



Systematic comparison of single-cell and single-nucleus RNA-sequencing methods

Jiarui Ding¹, Xian Adiconis^{1,9}, Sean K. Simmons^{1,9}, Monika S. Kowalczyk¹, Cynthia C. Hession¹, Nemanja D. Marjanovic¹, Travis K. Hughes^{1,2,3,4}, Marc H. Wadsworth^{1,2,3,4}, Tyler Burks¹, Lan T. Nguyen¹, John Y. H. Kwon¹, Boaz Barak⁵, William Ge ¹, Amanda J. Kedaigle ¹, Shaina Carroll^{1,2,3,4}, Shuqiang Li¹, Nir Hacohen^{1,6}, Orit Rozenblatt-Rosen¹, Alex K. Shalek ^{1,2,3,4}, Alexandra-Chloé Villani^{1,6,7}, Aviv Regev ^{1,4,8} and Joshua Z. Levin ¹✉

The scale and capabilities of single-cell RNA-sequencing methods have expanded rapidly in recent years, enabling major discoveries and large-scale cell mapping efforts. However, these methods have not been systematically and comprehensively benchmarked. Here, we directly compare seven methods for single-cell and/or single-nucleus profiling—selecting representative methods based on their usage and our expertise and resources to prepare libraries—including two low-throughput and five high-throughput methods. We tested the methods on three types of samples: cell lines, peripheral blood mononuclear cells and brain tissue, generating 36 libraries in six separate experiments in a single center. To directly compare the methods and avoid processing differences introduced by the existing pipelines, we developed scumi, a flexible computational pipeline that can be used with any single-cell RNA-sequencing method. We evaluated the methods for both basic performance, such as the structure and alignment of reads, sensitivity and extent of multiplets, and for their ability to recover known biological information in the samples.

Single-cell RNA sequencing (scRNA-seq) has emerged as a central tool for identifying and characterizing cell types, states, lineages and circuitry^{1–3}. The rapid growth in the scale and robustness of laboratory protocols and associated computational tools has opened the way to substantial scientific discoveries and to an international initiative, the Human Cell Atlas (HCA), to build comprehensive reference maps of all human cells⁴. Methods for scRNA-seq differ in how they tag transcripts for their cell of origin and generate libraries for sequencing. Low-throughput, plate-based methods^{5,6} sort a cell into a well of a multi-well plate. High-throughput, bead-based methods distribute a cell suspension into tiny droplets^{7–9} or wells^{10,11} containing reagents and barcoded beads to produce a single droplet or well with one cell and one bead that is used to mark all of the complementary DNA generated from that cell. Scalable, combinatorial indexing methods reverse transcribe and barcode messenger RNAs *in situ* inside each cell or nucleus, without physically isolating single cells^{12–14} (Extended Data Fig. 1).

scRNA-seq remains a rapidly evolving field¹⁵, with continued development of new methods and improvement of existing ones. There is thus a need to provide benchmarking information to help users make informed choices based on each method's capabilities and limitations, compare new methods with existing ones, identify shared weaknesses as targets for experimental improvement and allow computational method developers to create new data processing software packages. Previous comparisons of scRNA-seq methods^{16–21}, though useful, have several shortcomings. Many are

outdated, incomplete, inapplicable (for example, not actually performed with single cells) or insufficiently controlled (for example, performed using different biological samples for comparisons); others limit their assessment to basic technical factors, but do not assess the key benchmark of the ability to recover meaningful biological information, such as population heterogeneity and structure. In particular, comparisons often focused on cultured cell lines, even though in practice most scRNA-seq studies seek insights from tissues and primary cells.

Here, we systematically and directly compared seven methods (Fig. 1 and Extended Data Fig. 1), including two low-throughput plate-based methods (Smart-seq2 (ref. ⁵) and CEL-Seq2 (ref. ⁶)) and five high-throughput methods (10x Chromium⁹, Drop-seq⁸, Seq-Well¹⁰, inDrops⁷ and sci-RNA-seq¹²), producing expression profiles from ~92,000 cells overall. We selected representative methods that are more widely used and for which we had the expertise and resources to prepare libraries. We analyzed three sample types—a mixture of human and mouse cell lines, human peripheral blood mononuclear cells (PBMCs) and mouse cortex nuclei, each sample with two replicates—to generate a total of 36 different scRNA-seq libraries. For mouse cortex, we tested four single-nucleus RNA-seq methods^{9,12,22,23}. For each sample type, we characterized performance with basic metrics, and for PBMC and cortex libraries, we examined how well methods capture biological information, a critical part of most scRNA-seq studies and one that has not been evaluated in other benchmarking exercises that used relatively homogeneous

¹Broad Institute of MIT and Harvard, Cambridge, MA, USA. ²Institute for Medical Engineering and Science, Department of Chemistry, MIT, Cambridge, MA, USA. ³Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA. ⁴Koch Institute of Integrative Cancer Research, Cambridge, MA, USA.

⁵McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA. ⁶Center for Cancer Research, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. ⁷Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA, USA. ⁸Howard Hughes Medical Institute, Department of Biology, MIT, Cambridge, MA, USA. ⁹These authors contributed equally: Xian Adiconis, Sean K. Simmons. ✉e-mail: jlevin@broadinstitute.org

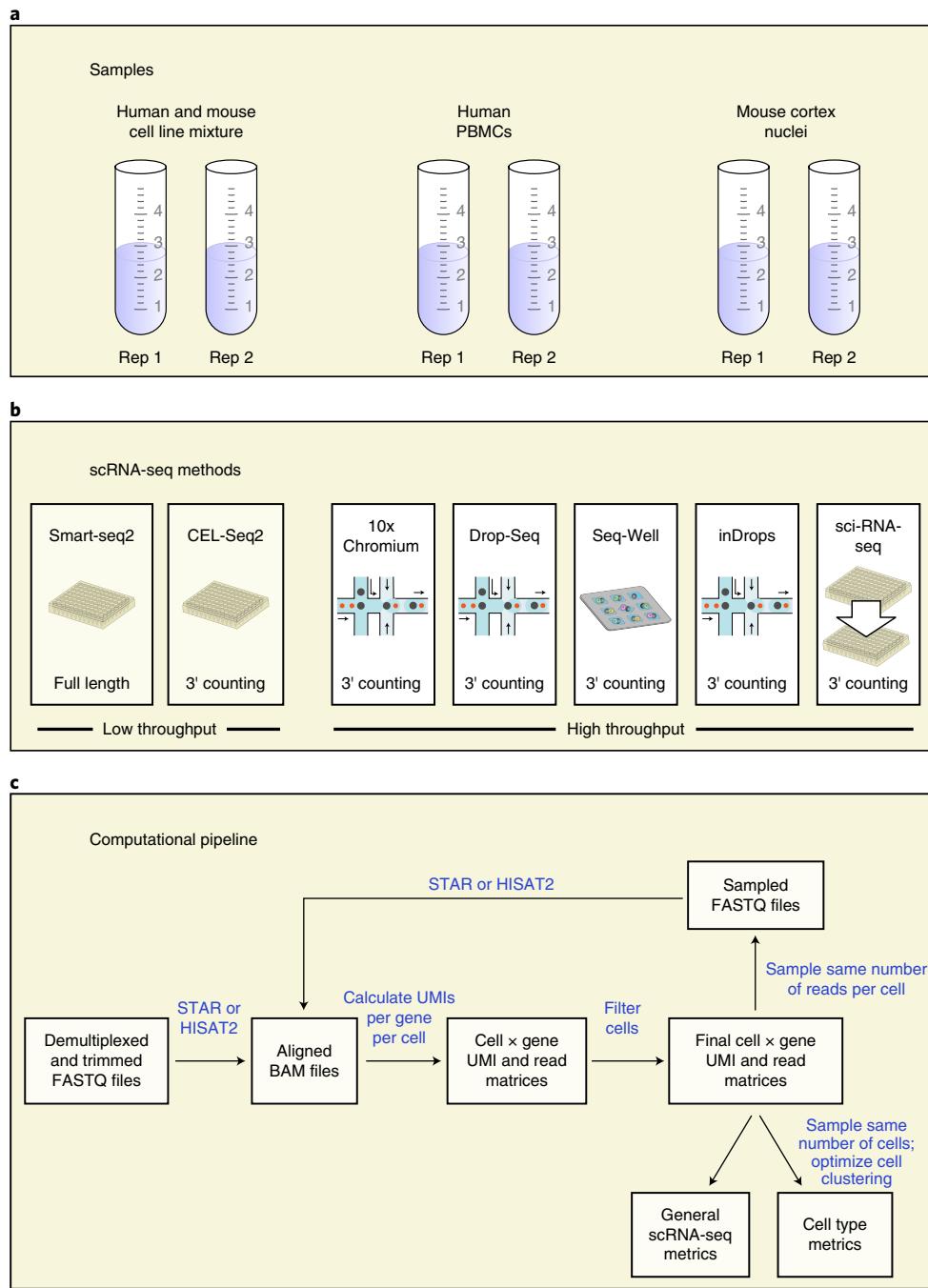


Fig. 1 | Study overview. **a**, Samples. **b**, scRNA-seq methods. **c**, Computational pipeline summary. Cell line mixtures were tested with all methods. PBMCs were tested with all methods except sci-RNA-seq. Cortex nuclei were tested with Smart-seq2, 10x Chromium, Drop-seq (aka DroNc-seq for nuclei) and sci-RNA-seq. Additional details can be found in Extended Data Figs. 1 and 2. Rep, repeat.

cell lines^{16,20}. Our study provides both immediate guidance on each method's relative performance, and an experimental and computational framework to assess future techniques. For the low-throughput methods, Smart-seq2 and CEL-Seq2 performed similarly, though the latter may be affected more by contaminating reads from other cells. Among the high-throughput methods, 10x Chromium was the top performer.

Results

A comparison of scRNA-seq methods. We selected seven scRNA-seq methods for comparison and tested each with up to three sample types: a mixture of mouse and human cell lines, human PBMCs

and mouse cortex nuclei (Fig. 1 and Extended Data Figs. 1 and 2). We chose to profile a cell line mixture with 50% human HEK293 and 50% mouse NIH3T3 cells (mixture) because (1) these cells are a common test^{8,9,12,14} for samples with relatively high amounts of RNA per cell; and (2) multiplets, two or more cells being sequenced together and assigned one cell barcode, can be detected when cell barcodes have a substantial fraction of reads from both species. We profiled frozen human PBMCs because (1) they are a heterogeneous mixture of cells, particularly with respect to their amount of RNA per cell, yet they do not require dissociation (a separate technical challenge); and (2) their cell types and associated expression patterns are well-studied. (We do not include data for sci-RNA-seq

with PBMCs because we detected very few genes (for example, <10 genes) per cell in several experiments (data not shown). We extended our study of single cells to single nuclei, as such samples have distinct properties including lower RNA input amounts. Using four methods that had previously been applied to nuclei^{12,22,23}, we profiled the mouse cortex because brain tissue is a major example of a tissue type commonly analyzed through single-nucleus RNA-seq. Each sample type was tested in two experiments (Mixture1, Mixture2, PBMC1, PBMC2, Cortex1 and Cortex2) run on different days to assess reproducibility.

In each experiment, we aimed to collect data from ~384 cells for the low-throughput methods, ~3,000 cells for the high-throughput methods and a bulk RNA-seq library as a control. We sequenced libraries together to similar coverage based on library type, except as noted (see Methods).

scumi computational pipeline allows unified analysis across any scRNA-seq method. Because each method had its own standard computational pipeline, we developed and used a ‘universal’ pipeline, to permit direct comparison of all of the experimental methods, and remove processing differences introduced by these existing pipelines (Extended Data Fig. 2). First, we developed the scumi software package (single-cell RNA-sequencing with UMI; Extended Data Fig. 2a), which starts from FASTQ files as input and generates gene × cell expression count matrices for downstream analyses.

Second, we addressed the major preprocessing challenge of filtering out low-quality cells before downstream analysis (Extended Data Fig. 2b). This is particularly important when comparing methods, to ensure that our approach is fair to all methods and less subjective. When selecting the cell barcodes with the largest number of reads or UMIs assigned to them, the challenge is to decide which threshold to choose for excluding lower-quality cells or barcodes likely reflecting ambient RNA rather than real cells (or nuclei). For the mixture experiments, we removed cells with low quantities of UMIs or reads per cell. For the more complex PBMCs and cortex samples, consisting of different cells with different characteristics, such a simple approach could bias against the recovery of cell types with relatively small amounts of RNA. Instead, we first looked at more cell barcodes than we expected to truly recover from experiments, did an initial clustering, identified differentially expressed genes to find cluster-specific marker genes and removed cells in clusters likely to be low quality (Extended Data Fig. 2b and see Methods).

Third, before calculating metrics that potentially show improvements with greater sequencing depths, such as the number of genes per cell or ability to detect known cell types, we sampled the same number of reads per cell for all of the methods of the same type, either low- or high-throughput, in a given experiment (see Methods and Extended Data Fig. 2c). This leads to better relative performance for methods that have a higher fraction of informative reads; that is, those reads aligning to genes and present in cells used for analysis. Note that because for most experiments we sequenced the poly(T) sequences that follow the cell barcode and UMI sequences in all methods except Smart-seq2, we tracked and removed reads without poly(T) at the expected positions because reads lacking the expected read structure are unlikely to be informative for further analysis.

Finally, we assessed the methods by several key metrics spanning (1) the structure and alignment of reads to the nuclear and mitochondrial genomes; (2) sensitivity in capturing RNA molecules; (3) extent of multiplets (assessed in mixture experiments); (4) their technical precision/reproducibility with respect to expression estimates; and (5) the ability to recover meaningful biological distinctions in cell types (for PBMC and cortex experiments).

Read structure and alignment reveal efficiency differences among methods. First, we characterized the methods by the distribution

of reads from each library with respect to their structure and alignment with the genome (Extended Data Fig. 3). These metrics inform about the ‘efficiency’ of methods in generating useful reads for downstream analysis. In the mixture experiments, we considered only uniquely mapped reads to minimize the effects of multi-mapped reads on calculating cell multiplet rates and other metrics. Methods differed with respect to reads without poly(T) at the expected positions (Supplementary Note 1). We next considered the distributions of reads across these categories: exonic, intronic, intergenic, overlapping different genes (ambiguous), multi-mapped and unmapped. Exonic reads are typically the only reads used in scRNA-seq studies of cells, whereas intronic reads are also used for studies with nuclei^{12,25}. In the mixture experiment, both replicates of Smart-seq2 and one replicate of inDrops had the highest fraction of exonic reads (51.0%, 53.7% and 56.9%, respectively), with sci-RNA-seq performing worst (28.7% and 29.4%; Extended Data Fig. 3a). Overall, the PBMC samples had a lower fraction of reads aligned to exons than the mixture samples (Extended Data Fig. 3a,b), with one replicate of inDrops having the highest fraction of exonic reads (46%), and Seq-Well having the lowest (20%; Extended Data Fig. 3b). To explore the origin of unmapped reads, we further analyzed the PBMC datasets and found that most of these reads are low quality, align to adapter sequences added during library construction or contain stretches of poly(A) (Supplementary Fig. 1a). Using Trimmomatic²⁶ to remove low-quality and adapter sequences from the unmapped reads (see Methods), we recovered <5% of the unmapped reads in each case, except for inDrops PBMC1 and Seq-Well PBMC2, for which we recovered ~8% and ~18%, respectively, of the initially unmapped reads (Supplementary Fig. 1b).

To determine the extent to which existing annotation limits recovery of reads aligning to genes²⁷, we used the PBMC1 and PBMC2 bulk RNA-seq to create a matched transcriptome and customized annotation (see Methods). The customized transcriptome annotations led to very few (<2%) additional reads aligning (Supplementary Table 1).

While the relative performance of each method was generally similar between the cortex nuclei and the other experiments, there was a higher ratio of intron-aligning reads to exon-aligning reads in nuclei than in cells (Extended Data Fig. 3), as expected because nuclei contain a higher proportion of unspliced transcripts than whole cells²⁸. We assessed whether reads aligned in the sense or antisense orientation for each method, except Smart-seq2, which is not strand-specific, and investigated antisense reads (Supplementary Note 2).

Method performance relative to mitochondrially aligned reads was also analyzed (Supplementary Note 3).

Similar relative ranking of method sensitivity across experiments. As scRNA-seq methods start with limited RNA inputs, a key quality metric is the sensitivity, or the ability to capture RNA molecules. We assessed the sensitivity of each method by measuring the number of detected UMIs or genes per cell in datasets sampled to the same number of reads per cell (see Methods and Supplementary Table 2). The only exception was Seq-Well PBMC1 with ~46,000 reads per cell compared with ~69,000 reads per cell for the other high-throughput methods in PBMC1 (Supplementary Table 2). For the mixture experiments, we report the results for mouse and human cells separately as the numbers of UMIs and genes per cell in the two cell types differ, such that differences in the ratio of human to mouse cells among the libraries (Supplementary Table 3) could skew the results, but the overall ranking of the methods is the same for both human and mouse cells (Fig. 2a,b and Extended Data Fig. 4a).

Overall, the low-throughput methods Smart-seq2 and CEL-Seq2 had the highest sensitivities, as expected¹⁶, whereas among high-throughput methods, 10x Chromium detected the most UMIs and genes per cell. In the mixture experiments, inDrops had the lowest

sensitivity and Seq-Well detected fewer genes per cell compared with 10x Chromium (v2) and sci-RNA-seq, but more genes per cell compared with Drop-seq and inDrops. The relative ranking of the methods was generally consistent when comparing the median number of detected UMIs per cell (Fig. 2a), detected genes per cell (Fig. 2b) or mean detected reads per cell (Extended Data Fig. 4a). Similarly, in PBMCs low-throughput methods detected more UMIs and genes per cell than the high-throughput methods (Fig. 3 and Extended Data Fig. 4b), with similar performance of Smart-seq2 (2,406 and 2,632 median number of genes detected) and CEL-Seq2 (2,717 and 2,545; Fig. 3b). Among the high-throughput methods, 10x Chromium (v3) had the highest median number of UMIs (4,494) and genes (1,482) per cell (Fig. 3), and inDrops (366 and 1,118 UMIs; 256 and 568 genes) and Seq-Well (844 and 577 UMIs; 513 and 372 genes) had the lowest (Fig. 3). In cortex nuclei, Smart-seq2 was the only low-throughput method tested and we sequenced to a slightly higher depth than for the other samples (Supplementary Tables 3–5) and used all of the reads. As expected¹⁶, Smart-seq2 detected more genes per cell than the high-throughput methods (Fig. 4 and Extended Data Fig. 4c). Among the high-throughput methods, 10x Chromium (v2) had the highest median number of UMIs (5,126 and 3,127) and genes (2,462 and 1,744) per cell (Fig. 4 and Extended Data Fig. 4c).

To explore how sensitivity varied with sequencing depth, we sampled fewer reads per cell from each method in the PBMC datasets (based on the molecular information matrices; see Methods). For each dataset, the relative ranking of the methods with respect to the median number of genes per cell (Extended Data Fig. 5a,b) or UMIs per cell (Extended Data Fig. 5c,d) detected remained the same at all sequencing depths tested. In addition, the number of genes detected may not have saturated at these sequencing depths, except for Seq-Well PBMC2, though it is possible that low levels of reads from other cells increasing the apparent number of genes detected may confound these analyses. Comparing this analysis sampling from the molecular information matrices versus from the raw reads for five of the PBMC1 libraries yielded indistinguishable results (Supplementary Fig. 2). Notably, the number of genes and UMIs per cell are highly correlated (Extended Data Fig. 5e).

We also performed comparisons with published datasets (Supplementary Note 4).

Mixture experiments enable detection of multiplets and reads from other cells. In the mixture experiments, we were able to assess the frequency of multiplets, because we started with a mixture of human and mouse cells. The observed multiplet rates were <3.5% for all seven tested methods (Fig. 2c), except for the first inDrops experiment, which also had a high fraction of reads without poly(T) (Supplementary Table 3). The multiplet rate depends on the number of cells used in each experiment⁹ and the ratio of mouse to human cells, but it was not possible to sequence the same number of cells or the same ratio of mouse and human cells with each method (Supplementary Table 3). The multiplet rates of low-throughput methods were the lowest (<1%), as expected as fluorescence-activated cell sorting was used to place a single cell in each well of a plate (Fig. 2c).

We also examined how the estimated multiplet rate varied with the number of detected UMIs per cell. Generally, multiplet rates were higher in cells with the largest number of UMIs (Fig. 2c), as expected because multiplets are expected to have more RNA input. While most cells with intermediate quantities of UMIs were not multiplets, cells with the lowest number of UMIs in some cases had higher rates, suggesting that these cells might be low quality or have more contributions from cell-free ambient RNA (Fig. 2c). In sci-RNA-seq experiment 2, the rate of multiplets decreased more gradually than for other methods for unknown reasons (Fig. 2c).

We also used the mixture experiments to ask whether the genes detected in a cell were actually from that cell instead of ‘contamination’ from other cells. As sequencing depth increased, more genes were detected from the ‘wrong’ species (Extended Data Fig. 6a,b), as reflected by the slope of a regression line along the cell barcodes adjacent to each axis (see Methods), such that the best-performing methods have the lowest slope. For the low-throughput methods, Smart-seq2 performed much better than CEL-Seq2. Among the high-throughput methods, inDrops had the best (lowest) slope and Seq-Well had the highest slope.

Technical precision, reproducibility and accuracy in gene expression quantification. To assess technical precision in the mixture experiment, which consisted of two homogeneous cell lines grown in controlled conditions in culture, we also compared the variation in scRNA-seq data, which we expect to be primarily driven in this case by technical variation¹⁶, although some intercellular heterogeneity may still have been present in our cell cultures. Previous studies have demonstrated that such technical variation generally follows Poisson distributions^{16,29,30}. CEL-Seq2, inDrops and Drop-seq consistently had relatively low extra Poisson coefficients of variation (Extended Data Fig. 7). Consistent with previous findings, Smart-seq2 data had the highest extra Poisson coefficient of variation, most likely because no UMIs were used (Extended Data Fig. 7).

We also compared reproducibility between replicates and accuracy using bulk and pseudo-bulk data (Supplementary Note 5).

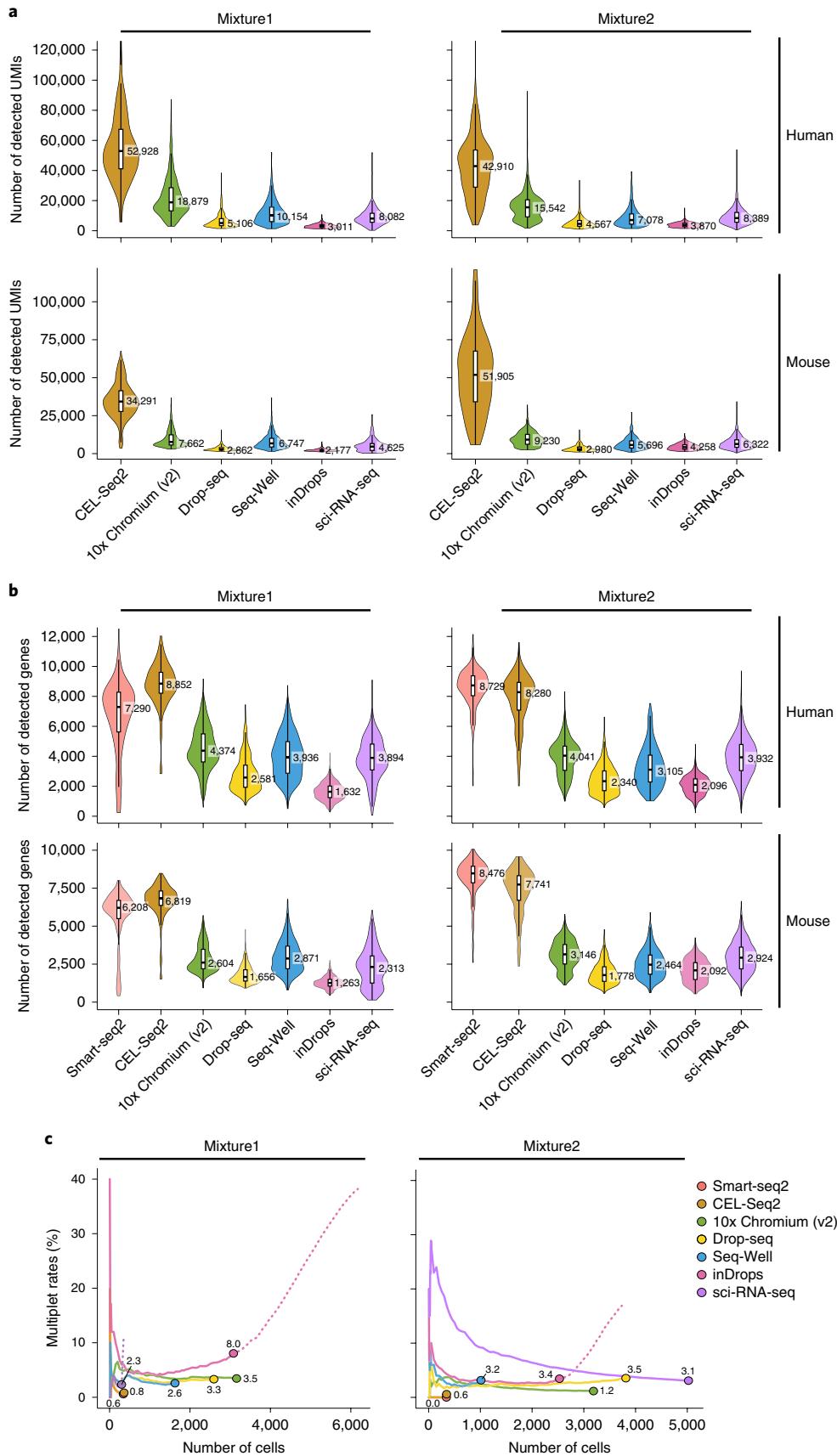
Methods vary in their ability to distinguish and recover cell types. A key consideration in choosing an scRNA-seq method is its ability to uncover the underlying biology of interest. Among the many biological features studied by scRNA-seq, one of the most prominent use cases is the identification of distinct cell types by clustering scRNA-seq profiles. Both the PBMC and mouse cortex datasets consist of diverse cell types, and were chosen to allow us to compare methods for this use case.

To this end, we processed the data with the goal of a fair and optimal assessment of each method. Not only did we sample the same number of reads per cell for each low- and high-throughput method in each experiment, as in ‘Similar relative ranking of method sensitivity across experiments’ above for the sensitivity metrics, we also performed another round of sampling to use the same number of cells from each low- and high-throughput method in each experiment (see Methods). The only exceptions were Seq-Well PBMC2, which had fewer cells (Supplementary Tables 2 and 4) because we

Fig. 2 | Performance metrics for mixture experiments. a,b, Distribution of the number of UMIs (a) or genes (b) in human (top) or mouse (bottom) cells in the two Mixture experiments ($n=1$ biologically independent sample per experiment). For a and b, median and box plots were based on all of the cells, but a few outlier cells were omitted in drawing the violin plots. Box plots denote the medians (labeled on the right) and the IQRs. The whiskers of each box plot are the lowest datum still within 1.5 IQR of the lower quartile and the highest datum still within 1.5 IQR of the upper quartile. Violin plot width is based on a Gaussian kernel density estimate of the data (estimated by the density function with standard parameters), scaled to have maximum width = 1. **c,** Multiplet frequency. We ordered cells based on the number of detected UMIs (or reads for Smart-seq2), from highest (left) to lowest (right). For a given number of cells (x axis value), the plot shows the percentage of cells that are multiplets. The dotted lines for sci-RNA-seq Mixture1 and inDrops Mixture1 and Mixture2 show the multiplet rate including low-quality cells that were not included in subsequent analysis.

used only one microwell array for that experiment, while we used two arrays for Seq-Well PBMC1, and DroNc-seq Cortex2, which had fewer cells for unknown reasons (Supplementary Tables 2 and 5).

For each dataset, we clustered the cells or nuclei based on their gene expression profiles to assess how well they detected the known cell types and their associated transcriptional profiles. For each



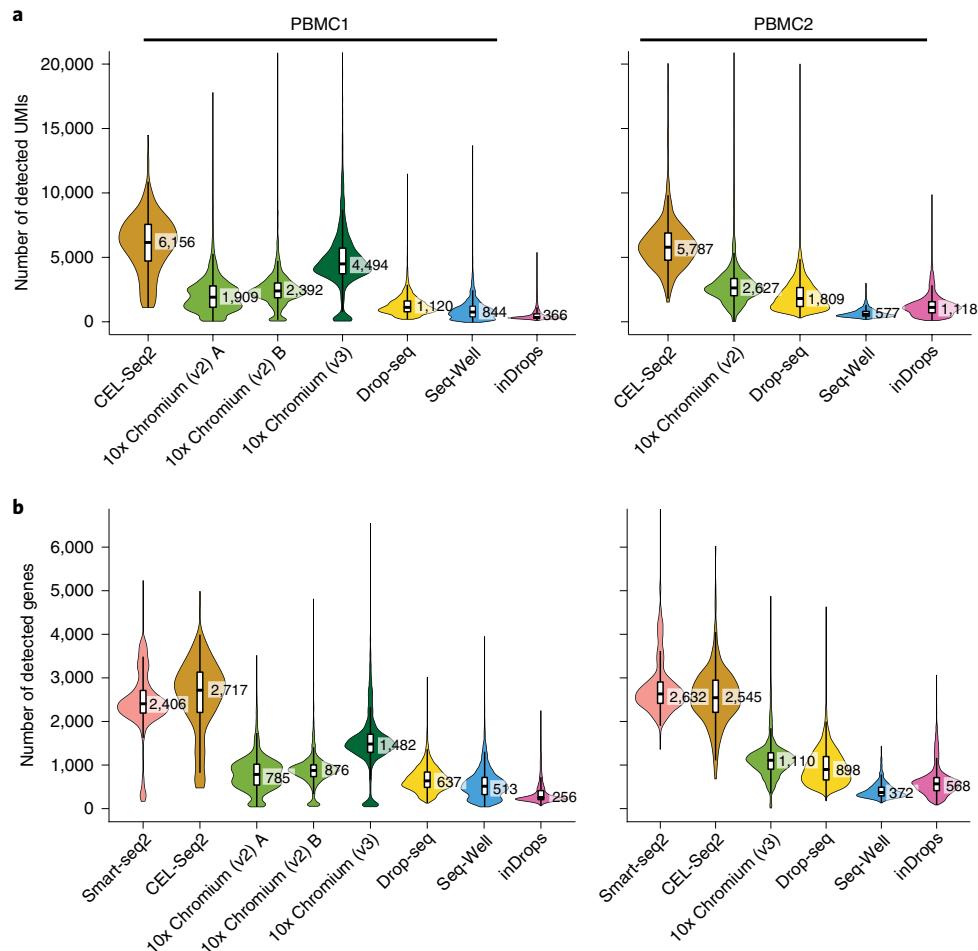


Fig. 3 | PBMCs sensitivity. **a,b**, Distribution of the number of UMIs (a) or genes (b) per cell for each method in the two experiments ($n=1$ biologically independent sample per experiment). Violin and box plot elements are defined as in Fig. 2.

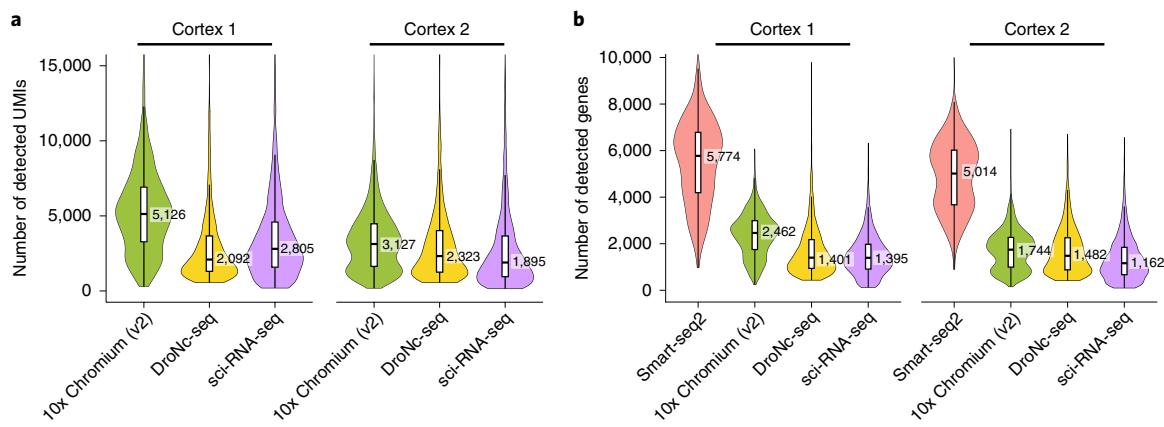


Fig. 4 | Cortex nuclei sensitivity. **a,b**, Distribution of the number of UMIs (a) or genes (b) per cell for each method in the two experiments ($n=1$ biologically independent sample per experiment). Violin and box plot elements are defined as in Fig. 2.

dataset, we searched a range of parameters to select the optimal clustering to recover each of the expected cell types (see Methods). We assigned each cluster a cell type identity based on known marker genes (see Methods). To quantify the quality of the clusters at separating cell types, we scored the expression of each cell for each cell type signature generated from known marker genes and calculated the area under the receiver operating characteristic curve (AUC) for

each cluster to estimate how well the cells in a cluster score for each cell type (see Methods). The AUC summarizes the performance of the gene signature scores in separating a cluster of cells from the rest of the cells, with $AUC=1$ for all cell types as the ideal outcome.

For PBMCs, methods varied in the ability to distinguish cell types, in the proportion of cell types recovered and, in some cases, in the recovery of certain cell types altogether. As expected,

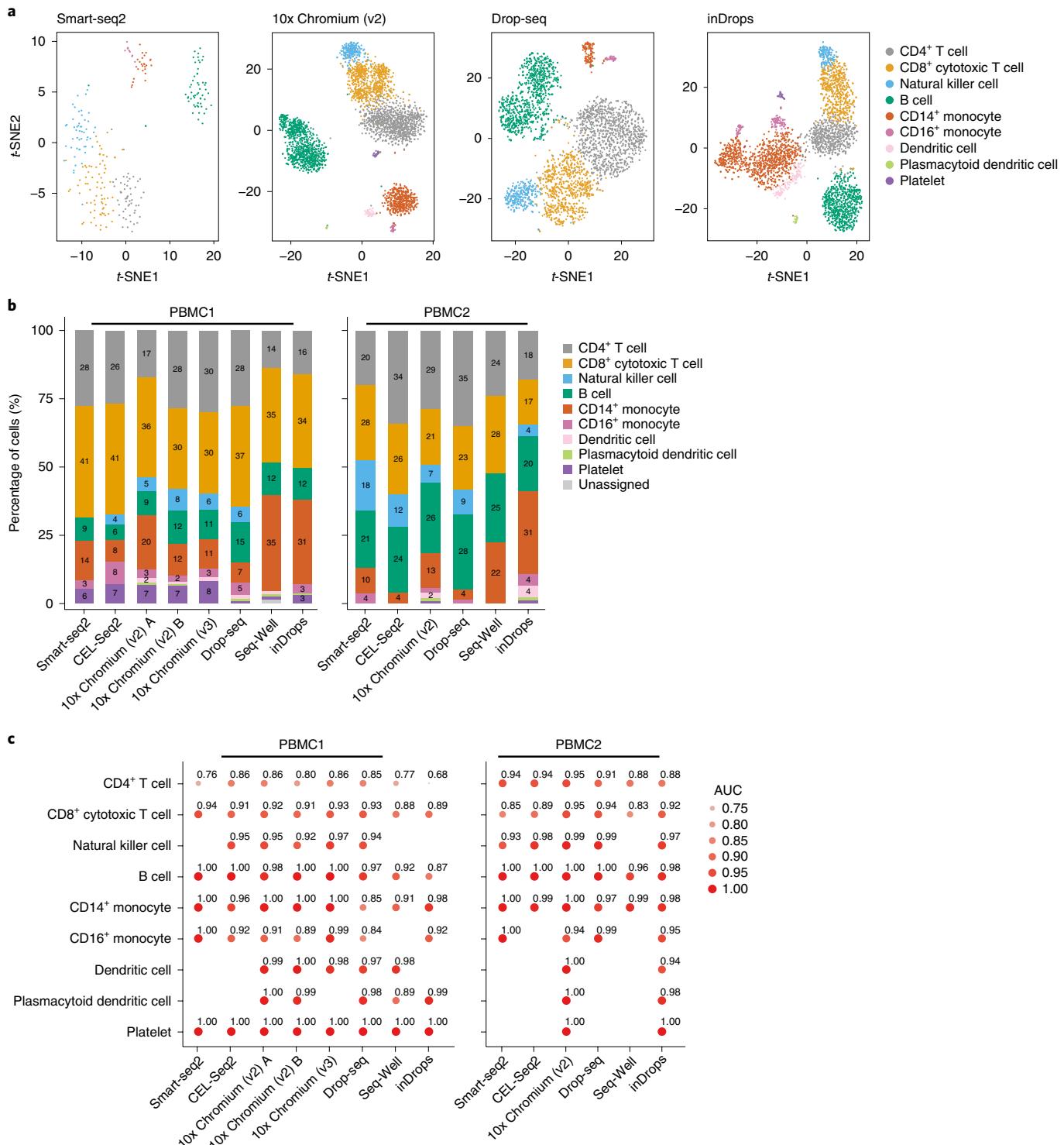


Fig. 5 | Cell type identification and assignment in PBMCs. **a**, t-SNE plots of single-cell profiles (dots) from representative PBMC2 libraries colored by cell type. **b**, Proportion of cells of each cell type (y axis) detected with different methods (x axis). Those not labeled with a number rounded to 1 or less. Sum does not always add to 100 due to this and rounding. **c**, The AUC (dot size, color and value) of each cluster from classifying the cell type to the cluster it was assigned for PBMC1 and PBMC2. See Supplementary Table 2 for the numbers of cells used ($n=1$ biologically independent sample per experiment).

methods had more difficulty in distinguishing transcriptionally related cell types, such as CD4⁺ T cells, CD8⁺ cytotoxic T cells and natural killer cells (Fig. 5a,b and Extended Data Fig. 8). From the t-distributed stochastic neighbor embedding (t-SNE) plots for PBMC2, we observed that 10x Chromium and inDrops performed well (Fig. 5a and Extended Data Fig. 8b). As all of the libraries for

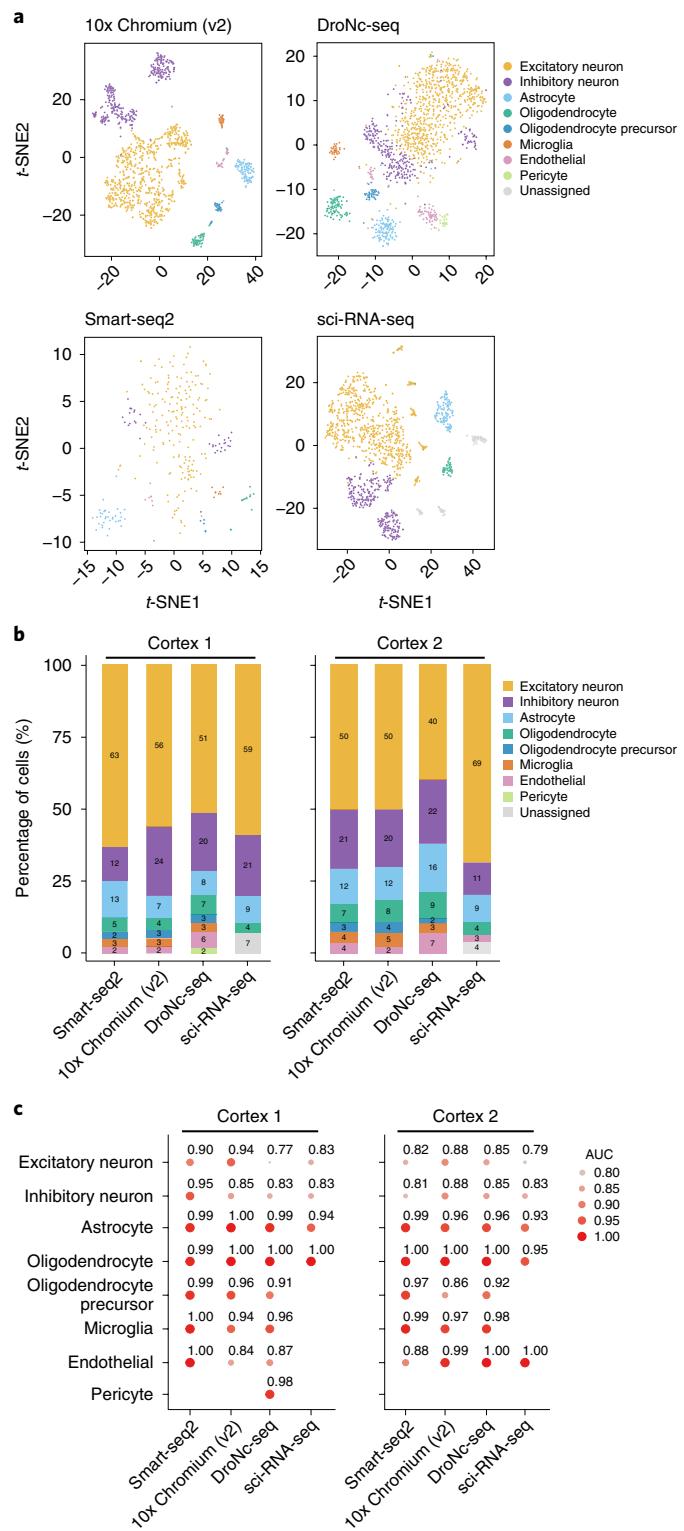
each experiment were generated from the same sample, we assessed the consistency across methods in the fraction of cells assigned to each cell type within an experiment (Fig. 5b). Generally, most methods successfully recovered the abundant cell types in PBMCs, but varied in the relative abundance of cell types. Methods also varied in whether cell types were detected, particularly for the rarer cell types,

such as plasmacytoid dendritic cells and contaminating platelets that were captured in various proportions across methods (Fig. 5b). Although platelets are not PBMCs, their presence in this dataset reflects incomplete purification of PBMCs from whole blood, as has been observed previously⁹. For the low-throughput methods, we did not profile a sufficient number of cells to recover the rarer cell types (Fig. 5b) and they performed similarly for the AUC measurements (Fig. 5c). In PBMC1 among the high-throughput methods, 10x Chromium (v2) showed the best quality for both the number of cell types identified and the average AUCs across cell types, followed by Drop-seq and 10x Chromium (v3), with Seq-Well and inDrops not identifying two cell types (Fig. 5c). In PBMC2, 10x Chromium (v2) and inDrops performed well, identifying all of the cell types (Fig. 5c). For Seq-Well PBMC2, the poor performance was strongly influenced by the low number of cells recovered in the experiment (Fig. 5c).

Similar to PBMCs, the mouse cortex has well-defined cell types, including excitatory and inhibitory neurons, astrocytes, oligodendrocytes, oligodendrocyte progenitor cells, microglia, endothelial cells and pericytes³¹. In both experiments for all of the methods, apart from sci-RNA-seq, we identified all of these cell types, except pericytes, a rare cell type only found in DroNc-seq Cortex1 (Fig. 6 and Extended Data Fig. 9). In the sci-RNA-seq datasets, we also could not find oligodendrocyte progenitor cells and microglia (Fig. 6). In the AUC analysis, Smart-seq2, 10x Chromium (v2) and DroNc-seq all had high AUCs, though their relative ability to detect the expected cells varied by cell type (Fig. 6c). Notably, even the small number of cells in the Smart-seq2 datasets (295 and 349) sufficed to find these cell types, in contrast to the PBMC datasets (Fig. 5). In the sci-RNA-seq datasets, we could not confidently assign cell types to some clusters of cells (7% and 4% of cells; Fig. 6a,b).

Pooled data analysis across methods enhances biological signal and consistency. Two general reasons may underlie the failure to detect certain cell types: (1) libraries did not contain cDNAs from these cell types due to experimental issues; or (2) data quality from these cells may not have been sufficient to identify them, given this depth of sequencing and number of cells. To distinguish between these possibilities, we combined for each PBMC experiment all of the sampled data together using Harmony³², re-clustered the cells (Extended Data Fig. 10a) and repeated our analysis. Following this analysis, all cell types were detected in each library, supporting the second possibility and showing the power of accruing data across methods (Extended Data Fig. 10b-d). Moreover, we determined in which cell type these missing cell types were originally (mis) assigned (Extended Data Fig. 10c,d). Although most of the combined and individual cell type assignments agree, some cell types seemed to be harder to distinguish. For example, in several libraries, such as Smart-seq2 and CEL-Seq2, the undetected dendritic cells were grouped with the CD14⁺ or CD16⁺ monocytes (Extended Data Fig. 10c,d). Overall, 10x Chromium (v2) was the most consistent between the combined- and individual-level clustering, followed closely by 10x Chromium (v3), and others having fairly high but variable levels of consistency. By contrast with the datasets with lower quality or numbers of cells, the AUC scores for cell clusters in the 10x Chromium datasets did not consistently improve when all of the datasets were used with the Harmony algorithm (Supplementary Fig. 3).

To check the cell types assigned by the combined analysis, we examined the cells assigned to cell types missing in our original analysis of each library separately. In most cases (20 of 25), we found that these cells could be assigned the same identity using our original AUC method (see Methods), with some exceptions for rare cell types with only ~1–2% of cells in a cluster (Supplementary Table 6). Thus, the failure to identify all of the relevant cell types was due, at least in part, to data quality issues, such as reads that could not



be used in the analysis (Supplementary Table 4), with the possible exception of two rarer cell types in our datasets, platelets and plasmacytoid dendritic cells, which may not have been present in some datasets. Another possible explanation is that low-quality cells were included that prevented identification of distinct cell types—this points to the difficulty in finding an optimal filtering threshold as well.

Similar analyses were done for mouse cortex nuclei (Supplementary Note 6).

Comparison of scumi with standard computational pipelines.

Although we used the scumi computational pipeline in this study to analyze each method's datasets in as similar a manner as possible, we also processed each of the datasets with its original method-specific pipeline for comparative purposes and found generally similar results (Supplementary Note 7).

Discussion

In this study, we systematically benchmarked seven methods across three major categories—plate-based, bead-based and combinatorial index-based methods—and summarized their relative merits (Supplementary Table 7). Our results were generally consistent in their ranking of the methods for sensitivity (Figs. 2–4), reproducibility (Supplementary Fig. 4), technical precision (Extended Data Fig. 7) and capturing biological information about cell types (Figs. 5 and 6, Extended Data Fig. 10 and Supplementary Fig. 5). Having a lower fraction of reads aligned to exons (Extended Data Fig. 3) could explain lower performance metrics in our comparisons, for example, in the Seq-Well PBMC datasets (Figs. 3 and 5). One limitation in our study is that our samples were not appropriate for pseudotime analysis³³.

All of the methods were able to generate useful data, but overall we found that 10x Chromium had the strongest consistent performance—similar to a more limited comparison of high-throughput methods²⁰. In our limited testing of 10x Chromium (v3), it had higher sensitivity (Fig. 3), but we did not detect improved cell type identification (Fig. 5), and had a higher fraction of reads aligned to mitochondrial genes (Supplementary Fig. 6). sci-RNA-seq, which has the ability to scale to much larger numbers of single cells¹³, may require optimization for use with some samples, such as PBMCs. We used the original version with two rounds of indexing¹². Moreover, its performance with cortex nuclei was not ideal as it could not assign an identity to some cells and did not detect all of the cell types present (Fig. 6 and Supplementary Fig. 5). For the low-throughput methods, Smart-seq2 and CEL-Seq2 performed similarly without a consistent pattern for which was better (Figs. 2–5). For studies that require the highest sensitivity, these two methods are clearly better than the high-throughput methods (Figs. 2–4), as shown previously¹⁶. Smart-seq2 has inherent advantages for genetic variant detection and studying RNA splicing isoforms because its sequencing is not limited to the 3' end of genes—along with the disadvantage of lacking UMIs. Note, however, that in CEL-Seq2 we cannot rule out the issue of contaminating reads from other cells (Extended Data Fig. 6)³⁴ and speculate that there is a step after cDNAs are pooled from different cells in which cell barcodes from one cell are switched with those from a different cell.

Looking beyond performance, we compared the time and reagent costs for each method as performed in this study (Supplementary Table 8). Drop-seq, Seq-Well and inDrops had the lowest costs and Smart-seq2 was the most expensive, primarily because there is no pooling during library preparation. Many of the methods, particularly sci-RNA-seq, would be more cost effective with larger numbers of single cells or nuclei¹³. The 10x Chromium method required the least time and Smart-seq2, CEL-Seq2 and inDrops took the most time. We did not utilize automation, but it could decrease hands-on time and affect cost.

Analyzing single nuclei rather than single cells is an important strategy, which addresses tissues that cannot be readily dissociated into a single-cell suspension (such as brain, skeletal muscle or adipose) and frozen samples, and also minimizes the alteration of gene expression that may be caused by dissociation^{35,36}. As in previous studies^{24,37}, we found that single-nucleus RNA-seq generally performed well for sensitivity (Fig. 4) and classification of cell types (Fig. 6). Even with the inclusion of intron-aligning reads in our analysis, a higher fraction of reads for 10x Chromium, and to a lesser extent for DroNc-seq, could not be analyzed because of the absence of a poly(T) sequence or aligning in an antisense orientation (Extended Data Fig. 3).

Our study, including the scumi pipeline (for its relative advantages, see Supplementary Note 8), data and approaches, will be a resource for future research in many fields where scRNA-seq methods are applied, and provides important guidance. First, using a coherently and reproducibly collected set of data, spanning three sample types, it provides direct guidance on key methods by a rich set of parameters and considerations—from technical to biological. It spans key and popular methods, including the comparison of single-nucleus RNA-seq methods. We chose two very different tissue types to make our framework and conclusions more generalizable, and show this by analysis of published datasets for three other tissues (Supplementary Fig. 7). Our study would allow future studies to reduce the number of protocols that need to be tested for new tissues. Second, the results presented here for each method could be used to further optimize and improve existing scRNA-seq methods. Third, our use of representative and easily accessible sample types should allow future studies, particularly those introducing new or improved methods, to make direct comparisons with this benchmark study. Indeed, all datasets were collected in a manner that allows open sharing, including the human PBMC data. Finally, we expect that our datasets will be valuable for computational method developers to benchmark algorithms and build pipelines for efforts such as HCA, the BRAIN Initiative, the Cancer Moonshot Human Tumor Atlas Network and other efforts to map cells in disease.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-020-0465-8>.

Received: 7 May 2019; Accepted: 24 February 2020;

Published online: 06 April 2020

References

1. Haque, A., Engel, J., Teichmann, S. A. & Lonnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **9**, 75 (2017).
2. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
3. Wu, A. R., Wang, J., Streets, A. M. & Huang, Y. Single-cell transcriptional analysis. *Annu. Rev. Anal. Chem. (Palo Alto Calif.)* **10**, 439–462 (2017).
4. Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
5. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
6. Hashimshony, T. et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77 (2016).
7. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
8. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
9. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
10. Gierahn, T. M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).

11. Han, X. et al. Mapping the mouse cell atlas by Microwell-seq. *Cell* **172**, 1091–1107 (2018).
12. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
13. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
14. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
15. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).
16. Ziegenhain, C. et al. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65**, 631–643 (2017).
17. Dueck, H. R. et al. Assessing characteristics of RNA amplification methods for single cell RNA sequencing. *BMC Genomics* **17**, 966 (2016).
18. Svensson, V. et al. Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14**, 381–387 (2017).
19. Bhargava, V., Head, S. R., Ordoukhianian, P., Mercola, M. & Subramaniam, S. Technical variations in low-input RNA-seq methodologies. *Sci. Rep.* **4**, 3678 (2014).
20. Zhang, X. et al. Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. *Mol. Cell* **73**, 130–142 (2019).
21. Wang, Y. J. et al. Comparative analysis of commercially available single-cell RNA sequencing platforms for their performance in complex human tissues. Preprint at *bioRxiv* 541433 (2019).
22. Habib, N. et al. Div-Seq: single-nucleus RNA-seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925–928 (2016).
23. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
24. Bakken, T. E. et al. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS ONE* **13**, e0209648 (2018).
25. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
26. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
27. Wallrapp, A. et al. The neuropeptide NMU amplifies ILC2-driven allergic lung inflammation. *Nature* **549**, 351–356 (2017).
28. Tilgner, H. et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).
29. Grun, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nat. Methods* **11**, 637–640 (2014).
30. Wagner, F., Yan, Y. & Yanai, I. K-nearest neighbor smoothing for high-throughput single-cell RNA-seq data. Preprint at *bioRxiv* 217737 (2018).
31. Zeisel, A. et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
32. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
33. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).
34. Arazi, A. et al. The immune cell landscape in kidneys of patients with lupus nephritis. *Nat. Immunol.* **20**, 902–914 (2019).
35. Lacar, B. et al. Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat. Commun.* **7**, 11022 (2016).
36. van den Brink, S. C. et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).
37. Lake, B. B. et al. A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **7**, 6031 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Single-cell or -nucleus experimental design. We performed two experiments with each single-cell method for the mixed cell lines and PBMCs, except as noted. To generate data for Seq-Well, we performed a second PBMC1 experiment on a different day with an aliquot identical to the one used in the main PBMC1 experiment to obtain a Seq-Well dataset with sufficient cells profiled for PBMCs. Similarly, we performed a third PBMC1 experiment with 10x Chromium (v2; designated as ‘B’) and (v3) on a different day. In addition, we performed two experiments with four methods for the mouse whole cortex nuclei. In all cases, each laboratory method was started at the same time by different researchers, so that the results would be directly comparable without any confounding due to the time cells or nuclei waited to start the experiment.

Additional experimental details are in the Supplementary Methods.

Computational methods. We present key steps in the analysis with the scumi pipeline here and other analyses can be found in the Supplementary Methods.

Annotating each cDNA read with its cell barcode and UMI. Different scRNA-seq protocols produce reads with different structures, especially the reads consisting of cell barcodes and/or UMI sequences. To address this issue, we used our scumi pipeline, which uses regular expressions (text strings defining search patterns) to extract the cell barcodes and UMI sequences from different FASTQ files and put them in the header of their corresponding cDNA reads. We started with the FASTQ files generated after de-multiplexing the BCL files from Illumina sequencers. For a typical 3’ tag-based scRNA-seq experiment, a cDNA sequence fragment is in one read of a FASTQ file, and its corresponding cell barcode sequences and UMI sequences are in a paired read of a separate FASTQ file. For example, for the reads generated from the Drop-seq platform⁸, the cDNA reads are in read 2 and the cell barcodes (base 1 to base 12) and UMIs (base 13 to base 20) are in read 1. Details about the location of cell barcodes and UMIs can be found in Supplementary Table 9.

The scumi pipeline also corrects for sequence errors in the indices used by bead-based methods. We have implemented code similar to the standard Drop-seq pipeline that overcomes the problems observed for some batches of Drop-seq beads, in which up to 20% of the cell barcodes have errors in the last base (base 12), mostly because the beads (encoding the cell barcodes) only synthesized 11 (or fewer) bases⁸. In such cases, base 12 of the cell barcode is actually the first base of the UMI sequence, and the last base of the UMI sequence is from the poly(T) sequence. This bead synthesizing error can be detected by calculating the frequencies of T bases in the UMI sequences. The scumi pipeline first detects possible erroneous cell barcodes and then merges these cell barcodes that are the same in their first 11 bases but differ in the 12th base. If more than one base of the UMI sequences (with the same cell barcode) had a high-frequency of T bases (more than 80%), these cell barcodes were removed from further analysis.

Mapping reads to a reference genome. We aligned the merged FASTQ files (each cDNA read with its cell barcode and UMI annotations) to a reference genome using STAR³⁸ v.2.6.1a, except for Smart-seq2 and CEL-Seq2. For those libraries, we used HISAT2 (ref. ³⁹) v.2.0.5 as it is better suited than STAR to paired-end read data such as Smart-seq2 because of the way it handles read pairs that do not both align to the same region of the genome—leading to more aligned reads and more detected genes per cell. Notably, HCA has adopted this aligner for its Smart-seq2 pipeline (<https://staging.data.humancellatlas.org/learn/userguides/data-processing-pipelines/smart-seq2-workflow/>). We also used it for CEL-Seq2 to facilitate better performance comparisons between the two low-throughput methods. For mixture data, we used the STAR reference available in the hg19 and mm10 v.2.1.0 Cell Ranger reference. For PBMC data, we used the STAR reference available in the GRCh38 v.1.2.0 Cell Ranger reference. For cortex data, we used the STAR reference available in the mm10 v.1.2.0 Cell Ranger reference. We downloaded Cell Ranger reference data from <https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/references>. For each sample type, we also generated a HISAT2 reference with the associated GTF and FASTA files.

Annotating each alignment with a gene tag. We use featureCounts⁴⁰ from the Subread package, v.1.6.2, to add a gene tag to each alignment. To count reads overlapping with introns for single-nucleus RNA-seq data, we used a two-step approach to first count the reads overlapping with exons. In the second step, the reads not overlapping with exons were recounted if they overlapped with introns. We only included reads aligning in the sense orientation with the genome annotation, except for Smart-seq2, which does not generate strand-specific data.

Counting transcripts of each gene in each cell. For the UMI-based methods, we used scumi to generate a cell by gene UMI count matrix. We included a multi-mapped read if all of its alignments overlapped with a single gene, similar to the Cell Ranger pipeline⁹. We collapsed UMIs in reads from the same gene from the same cell based on a Hamming distance of 1. To prevent over-collapsing UMIs⁴¹, we did not collapse two UMIs—in the same gene in the same cell—if they each had support from more than five reads. For Smart-seq2, we used a similar procedure to generate the count matrix used for the sensitivity and technical precision metrics, except we created a cell by gene read count matrix. For Mixture data, we did not include multi-mapped reads in subsequent analyses, and instead used

featureCounts to count uniquely mapped reads. For Smart-seq2, this resulted in each read of a paired-end read being counted separately in the Mixture data (though not in the other experiments), enabling us to count reads in pairs in which each aligned to a different gene. For clustering Smart-seq2 data and downstream analysis, we used RSEM⁴² v.1.3.0 to generate a cell by gene transcripts per million matrix, which was used instead of the UMI count matrix. We generated the RSEM reference using the FASTA and GTF files used for creating the STAR and HISAT2 references (see Mapping reads to a reference genome). When generating the RSEM reference for cortex data, we modified the GTF to include one unspliced transcript per gene that included all introns and exons in that gene. This allowed us to count reads that mapped to introns.

Selecting the number of cells. For an scRNA-seq experiment, we have a rough estimate of the number of cells, N , that can be recovered. A simple yet robust empirical method (used by Cell Ranger of the 10x Chromium pipeline) for cell barcode selection is to first estimate the library size, m (in reads or UMIs), by the 99th percentile of the top N cell barcodes in terms of the number of reads (or UMIs). The cell barcodes with reads (or UMIs) greater than $0.1 \times m$ are considered ‘cells’.

For the cell line mixture experiments, we used different filtering approaches depending on the dataset. For 10x Chromium, Drop-seq, Seq-Well, inDrops and sci-RNA-seq, we used this empirical rule for cell barcode selection. For Smart-seq2 and CEL-Seq2, we had a better estimation of the number of cells as we sorted individual cells into wells. We used a mixture of two Student’s t distributions⁴³ to model the read or UMI (\log_{10} transformed) count distributions of each cell, and removed the cells that were likely from the mixture component with fewer reads or UMIs (posterior probability ≥ 0.5). The parameters of the Student’s t mixture model were estimated by maximizing the posterior distribution using the expectation–maximization algorithm. For sci-RNA-seq and inDrops, the empirical rule tended to select low-quality cell barcodes. We therefore used this mixture model on the cell barcodes selected by the empirical rule to further filter out likely low-quality cells.

For all of the high-throughput PBMC datasets, we extracted two times the number of expected cell barcodes for each method, choosing the cells with the most reads. We removed cells with a high fraction of reads aligning to mitochondrial genes (names starting with ‘mt-’ for mouse and ‘MT-’ for human)—greater than 75th percentile + 3 × IQR of the mitochondrial ratios across the top returned cell barcodes, where IQR stands for interquartile range. For each cell, its UMI counts were divided by the total number of UMIs from that cell and then scaled by multiplying by 10,000 to get transcripts per 10,000 (TP10K). We then added 1 to these transcripts per 10,000 and log transformed by the natural log. We then performed principal-component analysis using all genes, did clustering analysis (Louvain clustering^{44,45} of the k -nearest neighbor (k -NN) graph built from the first 50 principal components of each single-cell dataset with parameter $k = 30$ and a resolution parameter used for Louvain optimization of 1.0, implemented in the Seurat package⁴⁶ v.2.3.4 (see Parameter selection for clustering analysis section in Supplementary Methods for more details)), followed by differential gene expression analysis with the FindAllMarkers command in Seurat to find cluster-specific (up-regulated) marker genes. To filter out clusters of cells likely derived from low-quality cells or empty droplets, we removed clusters with insufficient marker genes, as follows. First, we identified marker genes for each cluster as genes expressed in $\geq 25\%$ of cells in that cluster and with false discovery rate corrected P values < 0.01 (significantly highly expressed in the cluster compared with the cells not in that cluster). Second, we excluded ribosomal protein coding genes, MALAT1 and genes starting with MTRNR, as they could be erroneously identified as marker genes after normalization and scaling because, for cells with a small number of UMIs, the UMIs of highly expressed genes will be weighted more than those from cells with a large number of UMIs based on the scaling formula $x_j / \sum x_j$, where x_j is the expression of gene j of a cell. Third, we only kept the clusters in which $> 70\%$ of the top 15 marker genes (or 10 out of all markers genes in a cluster that had < 15 marker genes) were not mitochondrial protein coding genes as high expression of mitochondrial genes can indicate stressed cells or empty droplets. This process was repeated twice.

We used a modified strategy for Smart-seq2 and CEL-Seq2 because there were fewer cells to cluster, which potentially could have led to the low-quality cells not forming distinct clusters. The assumptions for these cell selections were that (1) there were enough low-quality cells to form distinct clusters; and (2) the clustering algorithm did not split high-quality cells of the same cell type into many distinct clusters because this could have led to some subclusters having too few marker genes. We therefore set $k = 5$ (the number of neighbors in building the k -NN graph) to detect small clusters with a low resolution parameter of 0.5 to prevent splitting cells of the same type into many clusters. We also only used the top 25 principal components as we did not expect to identify as many cell types from a smaller number of cells.

For each cortex dataset, the number of UMIs per cell barcode across all cell barcodes returned by scumi followed a bimodal distribution, with some cell barcodes having few UMIs and others having many. We therefore first used a mixture of two Student’s t distributions to fit the UMI count distribution across all of the returned cell barcodes. We considered the mixture component with a larger mean as the high-quality cell barcode component. We removed from further analyses the cell barcodes with posterior probabilities < 0.5 from the high-quality component. The cell barcodes passing this mixture model filtering (with posterior

probabilities from the high-quality cell barcode component ≥ 0.5) were then further filtered using the approach previously described for the PBMC data. We used the top 25 principal components and set the number of nearest neighbors k to 10 in clustering analyses (the choice of parameters was to help recover rarer cell types with lower numbers of UMIs per cell, such as the oligodendrocyte precursors). To prevent splitting big clusters due to the small k , we lowered the resolution to 0.8 instead of 1.0 used for PBMC high-throughput methods. For these nuclei libraries, the mitochondrial ratios were very low compared with those from cells, so that it was less likely to see low-quality nuclei with mitochondrial protein coding genes as the top marker genes. Therefore, in addition to using mitochondrial protein coding genes to remove poor-quality nuclei, we removed cluster-specific marker genes that were also expressed in $>70\%$ of the nuclei from the other clusters in a given library. We also used the top 20 marker genes in each cluster for filtering. For Smart-seq2 data, we only used the mixture model for cell selection as the remaining clusters could be assigned known mouse cortex cell types and applying the cluster-based filtering removed many cells that could be easily identified as known cell types.

Sampling reads. To correct for differences in sequencing depth between methods, we used seqtk v.1.0 (<https://github.com/lh3/seqtk>) to sample the sequencing data, so that for each method we could analyze nearly the same average number of reads per cell. Low- and high-throughput methods were sampled separately. For a given experiment, we first decided on the average number of reads per cell to sample. We usually set this equal to the lowest average number of reads per cell for a method in that experiment, except for Seq-Well PBMC1, which had a lower average number of reads per cell, so that we chose the library with the next lowest average number of reads per cell. For each library in a given experiment, we then derived a sampling ratio by dividing this sampling target by the original average number of reads per cell in that library. We sampled the FASTQ file for each library with the ‘seqtk sample’ command, using the sampling ratio calculated above and the random seed set to 100 (after combining FASTQ files from different sequencing runs). Although we aimed to have the exact same average number of reads per cell for each library, there were some small deviations from this in practice because the number of cells we identified in each library was not always the same before and after sampling, as well as due to the random nature of the sampling. Details of the sampling for each experiment are in Supplementary Table 2.

Automatically assigning cell types to clusters. We followed common practices for scRNA-seq data clustering. Specifically, cells were divided into nonoverlapping clusters by using the Louvain community detection algorithm^{44,45}. For each cell from a dataset, we computed its k -NNs in that dataset, and then built a directed k -NN graph using all of the cells from that dataset. This directed k -NN graph was further converted to an undirected weighed graph by using shared neighbors. The Louvain algorithm was used to partition the undirected weighted k -NN graph into nonoverlapping clusters.

We used marker genes for each cell type to compute a score for each cell and automatically assign cell types to clusters. Both human PBMC and mouse cortex have well-annotated cell types and marker genes for each cell type^{31,47,48}. We generated lists of marker genes for each tissue with manual curation (Supplementary Tables 10 and 11). The score of cell i for cell type m is a normalized version of the percentage of total counts from marker genes from cell type m . Assuming that there were N_m marker genes for cell type m , we considered these N_m genes combined as a ‘meta-gene’ with counts $f_i^m = \sum x_{ij}$ in cell i , where x_{ij} was the expression (UMI count) of marker gene j for cell type m . The meta-gene relative expression in cell i was its count divided by the total count C_i in cell i . We obtained the score of cell i for cell type m as $s_i^m = \log(f_i^m/C_i \times 10^4 + 1)$.

Based on the scores, we assigned cell types to clusters using the AUCs. Stated another way, for a given cluster and a given cell type c , a cell i in that cluster is a true positive if the score s_i^c is above a given threshold and a false negative otherwise. On the other hand, a cell not in that cluster is a false positive if it has a score above the threshold and a true negative otherwise. A receiver operating characteristic curve plots the true positive rate against the false positive rate at different score thresholds. The AUC is 1.0 for perfectly assigning a cell type to a cluster (we can find a threshold score perfectly separating a cluster from the rest), and around 0.5 for randomly assigning a score to a cell. Specifically, for each cluster, the cell type with the maximum AUC was assigned to that cluster. As the same type of cells can be split into several clusters, after initial assignment of cell types to clusters, we recomputed the AUC of a cluster for a cell type by excluding other clusters of cells that were assigned to that cell type. This process was repeated until there were no changes in the cluster assignment. We then calculated the AUC for a cell type by merging the cluster of cells that were assigned to that cell type.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

RNA-seq data generated in this project are available from the Gene Expression Omnibus with accession number [GSE132044](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132044) and the Single Cell Portal (https://portals.broadinstitute.org/single_cell). Source data for Figs. 2–6 are presented with the paper.

Code availability

The scumi Python package is available freely from bitbucket repository at <https://bitbucket.org/jerry00/scumi-dev/src/master/> and as Supplementary code. The R scripts (used to assign cell types to clusters based on a set of marker genes, for parameter selecting for clustering analysis and for filtering low-quality cells) are available from bitbucket repository at <https://bitbucket.org/jerry00/scumi-dev/src/master/>.

References

38. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
39. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
40. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
41. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. zUMIs—a fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience* **7**, giy059 (2018).
42. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
43. Ding, J. et al. Systematic analysis of somatic mutations impacting gene expression in 12 tumour types. *Nat. Commun.* **6**, 8554 (2015).
44. Levine, J. H. et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
45. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory E* **2008**, P10008 (2008).
46. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
47. Tasic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
48. Palmer, C., Diehn, M., Alizadeh, A. A. & Brown, P. O. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics* **7**, 115 (2006).

Acknowledgements

We especially thank M. Chatterjee, A. Ratner and S. Boswell of the Single Cell Core at Harvard Medical School for performing the inDrops experiments. We are grateful to A. Neumann, J. Lee, D. Dionne and N. Sharif for assistance with project coordination; A. Klein for helpful discussions and suggestions; R. Kirchner for advice on inDrops data analysis; D. Leib for advice on CEL-Seq2 data analysis; B. Li for advice on PBMC data analysis; K. Shekhar for precision analysis in cell line mixture data; M. Cuoco for sample transportation; Broad Flow Cytometry Facility for cell sorting; Broad Genomics Platform for sequencing; and L. Gaffney for assistance with figures. Work was supported by the Klarman Cell Observatory, the Manton Foundation and the BRAIN Initiative (grant no. 1U19 MH114821, A.R.). A.R. is an Investigator of the Howard Hughes Medical Institute. This publication is part of the Human Cell Atlas at www.humancellatlas.org/publications.

Author contributions

J.Z.L., A.K.S., O.R.-R. and A.R. conceived the research. X.A., C.C.H., N.D.M., T.K.H., M.H.W., T.B., L.T.N., J.Y.H.K., S.C. and S.L. performed the scRNA-seq experiments. X.A. and C.C.H. organized the sequencing. X.A. prepared the bulk RNA-seq libraries. J.D. created the scumi pipeline. J.D., S.K.S., A.-C.V., A.J.K. and J.Z.L. analyzed the data. M.S.K. contributed an optimized Smart-seq2 protocol. J.Y.H.K. prepared the cell lines. A.-C.V. and W.G. prepared the PBMCs. B.B. prepared the mouse cortex. J.Z.L., N.H., O.R.-R., A.K.S., A.-C.V. and A.R. provided supervisory guidance. J.D., X.A., S.K.S., C.C.H., T.K.H., M.H.W., T.B., J.Y.H.K., A.-C.V., A.R. and J.Z.L. wrote the paper. All authors assisted in editing the paper.

Competing interests

A.R. is a founder and equity holder in Celsius Therapeutics; an equity holder in Immunitas; and an SAB member of Syros Pharmaceuticals, Neogene Therapeutics and Thermo Fisher Scientific. A.K.S. is a founder of, and consultant for, Honeycomb Biotechnologies, Inc., which manufactures Seq-Well peripherals. A.K.S. and A.R. are also named inventors on patents filed by the Broad Institute related to either Drop-seq (A.R. and A.K.S.), DroNc-seq (A.R.) or Seq-Well (A.K.S.). The interests of A.K.S. and A.R. were reviewed and are subject to a management plan overseen by their institutions in accordance with their conflict of interest policies. The other authors declare no competing interests.

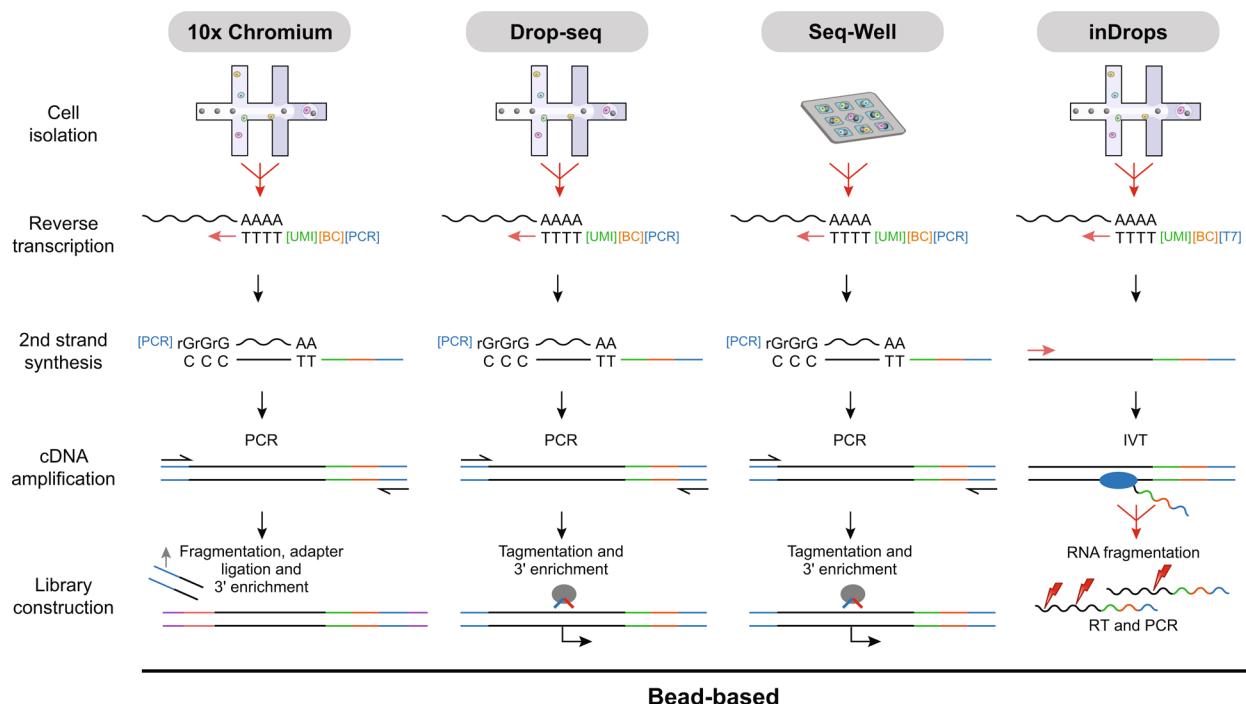
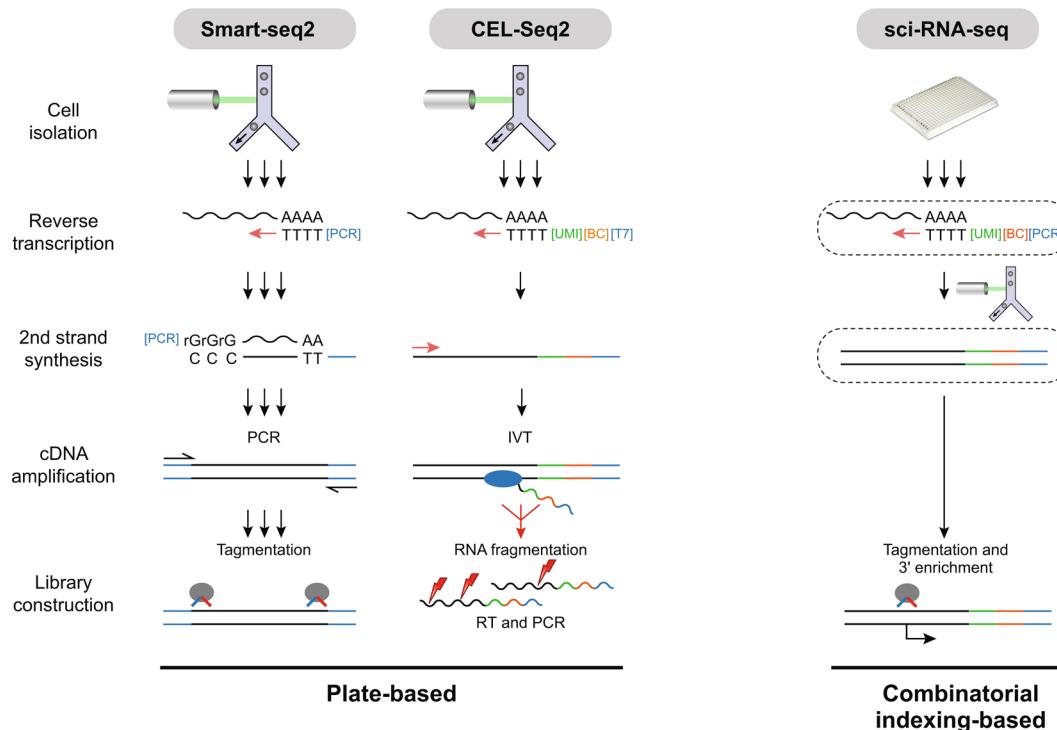
Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41587-020-0465-8>.

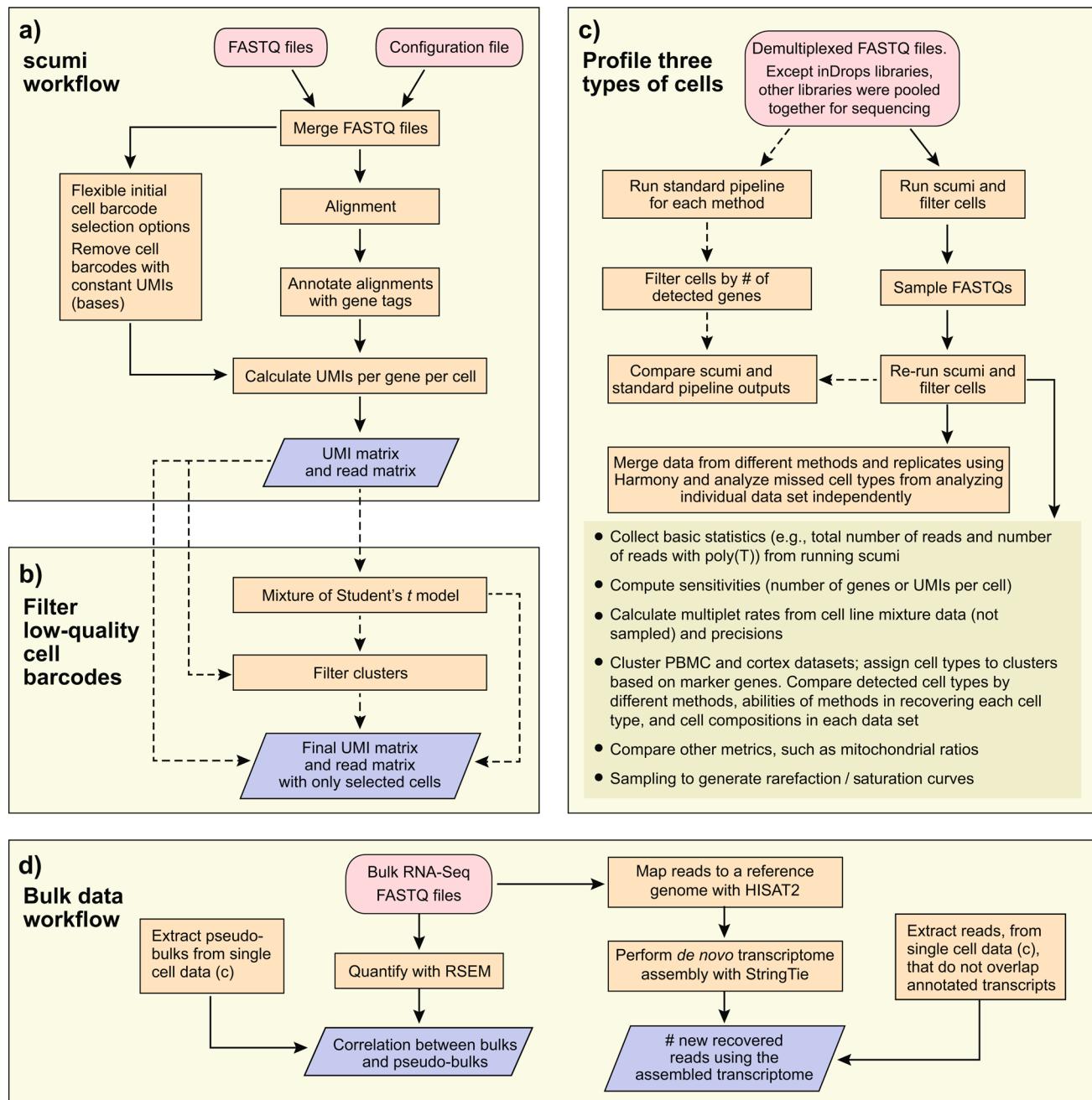
Supplementary information is available for this paper at <https://doi.org/10.1038/s41587-020-0465-8>.

Correspondence and requests for materials should be addressed to J.Z.L.

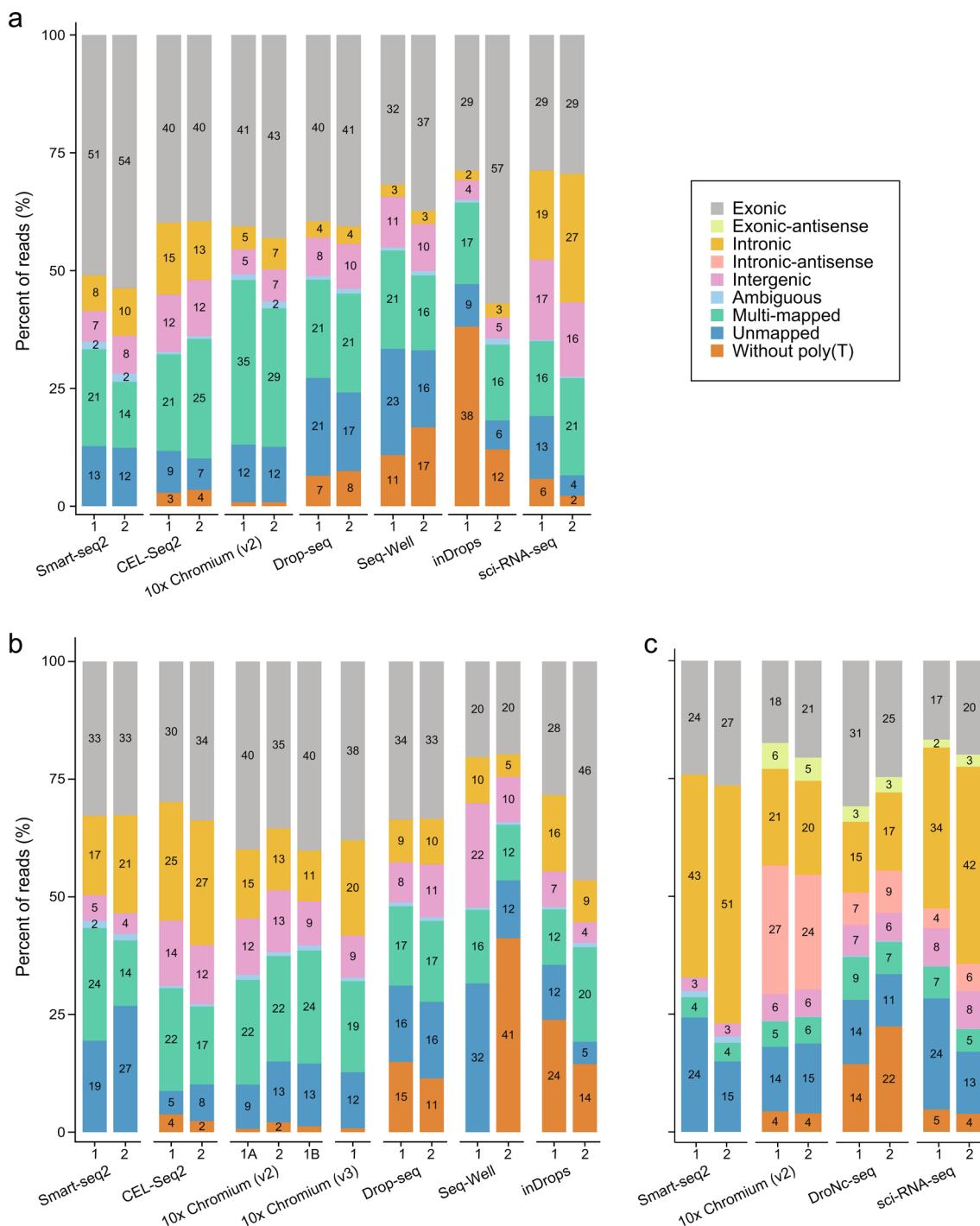
Reprints and permissions information is available at www.nature.com/reprints.



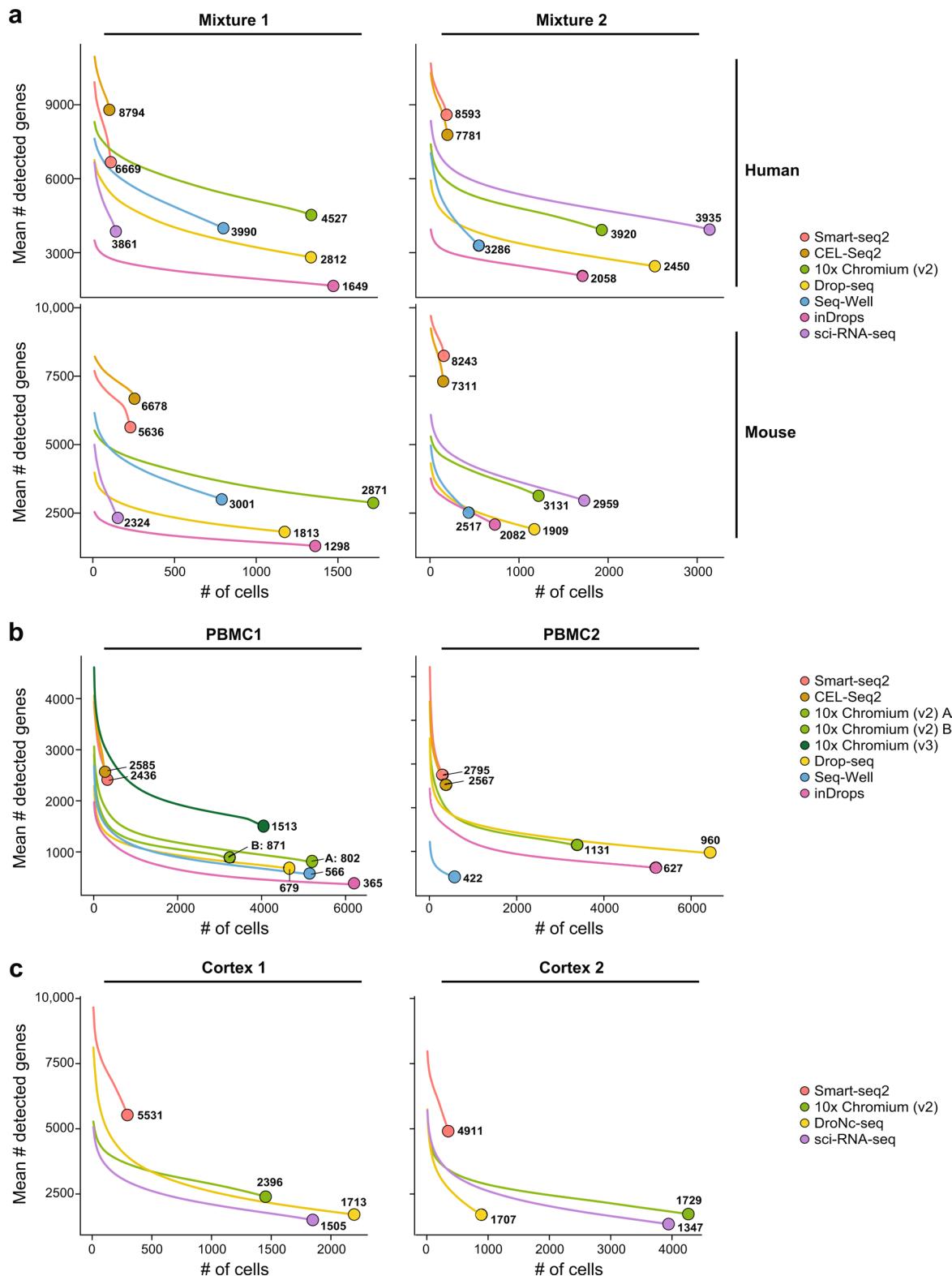
Extended Data Fig. 1 | Description of scRNA-seq methods evaluated. Salient details for seven protocols tested in this paper. Adapted from Figure 2 of ref. ¹⁶.



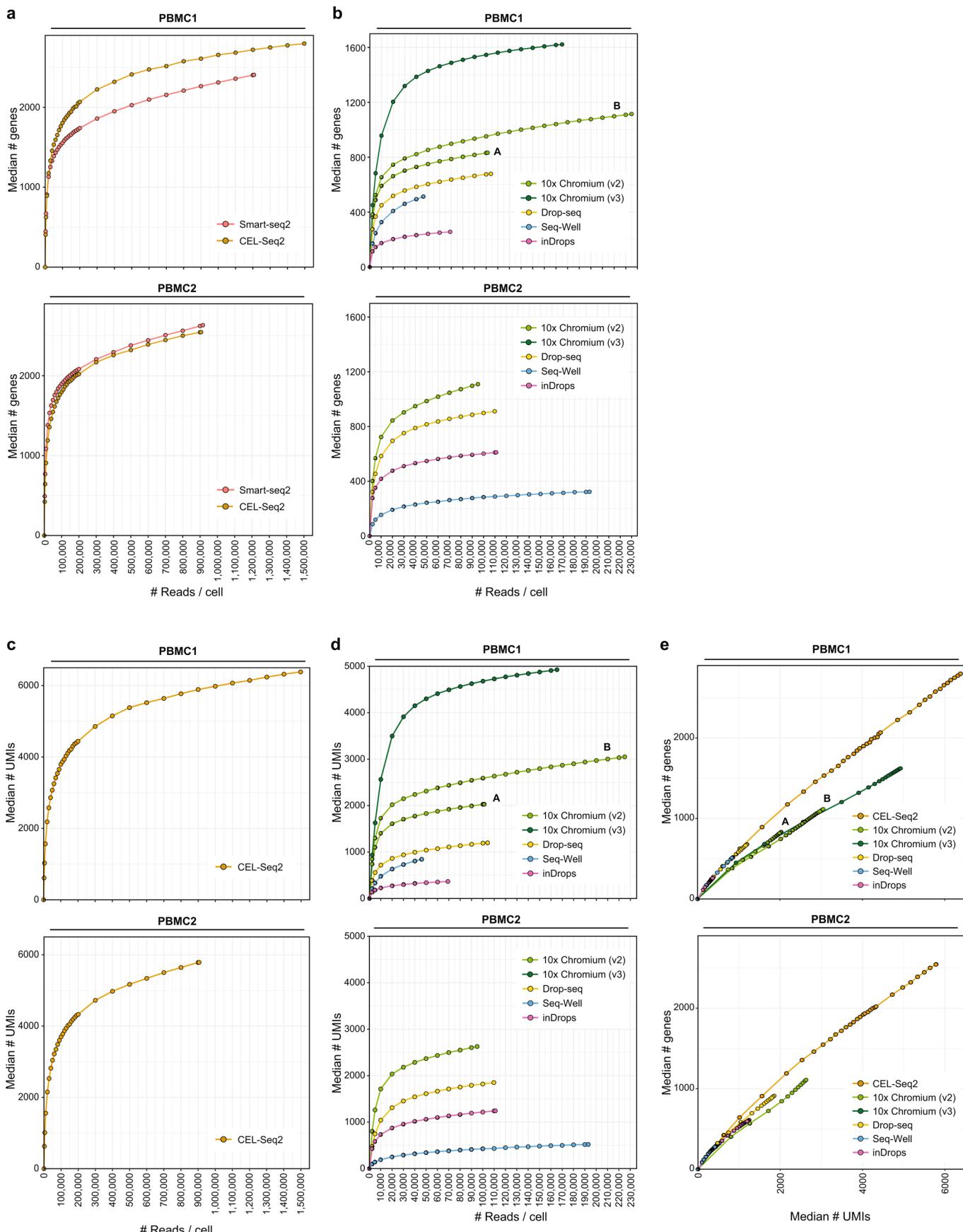
Extended Data Fig. 2 | Flowchart detailing computational analysis. **a**, scumi workflow, **b**, removing low quality cell barcodes, **c**, profiling samples, **d**, bulk data workflow.



Extended Data Fig. 3 | Characterization of genome alignments for sequence reads. **a**, Mixture, **b**, PBMCs, **c**, Cortex. For each pair of bar graphs, experiment 1 is on the left and experiment 2 is on the right. For Smart-seq2, there were no poly(T) reads due to the full transcriptome coverage and the library construction using transposase-based Nextera reagents to attach adapters to both ends of cDNA fragments. Reads were assigned in the following order: no poly(T), unmapped or multi-mapped, ambiguous (mapping to a single location that overlaps 2 or more genes), and then one of the remaining categories. Reads were assigned as antisense only for the cortex datasets (c). % of reads may not sum to 100 due to rounding and numbers not shown for fraction of reads in categories with <2%.

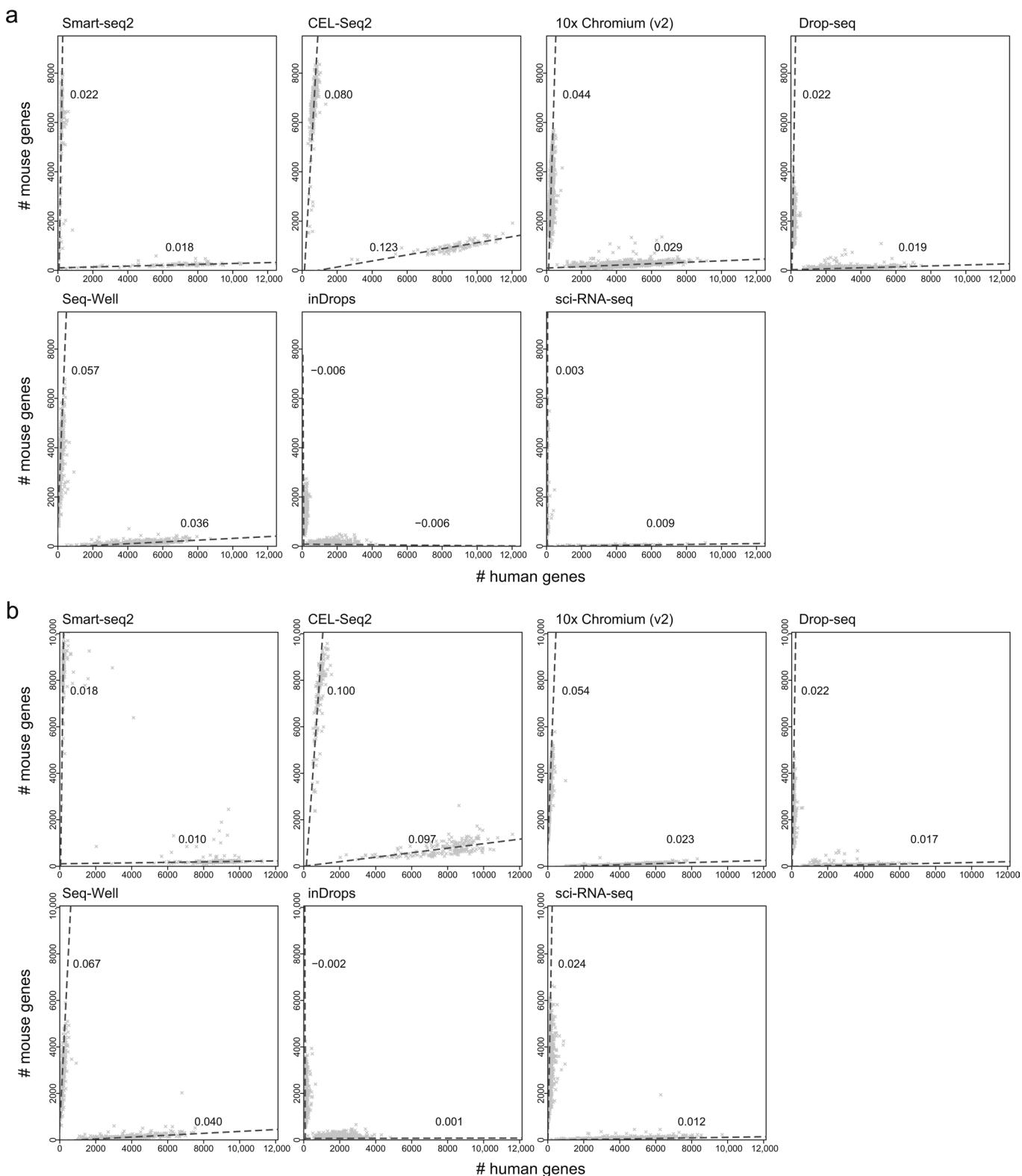


Extended Data Fig. 4 | Impact of number of cells on sensitivity. **a**, human cells and mouse cells from Mixture experiments. Multiplet cells are not shown in this plot. **b**, PBMC. **c**, cortex. The number of cells (x-axis) with a given mean number of genes detected (y-axis), when cells are ordered from highest (left) to lowest (right) total number of genes. The right most point at the end of each curve shows the average number of detected genes for the final selected number of cells in this study.

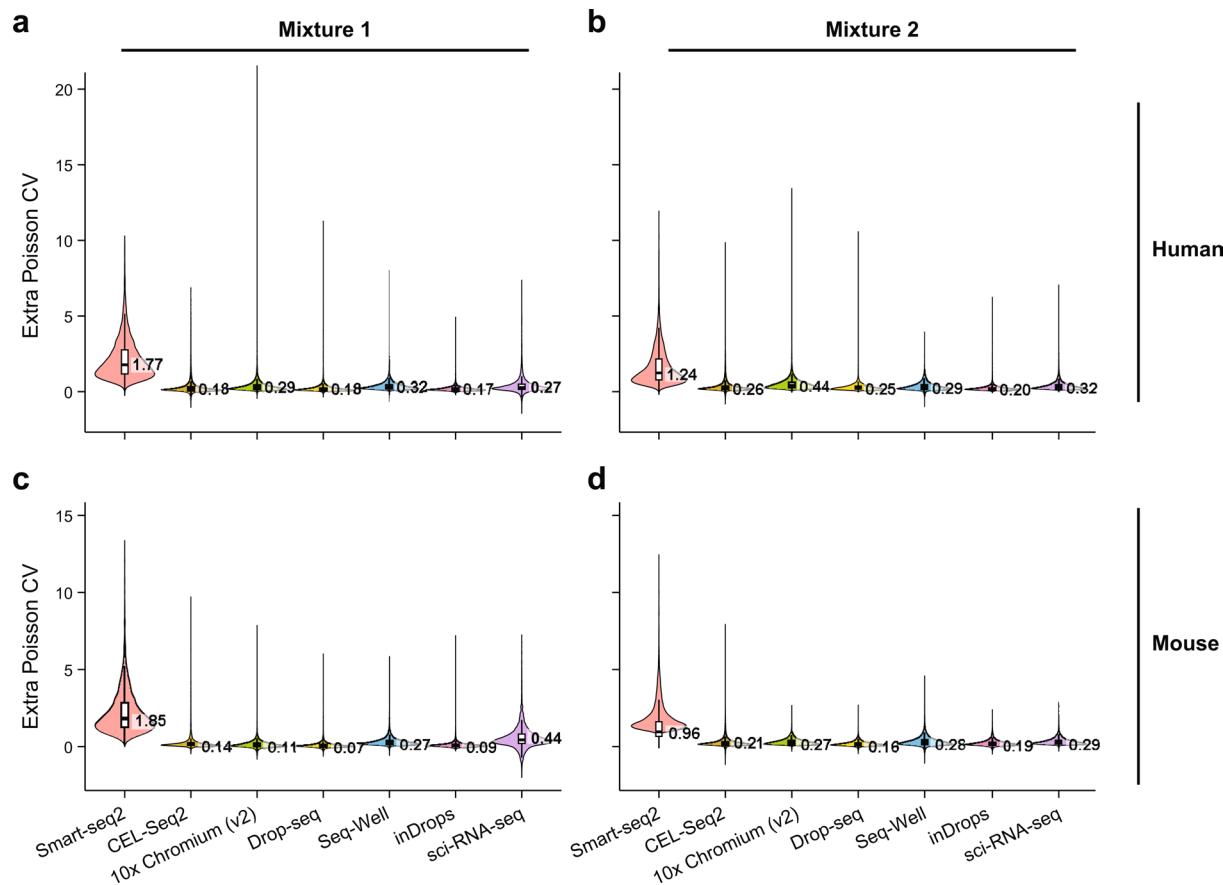


Extended Data Fig. 5 | See next page for caption.

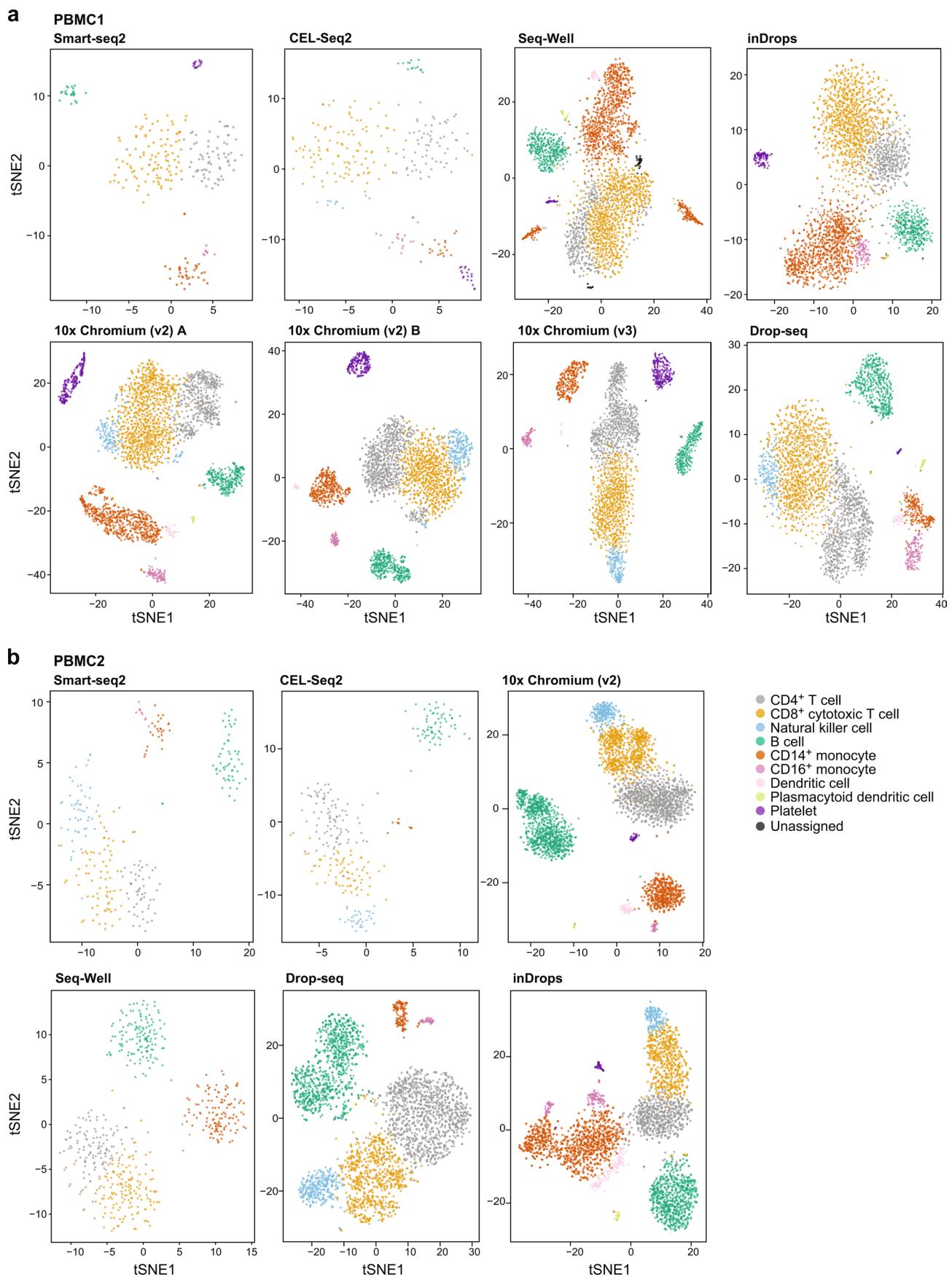
Extended Data Fig. 5 | Impact of sequencing depth on gene and UMI detection per cell in the PBMC datasets. **a-d**, The median number of genes (**a, b**, y-axis) and UMIs (**c, d**, y-axis) detected per cell at different sequencing depths (x-axis) for low-throughput (**a, c**) and high-throughput (**b, d**) methods from PBMC1 (left) and PBMC2 (right). Far right point of each curve: median number of detected genes per cell at full sequencing depth. **e**, Relation between median number of genes and UMIs per cell in PBMC1 (left) and PBMC2 (right). ($n=1$ biologically independent sample for each curve in each plot).



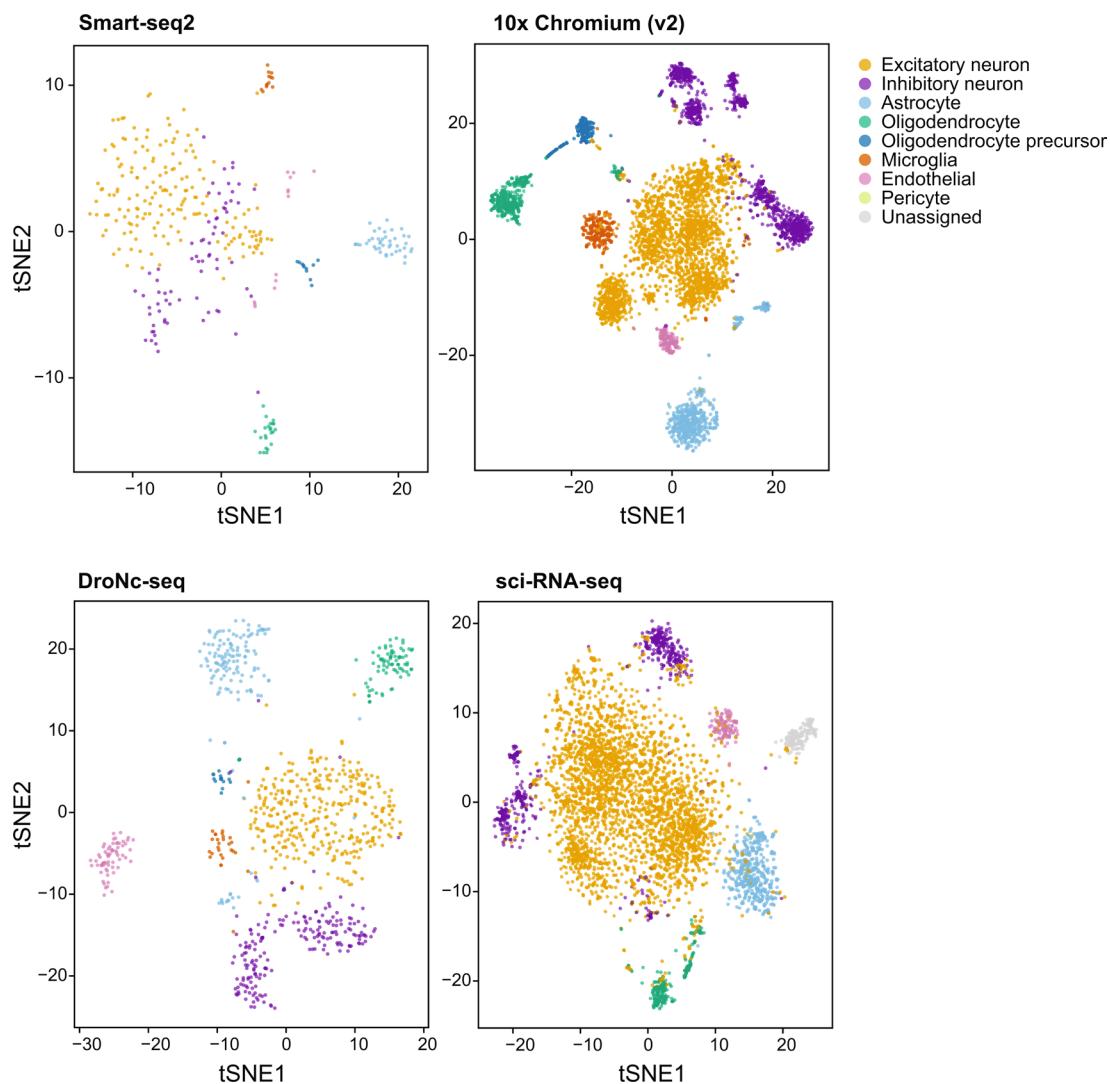
Extended Data Fig. 6 | Fraction of reads from each species in Mixture experiments. Fraction of UMIs (or reads for Smart-seq2) aligned to either mouse (y-axis) or human (x-axis) in each cell from the Mixture1 **a**, and Mixture2 **b**, experiments ($n=1$ biologically independent sample per panel). Each dot represents a cell. Dashed line and number: robust linear regression fitted line and its slope. Number of genes detected from the “wrong” species is higher in cells with more reads.



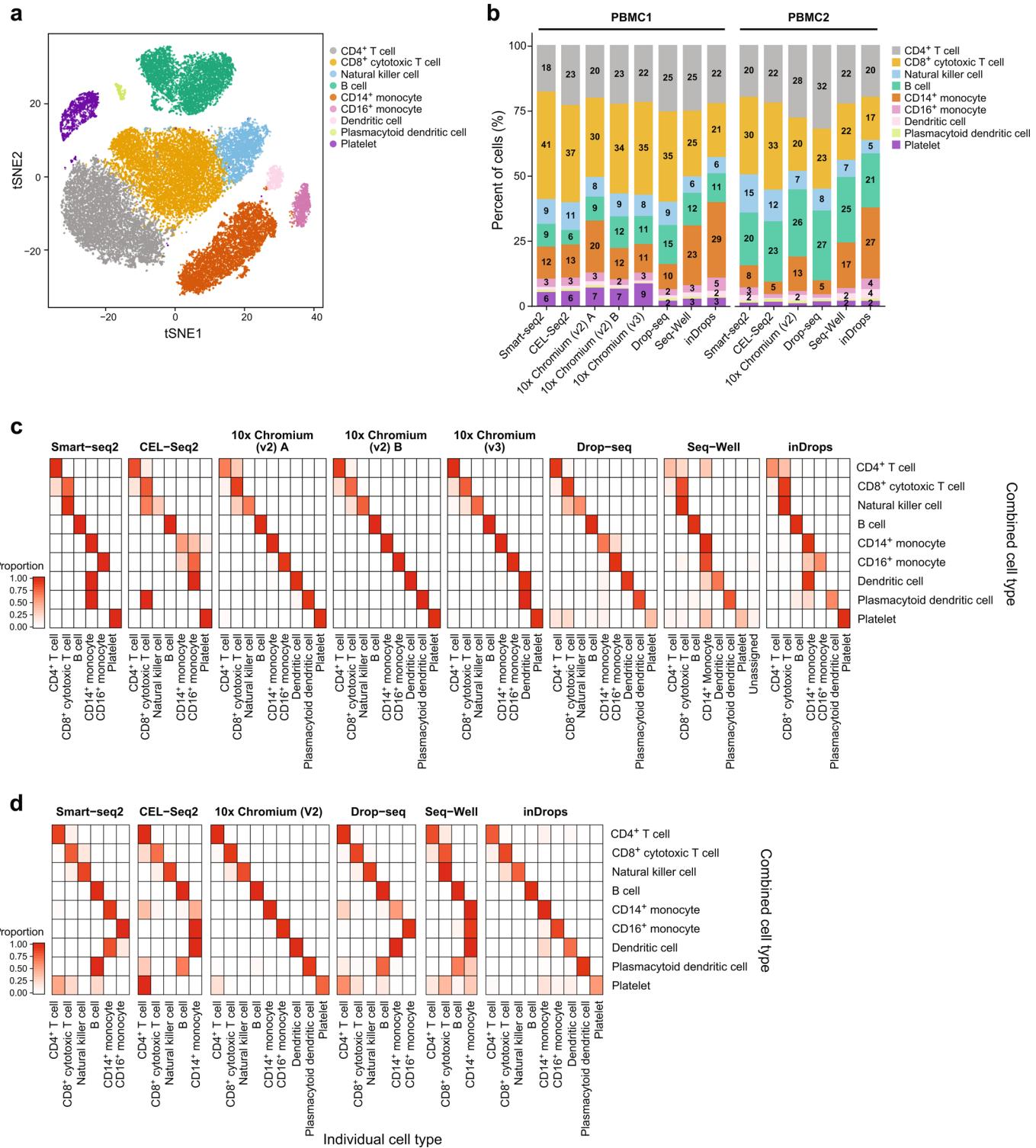
Extended Data Fig. 7 | Technical precision plots for mixture experiments. Distributions of the extra Poisson coefficients of variation (“Extra Poisson CV”, y-axis) from each method (x-axis). **a, b**, Human cells; **c, d**, mouse cells – from Mixture1 (left) and Mixture2 (right) ($n=1$ biologically independent sample per panel). Violin and box plot elements are defined as in Figure 2.



Extended Data Fig. 8 | Cell type analysis for each PBMC dataset. t-SNEs of single cell profiles (dots) from each method colored by cell type assignment from PBMC1 **a**, and PBMC2 **b**. ($n=1$ biologically independent sample per panel).



Extended Data Fig. 9 | Cell type analysis for Cortex2. t-SNEs of single nucleus profiles (dots) from each method colored by cell type assignment from Cortex2 ($n=1$ biologically independent sample).



Extended Data Fig. 10 | Cell type analysis of the combined PBMC datasets. **a**, t-SNE plot generated with Harmony clustering all PBMC cells in this study ($n=2$ biologically independent samples). **b**, All libraries contain cells of every cell type, according to this joint annotation. This differs from the individual level clustering results, in which many libraries are missing particular cell types ($n=2$ biologically independent samples). **c**, PBMC1 and **d**, PBMC2. For each annotated cell type and method in the jointly clustered dataset (y-axis), we calculated the percentage of cells from that cell type that come from each cell type in the individual level clustering results (x-axis). This is denoted by the color of the corresponding boxes ($n=1$ biologically independent sample for (c) and (d)).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

The demultiplexed FASTQ files from the sequencing center were shared with us via FTP

Data analysis

The scumi (v0.1.0, available from <https://bitbucket.org/jerry00/scumi-dev/src/master/>) Python package was used to profile single-cell datasets to generate gene by cell UMI (read) count matrices. We used STAR v2.6.1a to align reads to genomes and featureCounts v1.6.2 to annotate each alignment with a gene tag. RSEM v1.3.0 was used to generate TPM matrices for Smart-seq2 data and bulk RNA-Seq data. HISAT2 v2.0.5 was used to alignment scRNA-seq data from Smart-seq2 and CEL-Seq2. For de-novo transcript construction, we first aligned bulk RNA-Seq data to the genome using HISAT2 v2.0.5 and then used StringTie v0.1.18 for transcript construction. We also used Cell Ranger v2.0.0, Drop-seq pipeline v1.3, the inDrops pipeline (<https://github.com/indrops/indraps>, accessed on May 23 2018), and CEL-Seq2 pipeline (<https://github.com/yanailab/CEL-Seq-pipeline>, accessed on May 23 2018) to profile the data from their corresponding platforms. We used Cell Ranger 3.0.2 to profile 10x Chromium (v3) datasets. We used Seurat v2.3.4 R package for clustering analysis and t-SNE dimension reduction, and harmony (<https://github.com/immunogenomics/harmony>, accessed on Nov 9 2018) to merge datasets from different runs/platforms. Custom R scripts for filtering low-quality cell barcodes, assigning cell types to clusters, and optimizing clustering analysis parameters were available from bitbucket repo: <https://bitbucket.org/jerry00/scumi-dev/src/master/>. Trimmomatic v0.39 was used to trim adapter sequences and low quality bases from unmapped reads. Differential expression between methods and between batches was performed with the glm function in R, using the binomial family.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-Seq (both single cell/nucleus data and bulk) data generated in this project are available from the Gene Expression Omnibus with accession number GSE132044.

Figures 2–6 have associated source data included.

No restrictions on data availability.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

We chose the number of cells to profile based on the cell capture rate of lab methods, the frequencies of different cell types in the studied tissue, the costs, and our sense of the sample size of a typical experiment. In each experiment, we aimed to collect data from ~384 cells (or nuclei) for the low-throughput methods and ~3,000 cells (or nuclei) for the high-throughput methods. We aimed for 750,000 to 1,000,000 reads per cell for low-throughput methods and 50,000 to 100,000 reads per cell for high-throughput methods.

Data exclusions

No data was specifically excluded, but filtering was performed as described in the Methods section to exclude cells (cell barcodes) of low quality. In some datasets (cell line mixtures), we used a mixture of two Student's t distribution model to model the read or UMI (log10 transformed) count distributions of each cell, and removed the cells that were likely from the mixture component with fewer reads or UMIs. For other datasets (like PMBCs), we selected more cell barcodes than expected, clustered the data, and removed clusters likely to have resulted from low quality cells. The second strategy was chosen to avoid a bias against cells with lower RNA content. We used a combination of these approaches for the cortex nuclei datasets. In addition, for the standard computational pipeline analyses, we excluded cells using a filter of a minimum number of reads per cell -- see Supplementary Table 3.

Replication

For each sample type, we generated two replicates of data using each lab method. Generally, the results (the ranking of methods) from replicates were consistent.

Randomization

Not relevant to our study. The cells for each replicate were prepared on different days or from different sources. In our comparisons of scRNA-seq methods, we used essentially identical aliquots from the same source of cells. There were no human participants to randomize.

Blinding

Not relevant to our study. We used computational analysis methods that were intended to be unbiased in their evaluation of methods.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	<input checked="" type="checkbox"/> Involved in the study <input type="checkbox"/> Antibodies <input type="checkbox"/> <input checked="" type="checkbox"/> Eukaryotic cell lines <input checked="" type="checkbox"/> <input type="checkbox"/> Palaeontology <input type="checkbox"/> <input checked="" type="checkbox"/> Animals and other organisms <input checked="" type="checkbox"/> <input type="checkbox"/> Human research participants <input checked="" type="checkbox"/> <input type="checkbox"/> Clinical data
-----	--

Methods

n/a	<input checked="" type="checkbox"/> Involved in the study <input checked="" type="checkbox"/> <input type="checkbox"/> ChIP-seq <input checked="" type="checkbox"/> <input type="checkbox"/> Flow cytometry <input checked="" type="checkbox"/> <input type="checkbox"/> MRI-based neuroimaging
-----	--

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK293 and NIH3T3 were obtained from the American Type Culture Collection.
Authentication	None of the cell lines were authenticated.
Mycoplasma contamination	Both cell lines were tested for mycoplasma contamination using a PCR-based assay (ATCC, #30-1012K) three days prior to each RNA-Seq experiment and were negative.
Commonly misidentified lines (See ICLAC register)	None of these misidentified lines were used.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Mus musculus, C57BL/6, male, 1 month old animals.
Wild animals	None used
Field-collected samples	None used
Ethics oversight	All animal-related work was performed under the guidelines of the Division of Comparative Medicine, with the protocol (# 0416-050-1) approved by the Committee for Animal Care of the Massachusetts Institute of Technology, and was consistent with the Guide for Care and Use of Laboratory Animals (1996 edition).

Note that full information on the approval of the study protocol must also be provided in the manuscript.