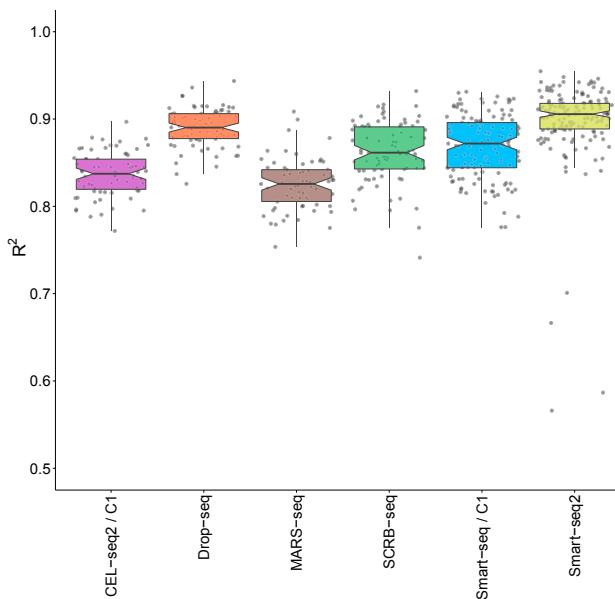
(C) Number of genes detected (counts  $\geq 1$ ) per cell. Each dot represents a cell and each box represents the median and first and third quartiles per replicate and method.

(D) Cumulative number of genes detected as more cells are added. The order of cells considered was drawn randomly 100 times to display mean  $\pm$  SD (shaded area). See also Figures S3 and S4.

### Smart-Seq2 Has the Highest Sensitivity

Taking the number of detected genes per cell as a measure of sensitivity, we found that Drop-seq and MARS-seq had the lowest

sensitivity, with a median of 4,811 and 4,763 genes detected per cell, respectively, while CEL-seq2/C1, SCRB-seq, and Smart-seq/C1 detected a median of 7,536, 7,906, and 7,572 genes per



**Figure 4. Accuracy of scRNA-Seq Methods**

ERCC expression values (counts per million reads for Smart-seq/C1 and Smart-seq2 and UMIs per million reads for all others) were correlated to their annotated molarity. Shown are the distributions of correlation coefficients (adjusted  $R^2$  of linear regression model) across methods. Each dot represents a cell/bead and each box represents the median and first and third quartiles. See also Figure S5.

cell (Figure 3C). Smart-seq2 detected the highest number of genes per cell with a median of 9,138. To compare the total number of genes detected across many cells, we pooled the sequence data of 65 cells per method, and we detected ~19,000 genes for CEL-Seq2/C1, ~17,000 for MARS-seq, ~18,000 for Drop-seq and SCRB-Seq, ~20,000 for Smart-seq/C1, and ~21,000 for Smart-seq2 (Figure 3D). While the majority of genes (~13,000) were detected by all methods, ~400 genes were specific to each of the 3' counting methods, and ~1,000 genes were specific to each of the two full-length methods (Figure S3A). This higher sensitivity of both full-length methods also was apparent when plotting the genes detected in all available cells, as the 3' counting methods leveled off below 20,000 genes while the two full-length methods leveled off above 20,000 genes (Figure 3D). Such a difference could be caused by genes that have 3' ends that are difficult to map. However, we found that genes specific to Smart-seq2 and Smart-seq/C1 map as well to 3' ends as genes with similar length distribution that are not specifically detected by full-length methods (Figure S3B). Hence, it seems that full-length methods turn a slightly higher fraction of transcripts into sequenceable molecules than 3' counting methods and are more sensitive in this respect. Importantly, method-specific genes are detected in very few cells (87% of genes occur in one or two cells) with very low counts (mean counts < 0.2, Figure S3C). This suggests that they are unlikely to remain method specific at higher expression levels and that their impact on conclusions drawn from scRNA-seq data is rather limited (Lun et al., 2016).

Next, we investigated how reads are distributed along the mRNA transcripts for all genes. As expected, the 3' counting

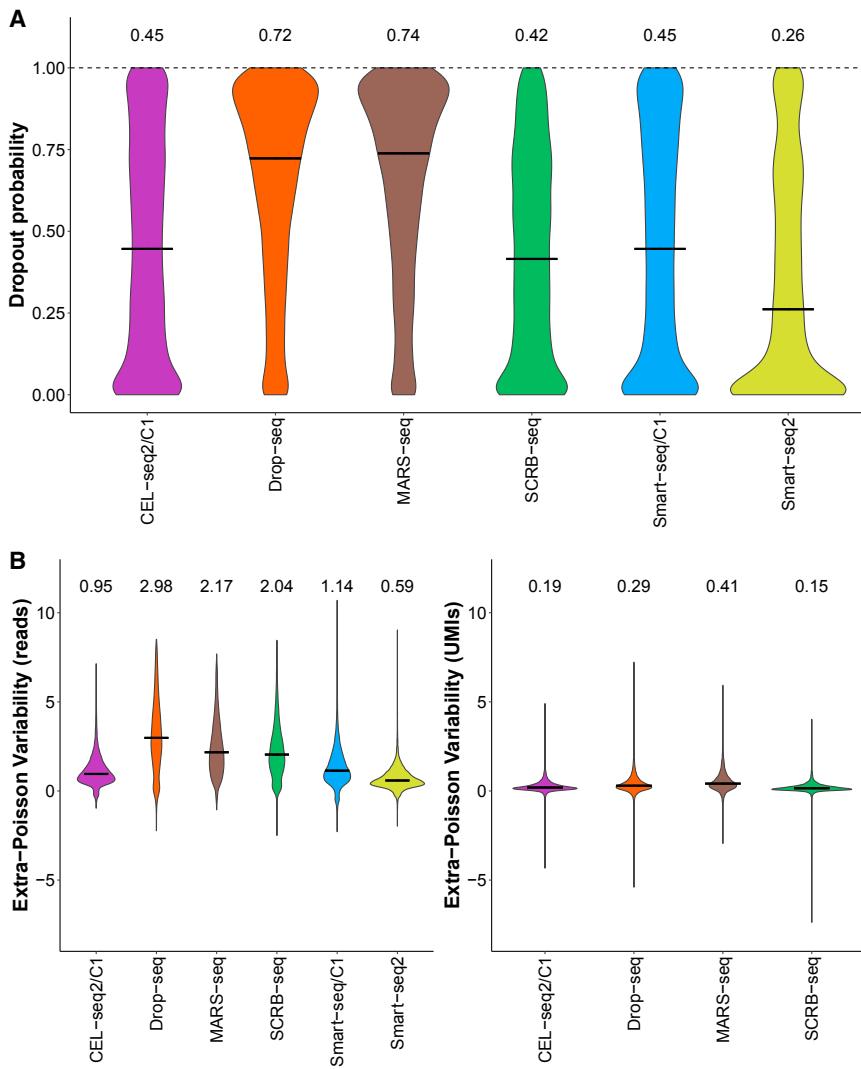
methods showed a strong bias of reads mapped to the 3' end (Figure S3D). However, it is worth mentioning that a considerable fraction of reads also covered other segments of the transcripts, probably due to internal oligo-dT priming (Nam et al., 2002). Smart-seq2 showed a more even coverage than Smart-seq, confirming previous findings (Picelli et al., 2013). A general difference in expression values between 3' counting and full-length methods also was reflected in their strong separation by the first principal component, explaining 37% of the total variance, and when taking into account that one needs to normalize for gene length for the full-length methods (Figure S4E).

As an absolute measure of sensitivity, we compared the probability of detecting the 92 spiked-in ERCCs, for which the number of molecules available for library construction is known (Figures S4A and S4B). We determined the detection probability of each ERCC RNA as the proportion of cells with at least one read or UMI count for the particular ERCC molecule (Marinov et al., 2014). For Drop-seq, we used the previously published ERCC-only dataset (Macosko et al., 2015), and for the other five methods, 2%–5% of the one million reads per cell mapped to ERCCs that were sequenced to complete saturation at that level (Figure S5B). A 50% detection probability was reached at ~7, 11, 14, 16, 17, and 28 ERCC molecules for Smart-seq2, Smart-seq/C1, CEL-seq2/C1, SCRB-seq, Drop-seq, and MARS-seq, respectively (Figure S4C). Notably, the sensitivity estimated from the number of detected genes does not fully agree with the comparison based on ERCCs. While Smart-seq2 was the most sensitive method in both cases, Drop-seq performed better and SCRB-seq and MARS-seq performed worse when using ERCCs. The separate generation and sequencing of the Drop-seq ERCC libraries could be a possible explanation for their higher sensitivity. However, it remains unclear why SCRB-seq and MARS-seq had a substantially lower sensitivity when using ERCCs. It has been noted before that ERCCs can be problematic for modeling endogenous mRNAs (Risso et al., 2014), potentially due to their shorter length, shorter poly-A tail, and their missing 5' cap (Grün and van Oudenaarden, 2015; Stegle et al., 2015). While ERCCs are still useful to gauge the absolute range of sensitivities, the thousands of endogenous mRNAs are likely to be a more reliable estimate for comparing sensitivities as we used the same cell type for all methods.

In summary, we find that Smart-seq2 is the most sensitive method, as it detects the highest number of genes per cell and the most genes in total across cells and has the most even coverage across transcripts. Smart-seq/C1 is slightly less sensitive per cell and detects almost the same number of genes across cells with slightly less even coverage. Among the 3' counting methods, CEL-seq2/C1 and SCRB-seq detect about as many genes per cell as Smart-seq/C1, whereas Drop-seq and MARS-seq detect considerably fewer genes.

#### Accuracy of scRNA-Seq Methods

To measure the accuracy of transcript level quantifications, we compared the observed expression values (counts per million or UMIs per million) with the known concentrations of the 92 ERCC transcripts (Figure S5A). For each cell, we calculated the coefficient of determination ( $R^2$ ) for a linear model fit (Figure 4). Methods differed significantly in their accuracy (Kruskal-Wallis

**Figure 5. Precision of scRNA-Seq Methods**

We compared precision among methods using the 13,361 genes detected in at least 25% of all cells by any method in a subsample of 65 cells per method.

(A) Distributions of dropout rates across the 13,361 genes are shown as violin plots, and medians are shown as bars and numbers.

(B) Extra Poisson variability across the 13,361 genes was calculated by subtracting the expected amount of variation due to Poisson sampling (square root of mean divided by mean) from the CV (SD divided by mean). Distributions are shown as violin plots and medians are shown as bars and numbers. For 349, 336, 474, 165, 201, and 146 genes for CEL-seq2/C1, Drop-seq, MARS-seq, SCRIB-seq, Smart-seq/C1, and Smart-seq2, respectively, no extra Poisson variability could be calculated. See also Figures S6 and S7.

test,  $p < 2.2e-16$ ), but all methods had a fairly high  $R^2$  ranging between 0.83 (MARS-seq) and 0.91 (Smart-seq2). This suggests that, for all methods, transcript concentrations across this broad range can be predicted fairly well from expression values. As expected, accuracy was worse for narrower and especially for lower concentration ranges (Figure S5C). It is worth emphasizing that the accuracy assessed here refers to absolute expression levels across genes within cells. This accuracy can be important, for example, to identify marker genes with a high absolute mRNA expression level. However, the small differences in accuracy seen here will rarely be a decisive factor when choosing among the six protocols.

#### Precision of Amplified Genes Is Strongly Increased by UMIs

While a high accuracy is necessary to compare absolute expression levels, one of the most common experimental aims is to compare relative expression levels to identify differentially expressed genes or different cell types. Hence, the precision (i.e.,

the reproducibility of the expression-level estimate) is a major factor when choosing a method. As we used the same cell type under the same culture conditions for all methods, the amount of biological variation should be the same in the cells analyzed by each of the six methods. Hence, we can assume that differences in the total variation among methods are due to differences in their technical variation. Technical variation is substantial in scRNA-seq data primarily because a substantial fraction of mRNAs is lost during cDNA generation and small amounts of cDNA get amplified. Therefore, both the dropout probability and the amplification noise need to be considered when quantifying variation.

Indeed, a mixture model including a dropout probability and a negative binomial distribution, modeling the overdispersion in the count data, have been shown to represent scRNA-seq data better than the negative binomial alone (Finak et al., 2015; Kharchenko et al., 2014).

To compare precision without penalizing more sensitive methods, we selected a common set of 13,361 genes that were detected in 25% of the cells by at least one method (Figure S6A). We then analyzed these genes in a subsample of 65 cells per method to avoid a bias due to unequal numbers of cells. We estimated the dropout probability as the fraction of cells with zero counts (Figure 5A; Figure S6B). As expected from the number of detected genes per cell (Figure 3C), MARS-seq had the highest median dropout probability (74%) and Smart-seq2 had the lowest (26%) (Figure 5A). To estimate the amplification noise of detected genes, we calculated the coefficient of variation (CV, SD divided by the mean, including zeros), and we subtracted the expected amount of variation due to Poisson sampling (i.e., the square root of the mean divided by the mean). This was possible

for 96.5% (MARS-seq) to 98.9% (Smart-seq2) of all the 13,361 genes. This extra Poisson variability includes biological variation (assumed to be the same across methods in our data) and technical variation, and the latter includes noise introduced by amplification (Brennecke et al., 2013; Grün et al., 2014; Stegle et al., 2015). That amplification noise can be a major factor is seen by the strong increase of extra Poisson variability when ignoring UMIs and considering read counts only (Figure 5B, left; Figure S7A). This is expected, as UMIs should remove amplification noise, which has been described previously for CEL-seq (Grün et al., 2014). For SCRIB-seq and Drop-seq, which are PCR-based methods, UMIs removed even more extra Poisson variability than for CEL-seq2/C1 and MARS-seq (Figure 5B), which is in line with the notion that amplification by PCR is more noisy than amplification by *in vitro* transcription. Of note, Smart-seq2 had the lowest amplification noise when just considering reads (Figure 5B, left), potentially because its higher sensitivity requires less amplification and, hence, leads to less noise.

In summary, Smart-seq2 detects the common set of 13,361 genes in more cells than the UMI methods, but it has, as expected, more amplification noise than the UMI-based methods. How the different combinations of dropout rate and amplification noise affect the power of the methods is not evident, neither from this analysis nor from the total coefficient of variation that ignores the strong mean variance and mean dropout dependencies of scRNA-seq data (Figure S7B).

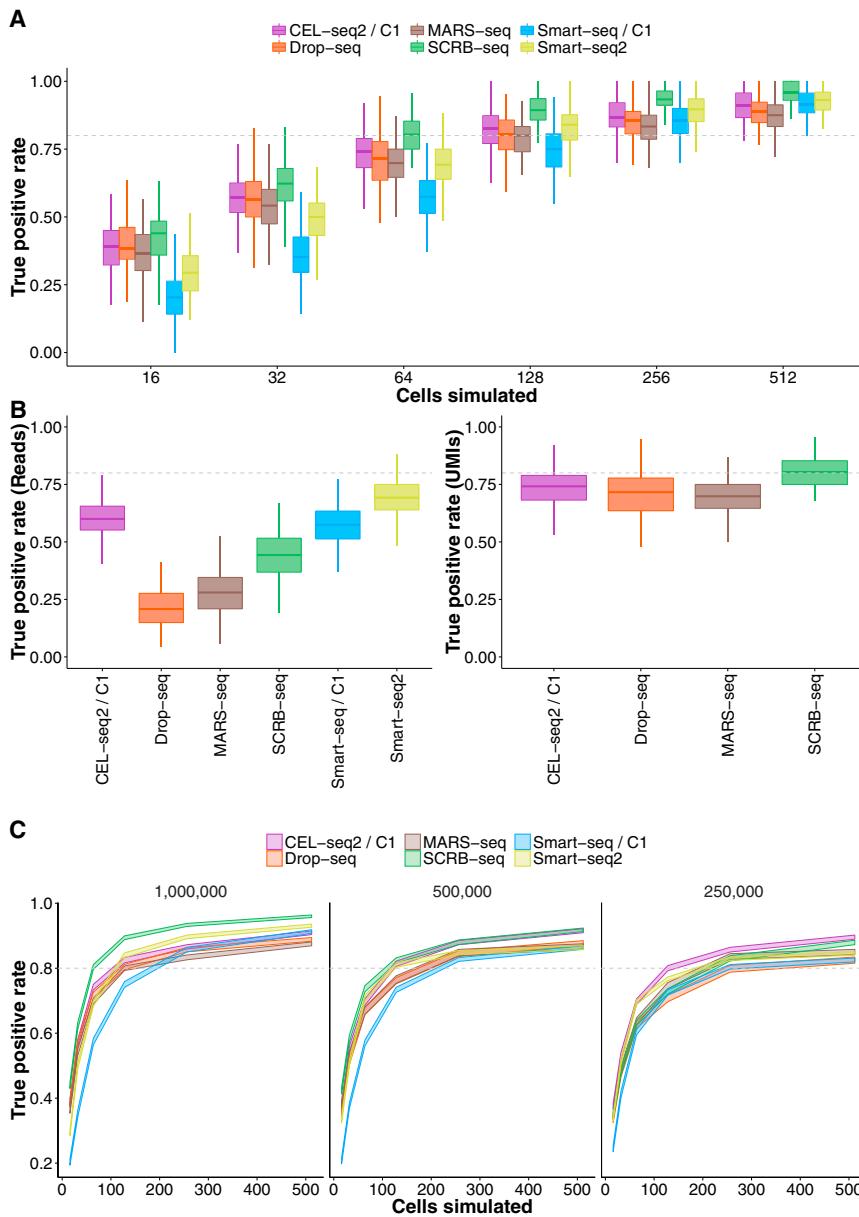
### Power Is Determined by a Combination of Dropout Rates and Amplification Noise and Is Highest for SCRIB-Seq

To estimate the combined impact of sensitivity and precision on the power to detect differential gene expression, we simulated scRNA-seq data given the observed dropout rates and variance for the 13,361 genes. As these depend strongly on the expression level of a gene, it is important to retain the mean variance and mean dropout relationships. To this end, we estimated the mean, the variance (i.e., the dispersion parameter of the negative binomial distribution), and the dropout rate for each gene and method. We then fitted a cubic smoothing spline to the resulting pairs of mean and dispersion estimates to predict the dispersion of a gene given its mean (Figure S8A). Furthermore, we applied a local polynomial regression model to account for the dropout probability given a gene's mean expression (Figure S8B). When simulating data according to these fits, we recovered distributions of dropout rates and variance closely matching the observed data (Figures S8C and S8D). To compare the power for differential gene expression among the methods, we simulated read counts for two groups of  $n$  cells and added log-fold changes to 5% of the 13,361 genes in one group. To mimic a biologically realistic scenario, these log-fold changes were drawn from observed differences between microglial subpopulations from a previously published dataset (Zeisel et al., 2015). Simulated datasets were tested for differential expression using limma (Ritchie et al., 2015), and the true positive rate (TPR) and the false discovery rate (FDR) were calculated. Of note, this does include undetected genes, i.e., the 2.5% (SCRIB-seq) to 6.8% (MARS-seq) of the 13,361 genes that had fewer than two measurements in a particular method (Figure S6B) and for which we could not estimate the variance. In our simulations, these

genes could be drawn as differentially expressed, and in our TPR they were then counted as false negatives for the particular method. Hence, our power simulation framework considers the full range of dropout rates and is not biased against more sensitive methods.

First, we analyzed how the number of cells affects TPR and FDR by running 100 simulations each for a range of 16 to 512 cells per group (Figure 6A). FDRs were similar in all methods ranging from 3.9% to 8.7% (Figure S9A). TPRs differed considerably among methods and SCRIB-seq performed best, reaching a median TPR of 80% with 64 cells. CEL-seq2/C1, Drop-seq, MARS-seq, and Smart-seq2 performed slightly worse, reaching 80% power with 86, 99, 110, and 95 cells per group, respectively, while Smart-seq/C1 needed 150 cells to reach 80% power (Figure 6A). When disregarding UMIs, Smart-seq2 performed best (Figure 6B), as expected from its low dropout rate and its low amplification noise when considering reads only (Figure 5B). Furthermore, power dropped especially for Drop-seq and SCRIB-seq (Figure 6B), as expected from the strong increase in amplification noise of these two methods when considering reads only (Figure 5B). When we stratified our analysis (considering UMIs) across five bins of expression levels, the ranking of methods was recapitulated and showed that the lowest expression bin strongly limited the TPR in all methods (Figure S9B). This ranking also was recapitulated when we analyzed a set of 19 genes previously reported to contain cell-cycle variation in the 2i/LIF culture condition (Kolodziejczyk et al., 2015b). The variance of these cell-cycle genes was clearly higher than the variance of 19 pluripotency and housekeeping (ribosomal) genes in all methods. The p value of that difference was lowest for SCRIB-seq, the most powerful method, and highest for Smart-seq/C1, the least powerful method (Figure S10D).

Notably, this power analysis, as well as the sensitivity, accuracy, and precision parameters analyzed above, includes the variation that is generated in the two technical replicates (batches) per method that we performed (Figure 1). These estimates were very similar among our technical replicates, and, hence, our method comparison is valid with respect to batch variations (Figures S10B–S10D). In addition, as batch effects are known to be highly relevant for interpreting scRNA-seq data (Hicks et al., 2015), we gauged the magnitude of batch effects with respect to identifying differentially expressed genes. To this end, we used limma to identify differentially expressed genes between batches (FDR < 1%), using 25 randomly selected cells per batch and method. All methods had significantly more genes differentially expressed between batches than expected from permutations (zero to four genes), with a median of 119 (Drop-seq) to ~1,135 (CEL-seq2/C1) differentially expressed genes (Figure S10A). Notably, genes were affected at random across methods, as there was no significant overlap among them (extended hypergeometric test [Kalinka, 2013],  $p > 0.84$ ). Hence, this analysis once more emphasizes that batches are important to consider in the design of scRNA-seq experiments (Hicks et al., 2015). While a quantitative comparison of the magnitude of batch effects among methods would require substantially more technical replicates per method, the methods differ in their flexibility to incorporate batch effect into the experimental design, which is an important aspect to consider as discussed below.



As a next step, we analyzed how the performance of the six methods depends on sequencing depth. To this end, we performed power simulations as above, but we estimated the mean dispersion and mean dropout relationships from data downsampled to 500,000 or 250,000 reads per cell. Overall, the decrease in power was moderate (Figure 6C; Table 1) and followed the drop in sensitivity at different sequencing depths (Figure 3B). While Smart-seq2 and CEL-seq2/C1 needed just 1.3-fold more cells at 0.25 million reads than at one million reads to reach 80% power, SCRB-seq and Drop-seq required 2.6-fold more cells (Table 1). In summary, SCRB-seq is the most powerful method at one million reads and half a million reads, but CEL-seq2/C1 is the most powerful method at a sequencing depth of 250,000 reads. The optimal balance between the number of cells and their sequencing depth depends on many factors,

### Figure 6. Power of scRNA-Seq Methods

Using the empirical mean/dispersion and mean/dropout relationships (Figures S8A and S8B), we simulated data for two groups of  $n$  cells each for which 5% of the 13,361 genes were differentially expressed, with log-fold changes drawn from observed differences between microglial subpopulations from a previously published dataset (Zeisel et al., 2015). The simulated data were then tested for differential expression using limma (Ritchie et al., 2015), from which the average true positive rate (TPR) and the average false discovery rate (FDR) were calculated (Figure S9A). (A) TPR for one million reads per cell for sample sizes  $n = 16$ ,  $n = 32$ ,  $n = 64$ ,  $n = 128$ ,  $n = 256$ , and  $n = 512$  per group. Boxplots represent the median and first and third quartiles of 100 simulations. (B) TPR for one million reads per cell for  $n = 64$  per group with and without using UMI information. Boxplots represent the median and first and third quartiles of 100 simulations. (C) TPRs as in (A) using mean/dispersion and mean/dropout estimates from one million (as in A), 0.5 million, and 0.25 million reads. Line areas indicate the median power with SE from 100 simulations. See also Figures S8–S10 and Table 1.

including the scientific questions addressed, the experimental design, or the sample availability. However, the monetary cost is certainly an important one, and we used the results of our simulations to compare the costs among the methods for a given level of power.

### Cost Efficiency Is Similarly High for Drop-Seq, MARS-Seq, SCRB-Seq, and Smart-Seq2

Given the number of cells needed to reach 80% power as simulated above for three sequencing depths (Figure 6C), we calculated the minimal costs to generate and sequence these libraries.

For example, at a sequencing depth of one million reads, SCRB-seq requires 64 cells per group to reach 80% power. Generating 128 SCRB-seq libraries costs ~260\$ and generating 128 million reads costs ~640\$. Note that the necessary paired-end reads for CEL-seq2/C1, SCRB-seq, MARS-seq, and Drop-seq can be generated using a 50-cycle sequencing kit, and, hence, we assume that sequencing costs are the same for all methods.

Calculating minimal costs this way, Drop-seq (690\$) is the most cost-effective method when sequencing 254 cells at a depth of 250,000 reads, and SCRB-seq (810\$), MARS-seq (820\$), and Smart-seq2 (1,090\$) are slightly more expensive at the same performance (Table 1). For Smart-seq2 it should be stressed that the use of in-house-produced Tn5 transposase (Picelli et al., 2014a) is required to keep the cost at this level, as

**Table 1. Cost Efficiency Extrapolation for Single-Cell RNA-Seq Experiments**

Method	TPR <sup>a</sup>	FDR <sup>a</sup> (%)	Cell per Group <sup>b</sup>	Library Cost (\$)	Minimal Cost <sup>c</sup> (\$)
CEL-seq2/C1	0.8	~6.1	86/100/110	~9	~2,420/2,310/2,250
Drop-seq	0.8	~8.4	99/135/254	~0.1	~1,010/700/690
MARS-seq	0.8	~7.3	110/135/160	~1.3	~1,380/1,030/820
SCRB-seq	0.8	~6.1	64/90/166	~2	~900/810/1,080
Smart-seq/C1	0.8	~4.9	150/172/215	~25	~9,010/9,440/11,290
Smart-seq2 (commercial)	0.8	~5.2	95/105/128	~30	~10,470/11,040/13,160
Smart-seq2 (in-house Tn5)	0.8	~5.2	95/105/128	~3	~1,520/1,160/1,090

See also [Figure 6](#).

<sup>a</sup>True positive rate and false discovery rate are based on simulations ([Figure 6](#); [Figure S9](#)).

<sup>b</sup>Sequencing depth of one, 0.5, and 0.25 million reads.

<sup>c</sup>Assuming \$5 per one million reads.

was done in our experiments. When instead using the Tn5 transposase of the commercial Nextera kit as described (Picelli et al., 2014b), the costs for Smart-seq2 are 10-fold higher. Even if one reduces the amount of Nextera transposase to a quarter, as done in the Smart-seq/C1 protocol, the Smart-seq2 protocol is still four times more expensive than the early barcoding methods. CEL-seq2/C1 is fairly expensive due to the microfluidic chips that make up 69% of the library costs, and Smart-seq/C1 is almost 13-fold less efficient than Drop-seq due to its high library costs that arise from the microfluidic chips, the commercial Smart-seq kit, and the costs for commercial Nextera XT kits.

Of note, these calculations are the minimal costs of the experiment and several factors are not considered, such as labor costs, costs to set up the methods, costs to isolate cells of interest, or costs due to practical constraints in generating a fixed number of scRNA-seq libraries with a fixed number of reads. In many experimental settings, independent biological and/or technical replicates are needed when investigating particular factors, such as genotypes or developmental time points, and Smart-seq/C1, CEL-seq2/C1, and Drop-seq are less flexible in distributing scRNA-seq libraries across replicates than the other three methods that use PCR plates. Furthermore, the costs are increased by unequal sampling from the included cells as well as from sequencing reads from cells that are excluded. In our case, between 6% (SCRB-seq) and 32% (Drop-seq) of the reads came from cell barcodes that were not included. While it is difficult to exactly calculate and compare these costs among methods, it is clear that they will increase the costs for Drop-seq relatively more than for the other methods. In summary, we find that Drop-seq, SCRB-seq, and MARS-seq are the most cost-effective methods, closely followed by Smart-seq2, if using an in-house-produced transposase.

## DISCUSSION

Here we have provided an in-depth comparison of six prominent scRNA-seq protocols. To this end, we generated data for all six compared methods from the same cells, cultured under the same condition in the same laboratory. While there would be many more datasets and methods for a comparison of the sensitivity and accuracy of the ERCCs (Svensson et al., 2016), our approach provides a more controlled and comprehensive com-

parison across thousands of endogenous genes. This is important, as can be seen by the different sensitivity estimates that we obtained for Drop-seq, MARS-seq, and SCRB-seq using the ERCCs. In our comparison, we clearly find that Smart-seq2 is the most sensitive method, closely followed by SCRB-seq, Smart-seq/C1, and CEL-seq2/C1, while Drop-seq and MARS-seq detect nearly 50% fewer genes per cell ([Figures 3B](#) and [3C](#)). In addition, Smart-seq2 shows the most even read coverage across transcripts ([Figure S3D](#)), making it the most appropriate method for the detection of alternative splice forms and for analyses of allele-specific expression using SNPs (Deng et al., 2014; Reinius et al., 2016). Hence, Smart-seq2 is certainly the most suitable method when an annotation of single-cell transcriptomes is the focus. Furthermore, we find that Smart-seq2 is also the most accurate method (i.e., it has the highest correlation of known ERCC spike-in concentrations and read counts per million), which is probably related to its higher sensitivity. Hence, differences in expression values across transcripts within the same cell predict differences in the actual concentrations of these transcripts well. All methods do this rather well, at least for higher expression levels, and we think that the small differences among methods will rarely be a decisive factor. Importantly, the accuracy of estimating transcript concentrations across cells (relevant, e.g., for comparing the total RNA content of cells) depends on different factors and cannot be compared well among the tested methods as it would require known concentration differences of transcripts across cells. However, it is likely that methods that can use UMIs and ERCCs (CEL-seq2/C1, MARS-seq, and SCRB-seq) would have a strong advantage in this respect.

How well relative expression levels of the same genes can be compared across cells depends on two factors. First, how often (i.e., in how many cells and from how many molecules) it is measured. Second, with how much technical variation (i.e., with how much noise, e.g., from amplification) it is measured. For the first factor (dropout probability), we find Smart-seq2 to be the best method ([Figure 5A](#)), as expected from its high gene detection sensitivity. For the second factor (extra Poisson variability), we find the four UMI methods to perform better ([Figure 5B](#)), as expected from their ability to eliminate variation introduced by amplification. To assess the combined effect of these two factors, we performed simulations for differential gene

expression scenarios (Figure 6). This allowed us to translate the sensitivity and precision parameters into the practically relevant power to detect differentially expressed genes. Of note, our power estimates include the variation that is caused by the two different replicates per method that constitutes an important part of the variation. Our simulations show that, at a sequencing depth of one million reads, SCRB-seq has the highest power, probably due to a good balance of high sensitivity and low amplification noise. Furthermore, amplification noise and power strongly depend on the use of UMIs, especially for the PCR-based methods (Figures 5B and 6B; Figure S7). Notably, this is due to the large amount of amplification needed for scRNA-seq libraries, as the effect of UMIs on power for bulk RNA-seq libraries is negligible (Parekh et al., 2016).

Perhaps practically most important, our power simulations also allow us to compare the efficiency of the methods by calculating the costs to generate the data for a given level of power. Using minimal cost calculations, we find that Drop-seq is the most cost-effective method, closely followed by SCRB-seq, MARS-seq, and Smart-seq2. However, Drop-seq costs are likely to be more underestimated, due to lower flexibility in generating a specified number of libraries and the higher fraction of reads that come from bad cells. Hence, all four UMI methods are in practice probably similarly cost-effective. In contrast, for Smart-seq2 to be similarly cost-effective it is absolutely necessary to use in-house-produced transposase or to drastically reduce volumes of commercial transposase kits (Lamble et al., 2013; Mora-Castilla et al., 2016).

Given comparable efficiencies of Drop-seq, MARS-seq, SCRB-seq, and Smart-seq2, additional factors will play a role when choosing a suitable method for a particular question. Due to its low library costs, Drop-seq is probably preferable when analyzing large numbers of cells at low coverage (e.g., to find rare cell types). On the other hand, Drop-seq in its current setup requires a relatively large amount of cells (>6,500 for 1 min of flow). Hence, if few and/or unstable cells are isolated by FACS, the SCRB-seq, MARS-seq, or Smart-seq2 protocols are probably preferable. Additional advantages of these methods over Drop-seq include that technical variation can be estimated from ERCCs for each cell, which can be helpful to estimate biological variation (Kim et al., 2015; Vallejos et al., 2016), and that the exact same setup can be used to generate bulk RNA-seq libraries. While SCRB-seq is slightly more cost-effective than MARS-seq and has the advantage that one does not need to produce the transposase in-house, Smart-seq2 is preferable when transcriptome annotation, identification of sequence variants, or the quantification of different splice forms is of interest. Furthermore, the presence of batch effects shows that experiments need to be designed in a way that does not confound batches with biological factors (Hicks et al., 2015). Practically, plate-based methods might currently accommodate complex experimental designs with various biological factors more easily than microfluidic chips.

We find that Drop-seq, MARS-seq, SCRB-seq, and Smart-seq2 (using in-house transposase) are 2- to 13-fold more cost efficient than CEL-seq2/C1, Smart-seq/C1, and Smart-seq2 (using commercial transposase). Hence, the latter methods

would need to increase in their power and/or decrease in their costs to be competitive. The efficiency of the Fluidigm C1 platform can be further increased by microfluidic chips with a higher throughput, as available in the high-throughput (HT) mRNA-seq integrated fluidic circuit (IFC) chip. While CEL-seq2/C1 has been found to be more sensitive than the plate-based version of CEL-seq2 (Hashimshony et al., 2016), the latter might be more efficient when considering its lower costs. Our finding that Smart-seq2 is the most sensitive protocol also hints toward further possible improvements of SCRB-seq and Drop-seq. As these methods also rely on template switching and PCR amplification, the improvements found in the systematic optimization of Smart-seq2 (Picelli et al., 2013) also could improve the sensitivity of SCRB-seq and Drop-seq. Furthermore, the costs of SCRB-seq libraries per cell can be halved when switching to a 384-well format (Soumillon et al., 2014). Similarly, improvements made for CEL-seq2 (Hashimshony et al., 2016) could be incorporated into the MARS-seq protocol. Hence, it is clear that scRNA-seq protocols will become even more efficient in the future. The results of our comparative analyses of six currently prominent scRNA-seq methods may facilitate such developments, and they provide a framework for method evaluation in the future.

In summary, we systematically compared six prominent scRNA-seq methods and found that Drop-seq is preferable when quantifying transcriptomes of large numbers of cells with low sequencing depth, SCRB-seq and MARS-seq is preferable when quantifying transcriptomes of fewer cells, and Smart-seq2 is preferable when annotating and/or quantifying transcriptomes of fewer cells as long one can use in-house-produced transposase. Our analysis allows an informed choice among the tested methods, and it provides a framework for benchmarking future improvements in scRNA-seq methodologies.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Published data
  - Single cell RNA-seq library preparations
  - DNA sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Basic data processing and sequence alignment
  - Power Simulations
  - ERCC capture efficiency
  - Cost efficiency calculation
- DATA AND SOFTWARE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information includes ten figures and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2017.01.023>.

**AUTHOR CONTRIBUTIONS**

C.Z. and W.E. conceived the experiments. C.Z. prepared scRNA-seq libraries and analyzed the data. B.V. implemented the power simulation framework and estimated the ERCC capture efficiencies. S.P. helped in data processing and power simulations. B.R. prepared the Smart-seq2 scRNA-seq libraries. A.G.-A. and H.H. established and performed the MARS-seq library preps. M.S. performed the cell culture of mESCs. W.E. and H.L. supervised the experimental work and I.H. provided guidance in data analysis. C.Z., I.H., B.R., and W.E. wrote the manuscript. All authors read and approved the final manuscript.

**ACKNOWLEDGMENTS**

We thank Rickard Sandberg for facilitating the Smart-seq2 sequencing. We thank Christopher Mulholland for assistance with FACS, Dominik Alterauge for help establishing the Drop-seq method, and Stefan Krebs and Helmut Blum from the LAFUGA platform for sequencing. We are grateful to Magali Soumillon and Tarjei Mikkelsen for providing the SCRIB-seq protocol. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through LMUexcellent and the SFB1243 (Subproject A01/A14/A15) as well as a travel grant to C.Z. by the Boehringer Ingelheim Fonds.

Received: August 8, 2016

Revised: December 1, 2016

Accepted: January 17, 2017

Published: February 9, 2017

**REFERENCES**

- Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., and Heisler, M.G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095.
- Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196.
- 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* **29**, 15–21.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
- Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278.
- Frazee, A.C., Jaffe, A.E., Langmead, B., and Leek, J.T. (2015). Polyester: simulating RNA-seq datasets with differential transcript expression. *Bioinformatics* **31**, 2778–2784.
- Gokce, O., Stanley, G.M., Treutlein, B., Neff, N.F., Camp, J.G., Malenka, R.C., Rothwell, P.E., Fuccillo, M.V., Südhof, T.C., and Quake, S.R. (2016). Cellular taxonomy of the mouse striatum as revealed by single-cell RNA-seq. *Cell Rep.* **16**, 1126–1137.
- Grün, D., and van Oudenaarden, A. (2015). Design and analysis of single-cell sequencing experiments. *Cell* **163**, 799–810.
- Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat. Methods* **11**, 637–640.
- 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.
- Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77.
- Hicks, S.C., Teng, M., and Irizarry, R.A. (2015). On the widespread and critical impact of systematic bias and batch effects in single-cell RNA-Seq data. *bioRxiv*. <http://dx.doi.org/10.1101/025528>.
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166.
- Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779.
- Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551.
- Kalinka, A.T. (2013). The probability of drawing intersections: extending the hypergeometric distribution. *arXiv*. [arXiv:1305.0717](https://arxiv.org/abs/1305.0717). <https://arxiv.org/abs/1305.0717>.
- Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11**, 740–742.
- Kim, J.K., Kołodziejczyk, A.A., Ilicic, T., Teichmann, S.A., and Marioni, J.C. (2015). Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat. Commun.* **6**, 8687.
- Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2011). Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74.
- 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.
- Kołodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015a). The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620.
- Kołodziejczyk, A.A., Kim, J.K., Tsang, J.C.H., Ilicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Bühl, M., Liu, P., et al. (2015b). Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* **17**, 471–485.
- Lamble, S., Batty, E., Attar, M., Buck, D., Bowden, R., Lunter, G., Crook, D., El-Fahmawi, B., and Piazza, P. (2013). Improved workflows for high throughput library preparation using the transposome-based Nextera system. *BMC Biotechnol.* **13**, 104.
- Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29.
- Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926.
- Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108.
- Lun, A.T.L., Bach, K., and Marioni, J.C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75.
- 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.
- Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M., and Wold, B.J. (2014). From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* **24**, 496–510.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12.
- Mora-Castilla, S., To, C., Vaezeslami, S., Morey, R., Srinivasan, S., Dumdie, J.N., Cook-Andersen, H., Jenkins, J., and Laurent, L.C. (2016). Miniaturization technologies for efficient single-cell library preparation for next-generation sequencing. *J. Lab. Autom.* **21**, 557–567.

- Nam, D.K., Lee, S., Zhou, G., Cao, X., Wang, C., Clark, T., Chen, J., Rowley, J.D., and Wang, S.M. (2002). Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc. Natl. Acad. Sci. USA* 99, 6152–6156.
- Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2016). The impact of amplification on differential expression analyses by RNA-seq. *Sci. Rep.* 6, 25533.
- Petropoulos, S., Edsgård, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* 165, 1012–1026.
- Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098.
- Picelli, S., Björklund, Å.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014a). Tn5 transposase and fragmentation procedures for massively scaled sequencing projects. *Genome Res.* 24, 2033–2040.
- Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014b). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.
- Reinius, B., Mold, J.E., Ramsköld, D., Deng, Q., Johnsson, P., Michaëlsson, J., Frisén, J., and Sandberg, R. (2016). Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat. Genet.* 48, 1430–1435.
- Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., and Kelso, J. (2015). deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* 31, 770–772.
- Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* 32, 896–902.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
- Soumillon, M., Caccharelli, D., Semrau, S., van Oudenaarden, A., and Mikkelsen, T.S. (2014). Characterization of directed differentiation by high-throughput single-cell RNA-seq. *bioRxiv*. <http://dx.doi.org/10.1101/003236>.
- Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16, 133–145.
- Svensson, V., Natarajan, K.N., Ly, L.-H., Miragaia, R.J., Labalette, C., Macaulay, I.C., Cvejic, A., and Teichmann, S.A. (2016). Power analysis of single cell RNA-sequencing experiments. *bioRxiv*. <http://dx.doi.org/10.1101/073692>.
- Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196.
- Vallejos, C.A., Richardson, S., and Marioni, J.C. (2016). Beyond comparisons of means: understanding changes in gene expression at the single-cell level. *Genome Biol.* 17, 70.
- Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* 34, 1145–1160.
- Wu, A.R., Neff, N.F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M.E., Mburu, F.M., Mantalas, G.L., Sim, S., Clarke, M.F., and Quake, S.R. (2014). Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* 11, 41–46.
- Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnberg, 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.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Esgro recombinant mouse LIF	Millipore	ESG1107
CHIR99021	Axon Med Chem	1386
PD0325901	Axon Med Chem	1408
2-Mercaptoethanol	Sigma-Aldrich	M3148
FBS	Sigma-Aldrich	F7524
Penicillin/Streptomycin	Sigma-Aldrich	P4333
MEM non-essential amino acids	Sigma-Aldrich	M7145
L-glutamine	Sigma-Aldrich	G7513
Dulbecco's modified Eagle's medium	Sigma-Aldrich	D6429
Perfluorooctanol	Sigma-Aldrich	370533
Maxima H- Reverse Transcriptase	Thermo Fisher Scientific	EP0753
SuperScript II	Life Technologies	18064071
Exonuclease I	New England Biolabs	M0293L
RNAprotect Cell Reagent	QIAGEN	76526
RNase inhibitor	Promega	N2515
RNase inhibitor	Lucigen	30281-2-LU
Phusion HF buffer	New England Biolabs	B0518S
Proteinase K	Ambion	AM2546
KAPA HiFi HotStart polymerase	KAPA Biosystems	KAPBKK2602
Phusion HF PCR Master Mix	Thermo Fisher Scientific	F531L
dNTPs	New England Biolabs	N0447L
Triton X-100	Sigma-Aldrich	T8787
SDS	Sigma-Aldrich	L3771
Tn5 transposase	<a href="#">Picelli et al., 2014a</a>	N/A
<b>Critical Commercial Assays</b>		
C1 Single-Cell System	Fluidigm	N/A
C1 IFC for Open App (10-17 µm)	Fluidigm	100-8134
C1 IFC for mRNA-seq (10-17 µm)	Fluidigm	100-6041
Nextera XT DNA Sample Preparation Kit	Illumina	FC-131-1096
SMARTer Ultra Low RNA Kit for Fluidigm C1	Clontech	634833
MinElute Gel Extraction Kit	QIAGEN	28606
<b>Deposited Data</b>		
single-cell RNA-seq data	This paper	GEO: GSE75790
Drop-seq ERCC data	<a href="#">Macosko et al., 2015</a>	GEO: GSE66694
<b>Experimental Models: Cell Lines</b>		
J1 mouse embryonic stem cells	<a href="#">Li et al., 1992</a>	N/A
<b>Sequence-Based Reagents</b>		
Nextera XT Index Kit	Illumina	FC-121-1012
SCRB-seq P5 primer, AATGATAACGGCGACCACCG AGATCTACACTTTCCCTACACGACGCTCTC CG*A*T*C*T, * PTO bond	IDT	N/A
SCRB-seq oligo-dT primer, Biotin-ACACTTTCCCT ACACGACGCTCTCCGATCT[BC6][N10][T30]VN	IDT	"TruGrade Ultramer"

(Continued on next page)

***Continued***

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SCRB-seq template-switch oligo, iCIGiCACACTCTTCC CTACACGGACGCrGrGrG	Eurogentech	N/A
Drop-seq P5 primer, AATGATAACGGCGACCACCGAGA TCTACACGCCT GTCCCGCGGAAGCAGTGGTATCAACG CAGAGT*A*C, * PTO bond	IDT	N/A
Drop-seq oligo-dT primer beads, Bead-Linker- TTTTTTAAGCAGTGGTATCAAC GCAGAGTAC[BC12][N8][T30]	Chemgenes	MACOSKO-2011-10
Drop-seq template-switch oligo, AAGCAGTGGTATCA ACGCAGAGTGAATrGrGrG	IDT	N/A
CEL-seq2 oligo-dT primer, GCCGGTAATACGACTCACTATA GGGAGTTCTACAGTCAGCAGTC[N6][BC6][T25]	Sigma-Aldrich	N/A
ERCC RNA Spike-In Mix	Ambion	4456740
<b>Software and Algorithms</b>		
STAR	Dobin et al., 2013	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
Drop-seq tools	Macosko et al., 2015	<a href="http://mccarrolllab.com/dropseq/">http://mccarrolllab.com/dropseq/</a>
featureCounts	Liao et al., 2013	<a href="https://bioconductor.org/packages/release/bioc/html/Rsubread.html">https://bioconductor.org/packages/release/bioc/html/Rsubread.html</a>
R	N/A	<a href="http://www.r-project.org">www.r-project.org</a>
<b>Other</b>		
Drop-seq PDMS device	Nanoshift	Drop-seq
2% E-Gel Agarose EX Gels	Life Technologies	G402002

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding author Wolfgang Enard ([enard@biologie.uni-muenchen.de](mailto:enard@biologie.uni-muenchen.de)).

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

J1 mouse embryonic stem cells (Li et al., 1992) 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  $\beta$ -mercaptoethanol (Sigma-Aldrich), 2 mM L-glutamine, 1x MEM non-essential amino acids, 100 U/ml penicillin, 100  $\mu$ g/ml streptomycin (Sigma-Aldrich), 1000 U/ml recombinant mouse LIF (Millipore) and 2i (1  $\mu$ M PD032591 and 3  $\mu$ M 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.

**METHOD DETAILS****Published data**

Drop-seq ERCC (Macosko et al., 2015) data were obtained under accession GEO: 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.

**Single cell RNA-seq library preparations****CEL-seq2/C1**

CEL-seq2/C1 libraries were generated as previously described (Hashimshony et al., 2016). Briefly, cells (200,000/ml), ERCC spike-ins, reagents and barcoded oligo-dT primers (Sigma-Aldrich) were loaded on a 10-17  $\mu$ m C1 Open-App microfluidic IFC (Fluidigm). Cell lysis, reverse transcription, second strand synthesis and in-vitro transcription were performed on-chip. Subsequently, harvested aRNA was pooled from 48 capture sites. After fragmentation and clean-up, 5  $\mu$ l of aRNA was used to construct final libraries by reverse transcription (SuperScript II, Thermo Fisher) and library PCR (Phusion HF, Thermo Fisher).

**Drop-seq**

Drop-seq experiments were performed as published (Macosko et al., 2015) and successful establishment of the method in our lab was confirmed by a species-mixing experiment (Figure S1A). For this work, J1 mES cells (100/ $\mu$ l) and barcode-beads (120/ $\mu$ l, Chem-Genes) were co-flowed in Drop-seq PDMS devices (Nanoshift) at rates of 4000  $\mu$ l/hr. Collected emulsions were broken by addition of perfluorooctanol (Sigma-Aldrich) and mRNA on beads was reverse transcribed (Maxima RT, Thermo Fisher). 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).

**MARS-seq**

To construct single cell libraries from polyA-tailed RNA, we applied massively parallel single-cell RNA sequencing (MARS-Seq) (Jaitin et al., 2014). Briefly, single cells were FACS-sorted into 384-well plates, containing lysis buffer and reverse-transcription (RT) primers. The RT primers contained the single cell barcodes and unique molecular identifiers (UMIs) for subsequent de-multiplexing and correction for amplification biases, respectively. Spike-in transcripts (ERCC, Ambion) were added, polyA-containing RNA was converted into cDNA as previously described and then pooled using an automated pipeline (liquid handling robotics). Subsequently, samples were linearly amplified by in vitro transcription, fragmented, and 3' ends were converted into sequencing libraries. The libraries consisted of 48 single cell pools.

**SCRB-seq**

RNA was stabilized by resuspending cells in RNAsprotect Cell Reagent (QIAGEN) and RNase inhibitors (Promega). Prior to FACS sorting, cells were diluted in PBS (Invitrogen). Single cells were sorted into 5  $\mu$ l lysis buffer consisting of a 1/500 dilution of Phusion HF buffer (New England Biolabs) and ERCC spike-ins (Ambion), spun down and frozen at -80°C. Plates were thawed and libraries prepared as described previously (Soumillon et al., 2014). Briefly, RNA was desiccated after protein digestion by 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 using the SMARTer Ultra Low RNA Kit (Clontech) according to the manufacturer's protocol. Cells were loaded on a 10-17  $\mu$ m RNA-seq microfluidic IFC at a concentration of 200,000/ml. Capture site occupancy was surveyed using the Operetta (Perkin Elmer) automated imaging platform.

**Smart-seq2**

mESCs were sorted into 96-well PCR plates containing 2  $\mu$ l lysis buffer (1.9  $\mu$ l 0.2% Triton X-100; 0.1  $\mu$ l RNase inhibitor (Lucigen)) and spike-in RNAs (Ambion), spun down and frozen at -80°C. To generate Smart-seq2 libraries, priming buffer mix containing dNTPs and oligo-dT primers was added to the cell lysate and denatured at 72°C. cDNA synthesis and pre-amplification of cDNA was performed as described previously (Picelli et al., 2014b, 2013). Sequencing libraries were constructed from 2.5 ng of pre-amplified cDNA using an in-house generated Tn5 transposase (Picelli et al., 2014a). Briefly, 5  $\mu$ l cDNA was incubated with 15  $\mu$ l tagmentation mix (1  $\mu$ l of Tn5; 2  $\mu$ l 10x TAPS MgCl<sub>2</sub> Tagmentation buffer; 5  $\mu$ l 40% PEG8000; 7  $\mu$ l water) for 8 min at 55°C. Tn5 was inactivated and released from the DNA by the addition of 5  $\mu$ l 0.2% SDS and 5 min incubation at room temperature. Sequencing library amplification was performed using 5  $\mu$ l Nextera XT Index primers (Illumina) that had been first diluted 1:5 in water and 15  $\mu$ l PCR mix (1  $\mu$ l KAPA HiFi DNA polymerase (KAPA Biosystems); 10  $\mu$ l 5x KAPA HiFi buffer; 1.5  $\mu$ l 10mM dNTPs; 2.5  $\mu$ l water) in 10 PCR cycles. Barcoded libraries were purified and pooled at equimolar ratios.

**DNA sequencing**

For SCRB-seq and Drop-seq, 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.

Smart-seq/C1, CEL-seq2/C1, Drop-seq and SCRB-seq library pools were sequenced on an Illumina HiSeq1500. Smart-seq2 pools were sequenced on Illumina HiSeq2500 (Replicate A) and HiSeq2000 (Replicate B) platforms. MARS-seq library pools were sequenced on an Illumina HiSeq2500 using the Rapid mode. Smart-seq/C1 and Smart-seq2 libraries were sequenced 45 cycles single-end, whereas CEL-seq2/C1, Drop-seq and SCRB-seq libraries were sequenced paired-end with 15-20 cycles to decode cell barcodes and UMI from read 1 and 45 cycles into the cDNA fragment. MARS-seq libraries were paired-end sequenced with 52 cycles on read 1 into the cDNA and 15 bases for read 2 to obtain cell barcodes and UMIs. Similar sequencing qualities were confirmed by FastQC v0.10.1 (Figure S1B).

**QUANTIFICATION AND STATISTICAL ANALYSIS****Basic data processing and sequence alignment**

Smart-seq/C1/Smart-seq2 libraries (i5 and i7) and CELseq2/C1/Drop-seq/SCRB-seq pools (i7) were demultiplexed from the Illumina barcode reads using deML (Renaud et al., 2015). MARS-seq library pools were demultiplexed with the standard Illumina pipeline. All reads were trimmed to the same length of 45 bp by cutadapt (Martin, 2011) (v1.8.3) and 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 ([Dobin et al., 2013](#)) using all default parameters.

For libraries containing UMIs, cell- and gene-wise count/UMI tables were generated using the published Drop-seq pipeline (v1.0) ([Macosko et al., 2015](#)). We discarded the last 2 bases of the Drop-seq cell and molecular barcodes to account for bead synthesis errors. For Smart-seq/C1 and Smart-seq2, features were assigned and counted using the Rsubread package (v1.20.2) ([Liao et al., 2013](#)).

### Power Simulations

We developed a framework in R for statistical power evaluation of differential gene expression in single cells. For each method, we estimated the mean expression, dispersion and dropout probability per gene from the same number of cells per method. In the read count simulations, we followed the framework proposed in Polyester ([Frazee et al., 2015](#)), i.e., we retained the observed mean-variance dependency by applying a cubic smoothing spline fit to capture the heteroscedasticity observed. Furthermore, we included a local polynomial regression fit for the mean-dropout relationship. In each iteration, we simulated count measurements for the 13,361 genes for sample sizes of  $2^4$ ,  $2^5$ ,  $2^6$ ,  $2^7$ ,  $2^8$  and  $2^9$  cells per group. The read count for a gene  $i$  in a cell  $j$  is modeled as a product of a binomial and negative binomial distribution:

$$X_{ij} \sim B(p = 1 - p_0) * NB(\mu, \theta).$$

The mean expression magnitude  $\mu$  was randomly drawn from the empirical distribution. 5 percent of the genes were defined as differentially expressed with an effect size drawn from the observed fold changes between microglial subpopulations in Zeisel et al. ([Zeisel et al., 2015](#)). The dispersion  $\theta$  and dropout probability  $p_0$  were predicted by above mentioned fits.

For each method and sample size, 100 RNA-seq experiments were simulated and tested for differential expression using limma ([Ritchie et al., 2015](#)) in combination with voom ([Law et al., 2014](#)) (v3.26.7). The power simulation framework was implemented in R (v3.3.0).

### 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 ([Marinov et al., 2014](#)). Detections are independent from each other and are thus regarded as independent Bernoulli trials. We recorded the number of cells with nonzero and zero read or UMI counts 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.

### Cost efficiency calculation

We based our cost efficiency extrapolation on the power simulations starting from empirical data at different sequencing depths (250,000 reads, 500,000 reads, 1,000,000 reads; [Figure 6C](#)). We determined the number of cells required per method and depth for adequate power (80%) by an asymptotic fit to the median powers. 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. We used prices in Euro as a basis and consider an exchange course of 1:1 for the given prices in USD.

### DATA AND SOFTWARE AVAILABILITY

The accession number for the raw and analyzed scRNA-seq data reported in this paper is GEO: GSE75790.

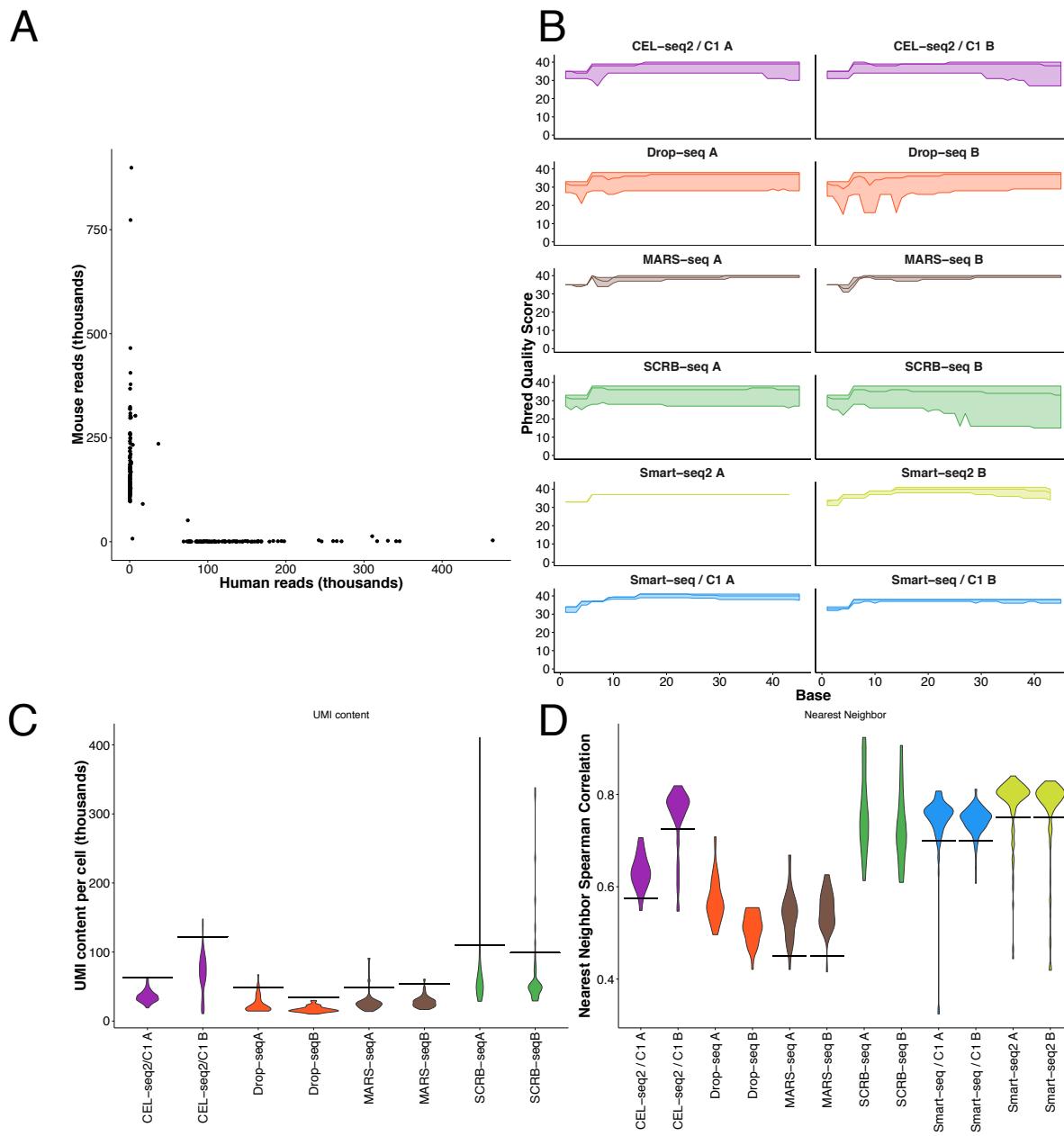
**Supplemental Information**

**Comparative Analysis  
of Single-Cell RNA Sequencing Methods**

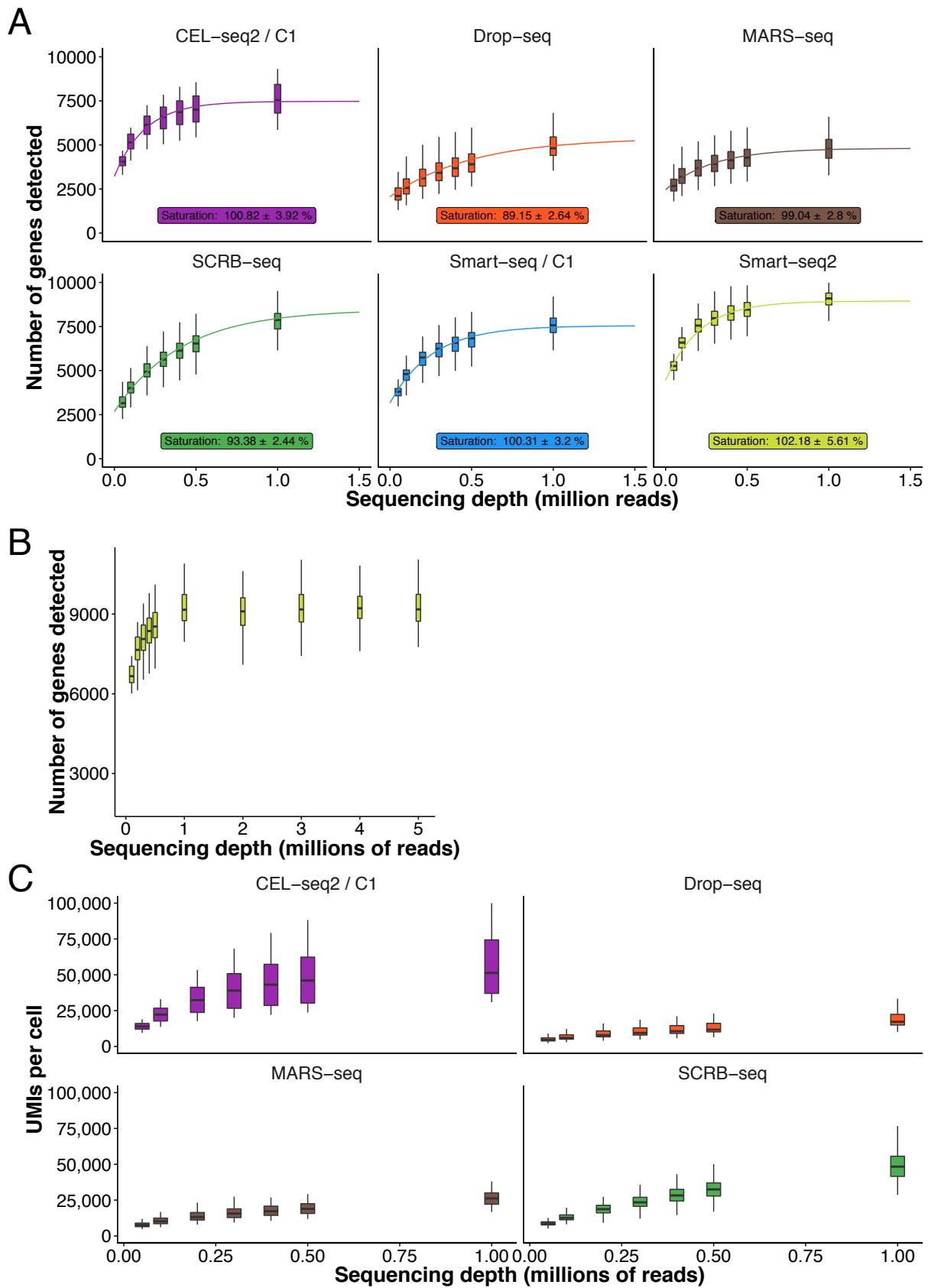
**Christoph Ziegenhain, Beate Vieth, Swati Parekh, Björn Reinius, Amy Guillaumet-Adkins, Martha Smets, Heinrich Leonhardt, Holger Heyn, Ines Hellmann, and Wolfgang Enard**

## Supplementary Figures

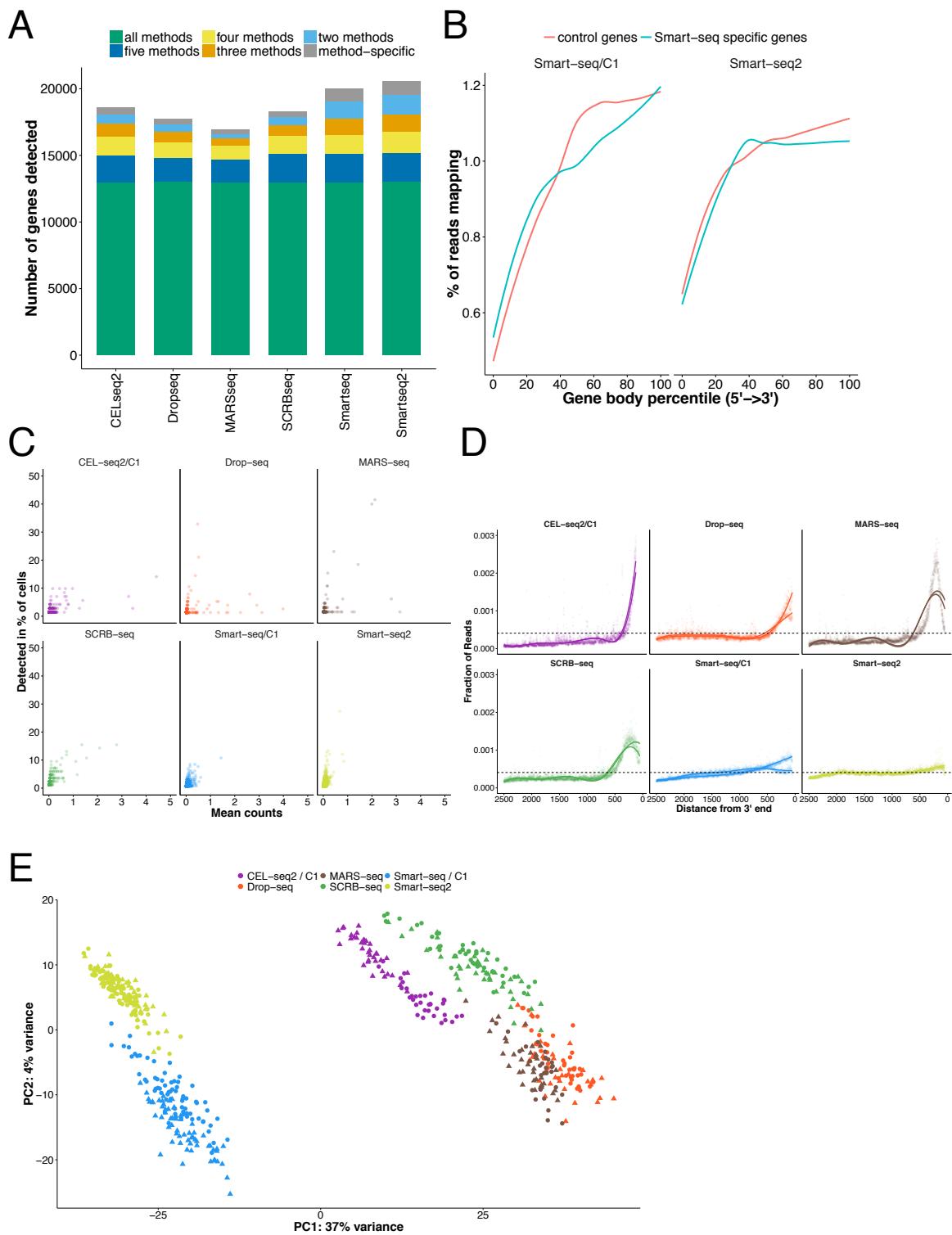
**Figure S1**



**Figure S2**

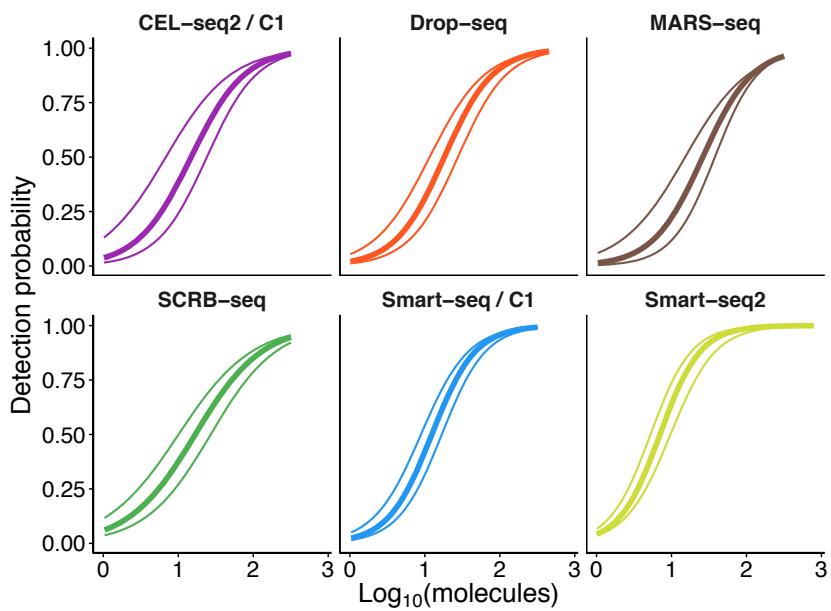


**Figure S3**

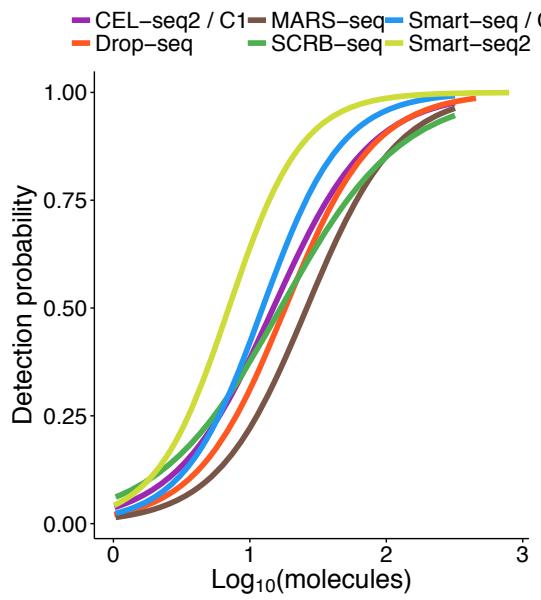


**Figure S4**

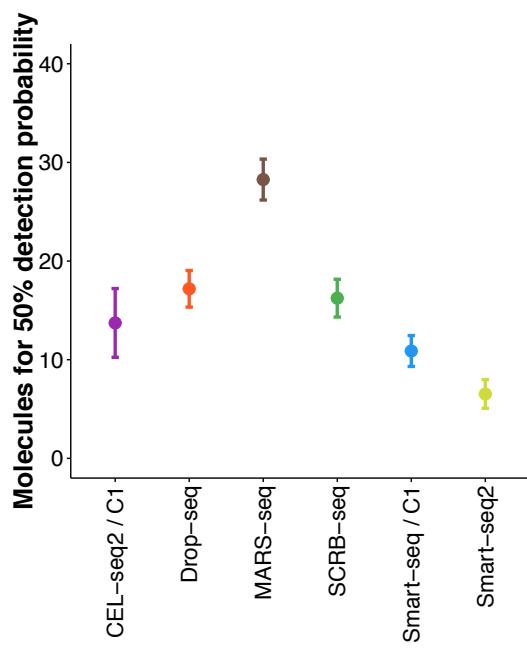
**A**



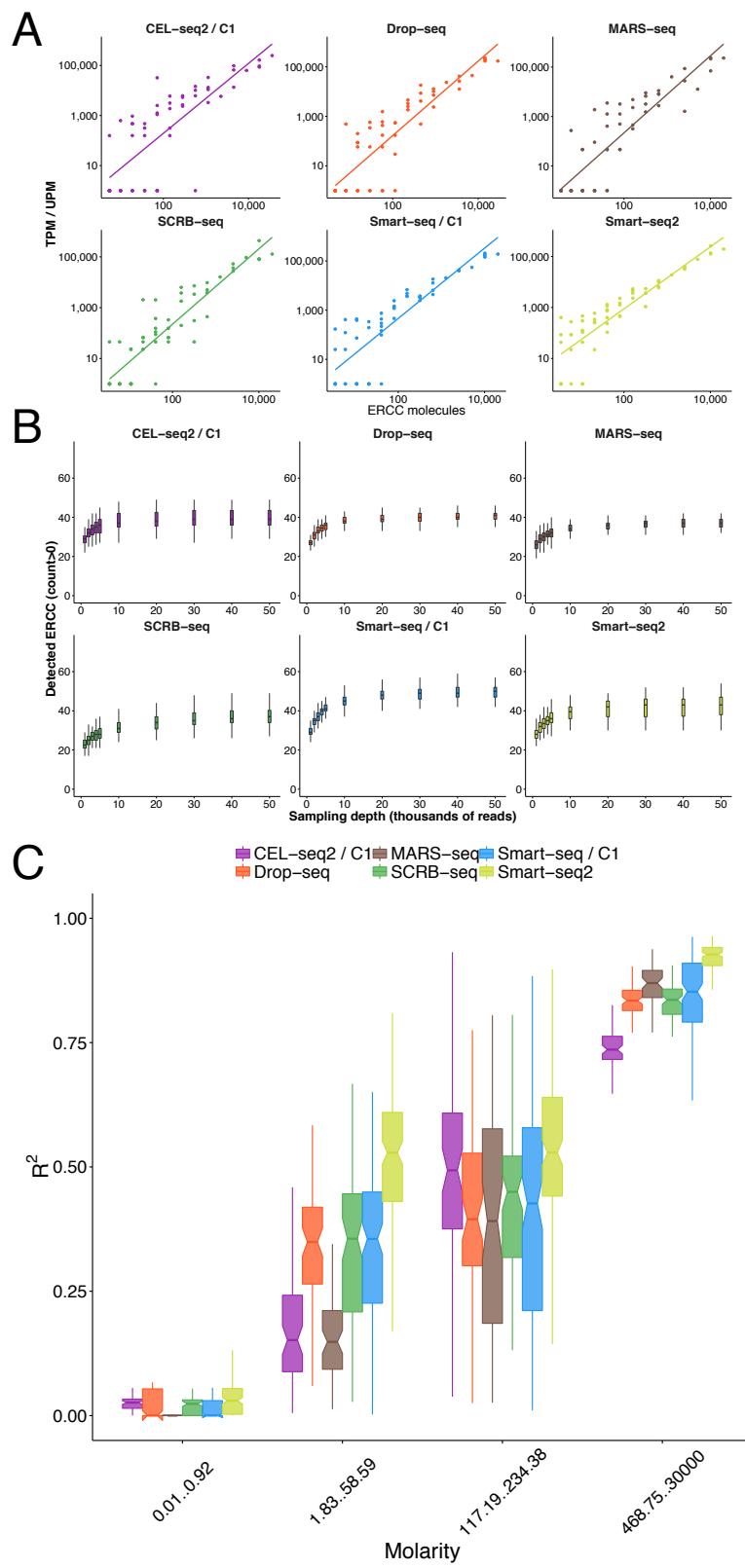
**B**



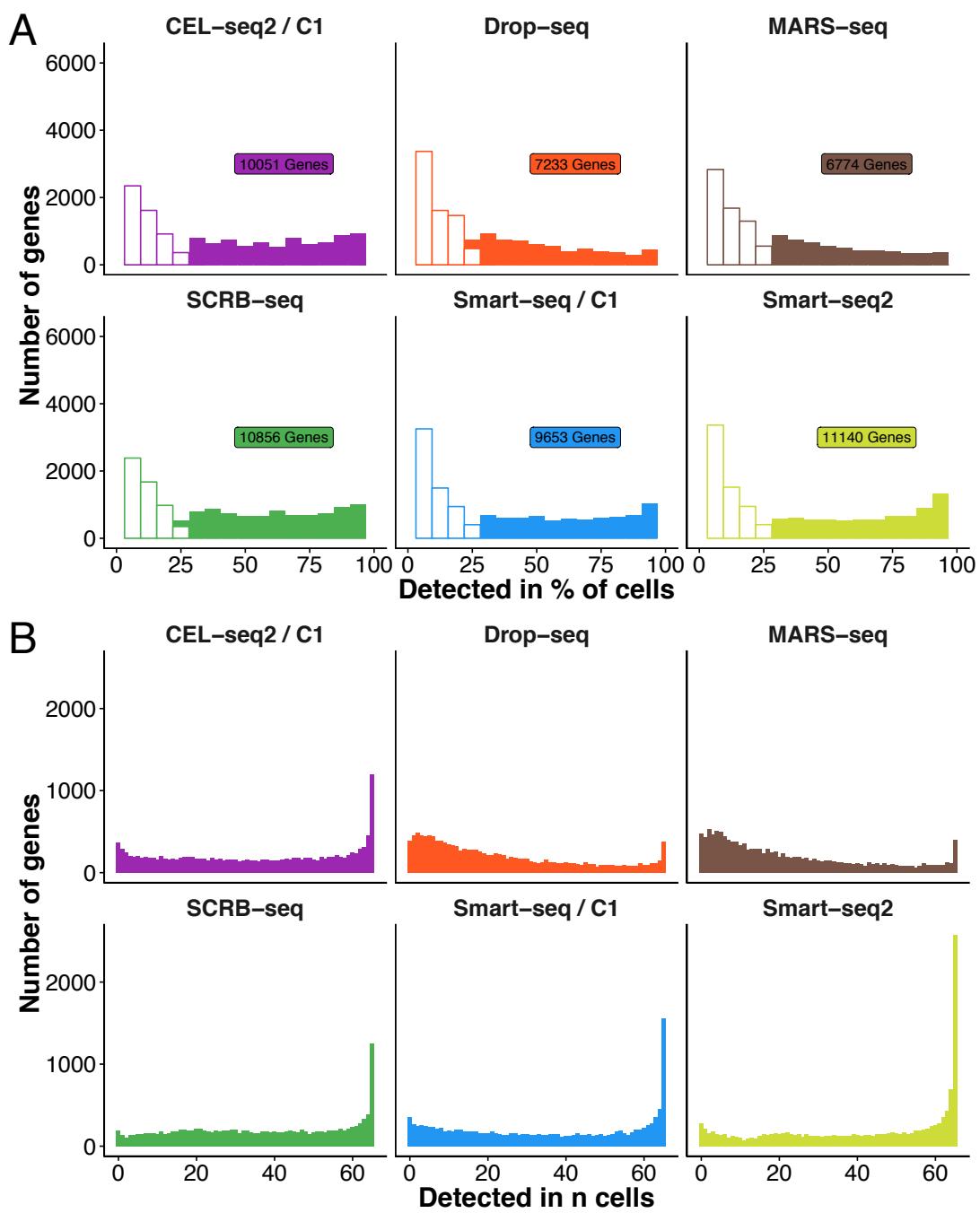
**C**



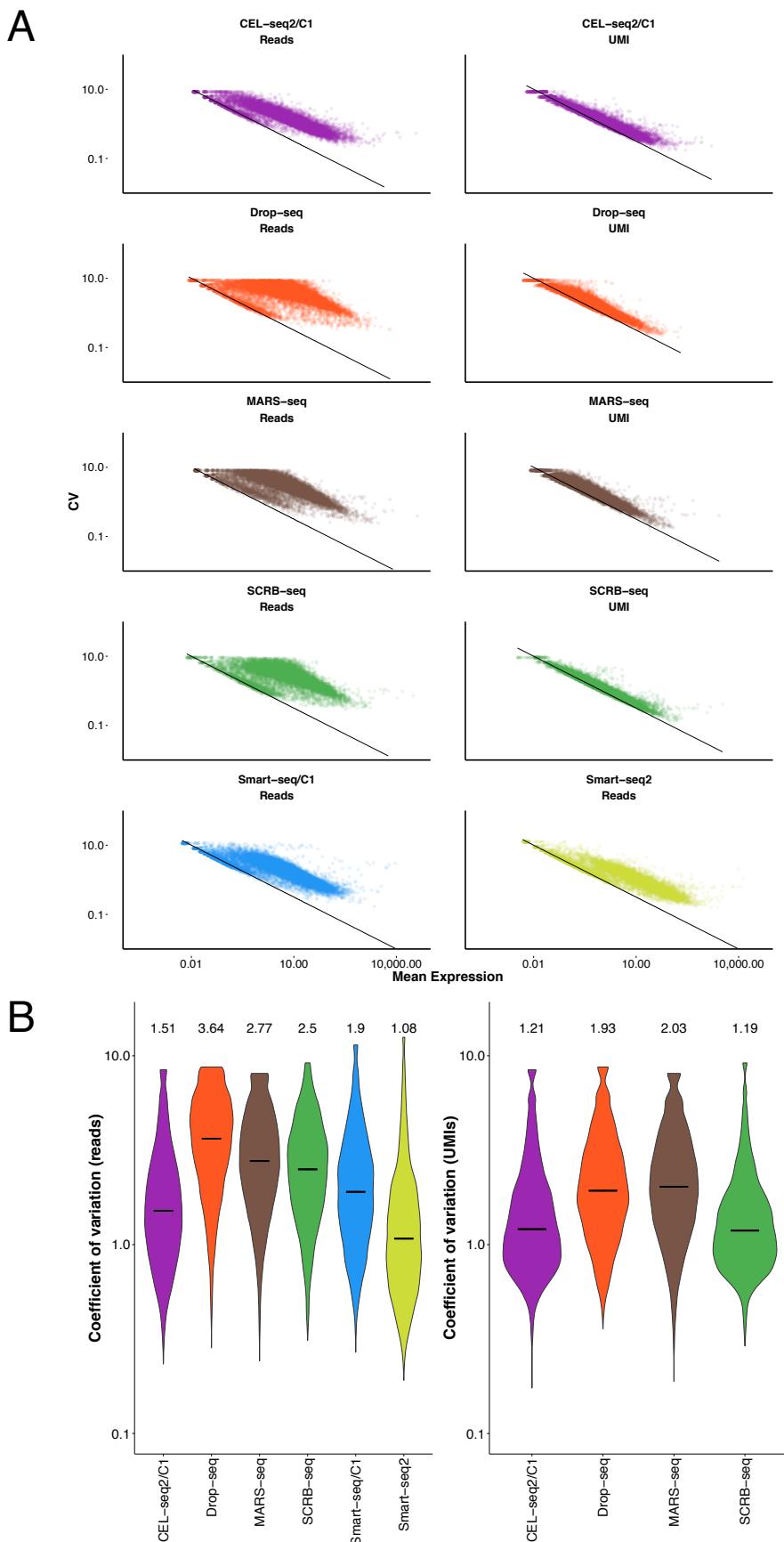
**Figure S5**



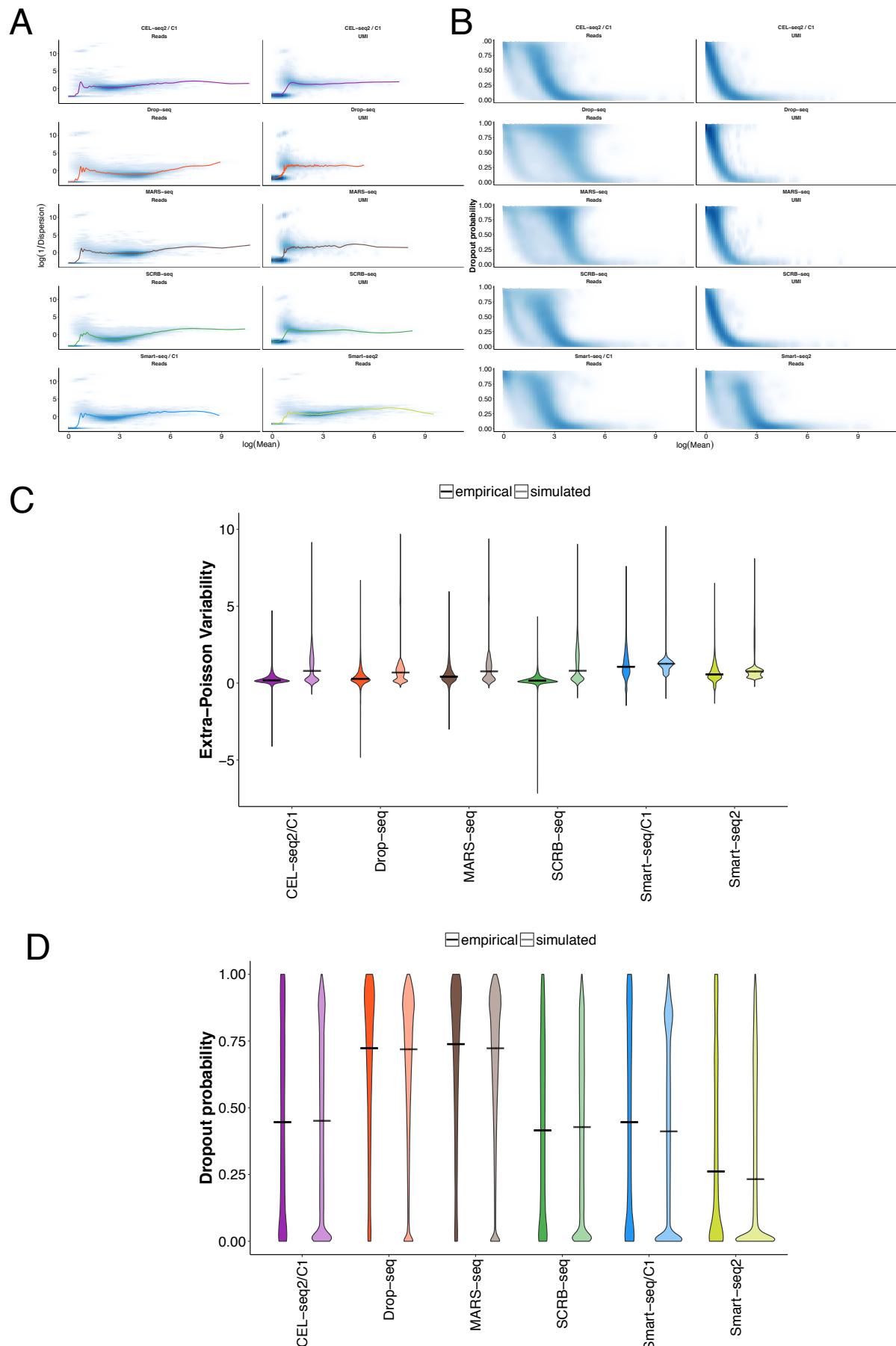
**Figure S6**



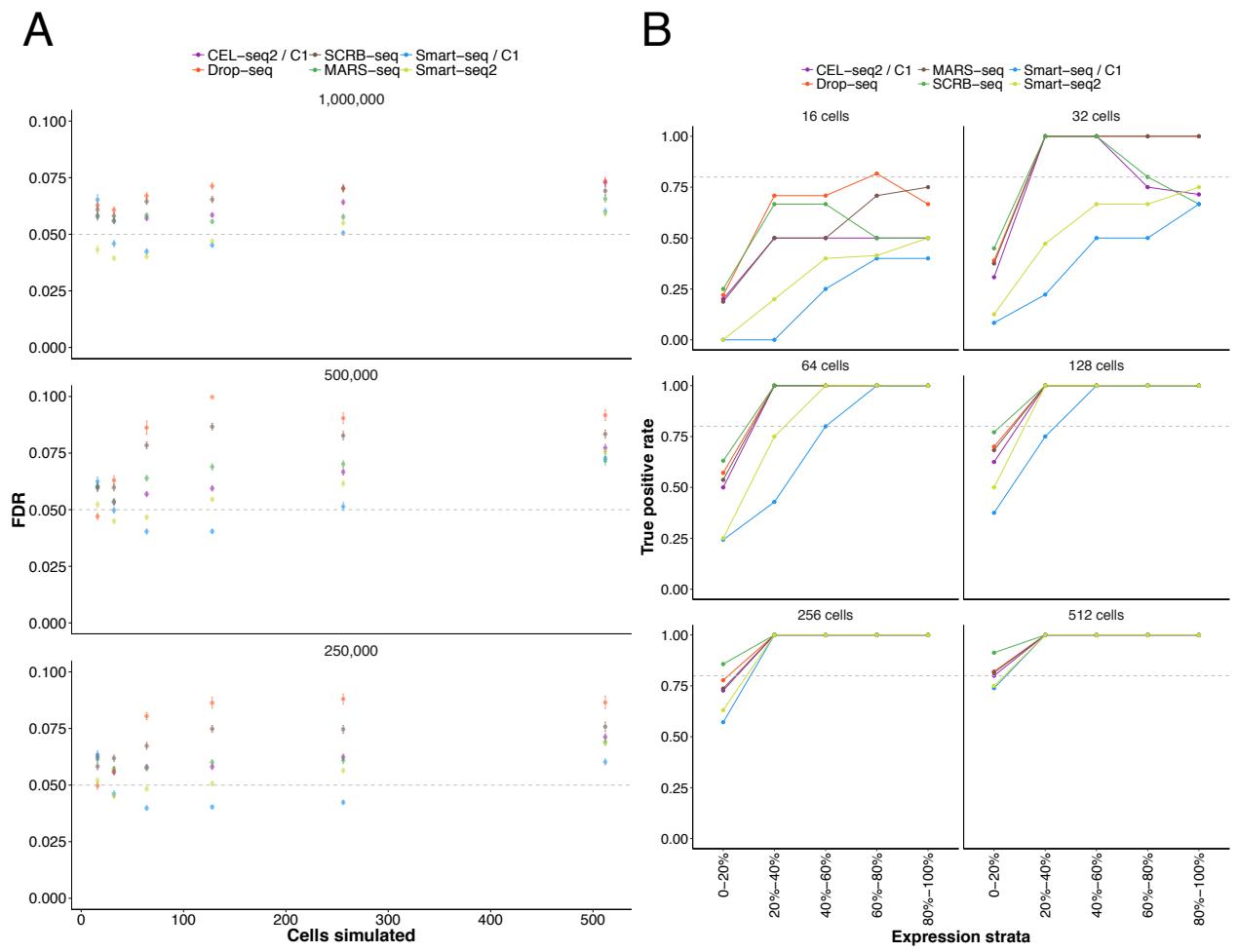
**Figure S7**



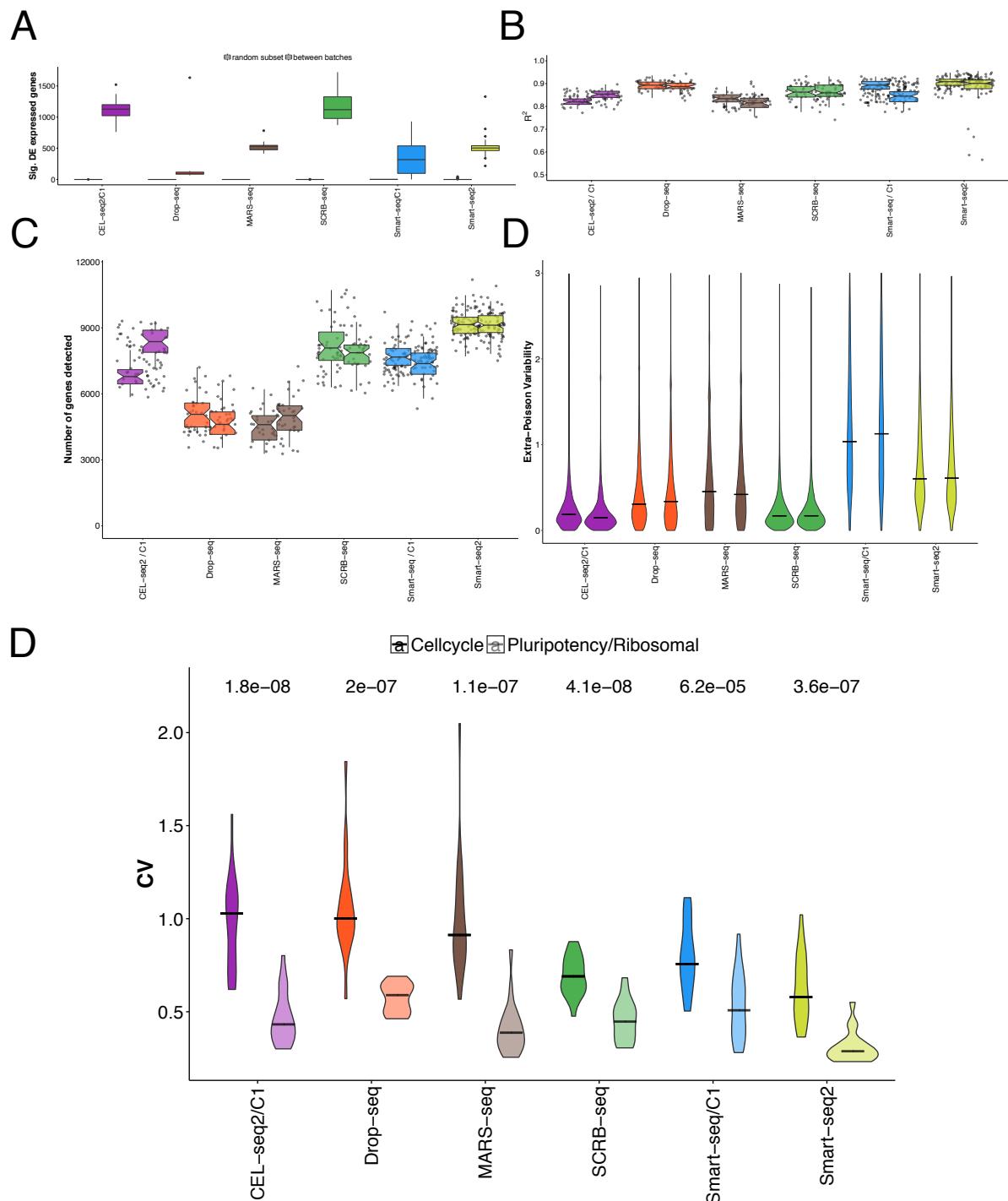
**Figure S8**



**Figure S9**



**Figure S10**



## Supplementary Figure Legends

**Figure S1 (related to Figure 1)** | Quality control and filtering. **A** Drop-seq species mixing experiment using human and murine T-cells. For each cell-barcode human- and mouse read numbers are plotted. **B** Per-base quality scores were summarized using FastQC. Lines indicate median Phred quality score with upper and lower quartile shaded. **C** Total UMI content per cell, with the filter cutoff (two times mean) shown as black lines. Violin plots indicate the density of the UMI content distribution per replicate. **D** Nearest-neighbor filtering based on the maximum pairwise Spearman's rho for each cell. Violin plots indicate the density of rho distribution per replicate. Black lines indicate the employed cutoffs.

**Figure S2 (related to Figure 1)** | Downsampling of scRNA-seq libraries. **A** Detected genes ( $\geq 1$  count) in relation to indicated sequencing depths. The ranges of the boxes indicate the upper and lower quartiles of cells and horizontal bars indicate the medians. **B** Boxplots of the number of detected genes in high-depth sequencing of Smart-seq2 libraries, showing a plateau above 1 million reads. **C** Boxplots of the number of detected UMIs per cell in relation to indicated sequencing depths.

**Figure S3 (related to Figure 3)** | Sensitivity **A** The overlap of detected genes ( $\geq 1$  count) between methods for 65 random cells is displayed as a barplot. Colors indicate the level of overlap: Green (detected in all methods), dark blue (detected in five methods), yellow (detected in four methods), orange (detected in three methods), light blue (detected in two methods), grey (method-specific detection). **B** Gene body coverage (left to right equalling 5' to 3') of  $\sim 3000$  genes detected by Smart-seq/C1 and/or Smart-seq2 (right panel) versus a random control set of 3000 genes detected by all methods. **C** Method-specific detected genes are shown as scatter plots with their rate of detection and mean counts over all cells. **D** For genes and their transcript variants of at least 2 kb length, we calculated the fraction of reads mapping to positions relative to the 3' end. For each method, we show mapping positions and a fit line per replicate. The dashed line indicates theoretical even distribution of reads across the 2.5 kb window. **(E)** Gene expression values were normalized as transcripts per million TPM or UMIs per million UPM. Principal component analysis was performed on the 1000 most variable genes to display the major variance between single cells. The 200 genes with the highest loading for PC1 were analysed and neither showed significant enrichment in GO categories (GOrilla) nor in technical properties such as gene length or GC content.

**Figure S4 (related to Figure 3)** | Detection probabilities were estimated from ERCC dropouts, where the RNA molecule number is known. **A** Thick lines indicate the maximum-likelihood estimate of the detection probability with the thin lines showing the 95% confidence interval of the fit. **B** Shown are per-method maximum-likelihood estimates of mRNA detection probabilities. **C** Sensitivity per method estimated as the 50% probability to detect a transcript. The 95% confidence interval of estimate is displayed as error bars.

**Figure S5 (related to Figure 4)** | **A** Exemplary correlations of ERCC expression values (transcripts per million TPM or UMIs per million UPM) with annotated concentrations. For each method, we chose a representative cell/bead with a linear model correlation coefficient close to the median of all cells. **B** Detection of ERCC genes ( $\geq 1$  count) in relation to sampling depth. Each boxplot represents the median, upper and lower quartile of all cells within each method. **C** Accuracy of scRNA-seq methods. ERCC expression values were correlated to their annotated molarity. Shown are the distributions of correlation coefficients (adjusted  $R^2$  of linear regression model) across methods for bins of ERCC molarity. Each boxplot represents the median, first and third quartile for the  $R^2$  in the indicated bin.

**Figure S6 (related to Figure 5)** | Gene detection sparsity. **A** For all detected genes ( $\geq 1$  CPM) per method, we calculated the rate of detection. Histograms show this measure for detection sparsity. Filled bars represent the genes detected in at least 25% of cells of each method along with the number of these reproducibly detected genes. **B** For genes detected in at least 25% of cells of any method, we calculate the rate of detection in 65 random cells.

**Figure S7 (related to Figure 5)** | Variation in scRNA-seq data. **A** Gene-wise mean and coefficient of variation from all cells are shown as scatterplots for all methods. The black line indicates variance according to the poisson distribution. The two populations of genes seen for read-count data are unamplified genes (close to Poisson, one or very few reads per UMI) and amplified genes (higher CV for a given mean, several reads per UMI). **B** Gene-wise coefficient of variation (CV) of scRNA-seq data were calculated for all cells including detection dropouts. Violin plots are shown for UMI and read-count based quantification indicating the density of the distribution.

**Figure S8 (related to Figure 6)** | **A-B** Power simulation parameters estimated from 1 million reads per cell. **A** Mean expression and size parameters were estimated for each method and their functional relation was approximated by a smooth spline fit. **B** The dropout probability  $p_0$  was calculated per gene and shown in relation to mean expression levels. We

fitted this relationship using a local polynomial regression. **C-D** Validation of power simulation framework. **C** Gene-wise Extra-Poisson Variability was calculated from empirical data and simulated data without addition of differentially expressed genes. Shown are the distributions with the black line indicating the median. **D** Gene-wise dropout rate distributions are shown from empirical data and simulated data. The black line indicates the median dropout rate.

**Figure S9 (related to Figure 6 and Table 1) | A** FDR. Simulations were performed using empirical mean, dispersion and dropout relationships (see Figure S8). For variable sample sizes of n=16, n=32, n=64, n=128, n=256 and n=512, we show points representing the mean FDR of 100 simulations with standard error. **B** I Stratified analysis of power. Shown are TPR for 1 million reads per cell for sample sizes n=16, n=32, n=64, n=128, n=256 and n=512 per group. Genes are grouped in five percentiles of mean expression with lines representing the median TPR of 100 simulations.

**Figure S10 (related to Figure 6) | A-D** Batch effects **A** For each method, we test for differential expression between random subsets of 25 cells per group (left box) and subsets of 25 cells of each batch (right box) in 20 permutations using limma. Shown are the number of significantly differentially expressed genes (FDR <0.01) as boxplots. **B** Sensitivity is shown as the number of detected genes (>= 1 count) per batch. **C** Accuracy is shown per batch as the correlation coefficient of observed expression (TPM/UPM) to annotated ERCC molecule numbers. **D** Precision is shown per batch as the Extra-Poisson Variability for the common 13,361 genes. For 3' counting methods, UMI quantification is shown. The distribution was only shown between values of 0 and 3 to make differences more visible. **D** Cell cycle analysis. For each method, we show the coefficient of variation (CV) for a set of 19 cell cycle genes previously found to be variable in 2i/LIF cultured mESCs (Kolodziejczyk, 2015) (left violin) compared to 19 ribosomal and pluripotency genes. Numbers above the violins indicate p-values of a t-test between the two groups.

## Supplementary Tables

Method	CEL-seq2/C1	Drop-seq	MARS-seq	SCRB-seq	Smart-seq/C1	Smart-seq2
Single-cell isolation	automated in the C1 system	droplets	FACS	FACS	automated in the C1 system	FACS
ERCC spike-ins	yes	no	yes	yes	yes	yes
UMI	6 bp	8 bp	8 bp	10 bp	no	no
Full-length coverage	no	no	no	no	yes	yes
1st strand synthesis	oligo-dT	oligo-dT	oligo-dT	oligo-dT	oligo-dT	oligo-dT
2nd strand synthesis	RNaseH / DNA Pol	template switching	RNaseH / DNA Pol	template switching	template switching	template switching
Amplification	IVT	PCR	IVT	PCR	PCR	PCR
Imaging of cells possible	yes	no	no	no	yes	no
Protocol usable for bulk	yes	no	yes	yes	yes	yes
Sequencing	paired-end	paired-end	paired-end	paired-end	single-end	single-end
Library cost /cell	~9.5€	~0.1€	~1.3€	~2€	~25€	~3/30*

**Table S1 (related to Figure 2):** Overview of single-cell RNA-seq methods.

\* in-house produced Tn5 / commercial Tn5