Systematic comparison of high throughput Single-Cell RNA-Seq platforms in complex tissues

Author suggestions:

Yolanda Colino-Sanguino[1,2], Laura Rodriguez de la Fuente[1,3,4], Brian Gloss[5], Andrew M. K. Law[3], Kristina Handler[6], Marina Pajic[3,4], Robert Salomon[7,8], David Gallego-Ortega[3,4,9*] ✉ and Fatima Valdes-Mora[1,2*]✉

1. Cancer Epigenetic Biology and Therapeutics Laboratory, Children's Cancer Institute, Lowy Cancer Centre, Kensington, NSW, Australia
2. School of Clinical Medicine, University of New South Wales Sydney, NSW, Australia
3. The Kinghorn Cancer Centre. Garvan Institute of Medical Research, Darlinghurst, NSW, Australia.
4. St. Vincent's Clinical School, Faculty of Medicine, University of New South Wales Sydney, NSW, Australia
5. Westmead Research Hub, Westmead Institute for Medical Research, Sydney, NSW, Australia
6. Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland
7. Institute for Biomedical Materials & Devices (IBMD), Faculty of Science, University of Technology Sydney, Sydney, NSW, Australia
8. ACRF Liquid Biopsy Program, Children's Cancer Institute, Lowy Cancer Centre, Kensington, NSW, Australia
9. School of Biomedical Engineering, Faculty of Engineering and Information Technology. University of Technology Sydney, NSW, Australia.

* Contributed equally

✉ Corresponding author: David.GallegoOrtega@uts.edu.au , FValdesMora@ccia.org.au

Abstract:

Single-cell transcriptomics has emerged as the preferred tool to define cell identity through the analysis of gene expression signatures. However, there are limited studies that have comprehensively compared the performance of different scRNAseq systems in complex tissues. Here, we present a systematic comparison of three well-established high throughput 3'-scRNAseq platforms: Drop-seq, 10x Chromium and BD Rhapsody; using tumours that present high cell diversity. Our experimental design includes both fresh and artificially damaged samples from the same tumours, which also provides a comparable dataset to examine their performance under challenging conditions. The performance metrics used in this study consist of gene sensitivity, mitochondrial content, reproducibility, clustering capabilities, cell type representation and ambient RNA contamination. These analyses showed that BD Rhapsody and 10x Chromium have similar but higher gene sensitivity than Drop-seq, while BD Rhapsody has the highest mitochondrial content. Interestingly, we found cell type detection biases between platforms, including a lower proportion of endothelial and myofibroblast cells in BD Rhapsody and lower gene sensitivity in granulocytes for 10x

Chromium. Moreover, the source of the ambient noise was different between plate-based and droplet-based platforms. In conclusion, our reported platform differential performance should be considered for the selection of the scRNAseq method during the study experimental designs.

Introduction

Single-cell RNA sequencing (scRNAseq) enables simultaneous profiling of gene expression of individual cells and has emerged as the tool of preference to define cell phenotypes and discover cell subsets and states (1). Single-cell transcriptomics provides an unprecedented resolution of the composition and functionality of cellular niches in homeostatic tissues and during disease (2,3). For example, scRNAseq has been used to unravel multicellular dynamic processes during embryogenesis and cell differentiation, tissue regeneration and morphogenesis, disease initiation and progression, and response to stimuli including drug treatments (4-12).

The ability to sequence genomic material combined with barcoding strategies to label each cell and RNA molecule extended the possibility of simultaneous analysis of barcoded cells in the same reaction, dramatically reducing cost and labour (13,14). The subsequent application of microfluidic devices to encapsulate individual cells in nano-droplet-sized bioreactors, like Drop-seq (6,7), or the development of high-density microwell plates to partition and capture individual cells, as Microwell-Seq (15,16) has dramatically increased the scRNAseq scale, enabling the analysis of tens-of-thousands of cells in a single reaction. Typically, these high throughput scRNAseq methods are technically challenging and difficult to control to deliver consistent outputs (17); however, commercial solutions have streamlined these processes, easing the technical demand and standardising procedures while ensuring consistent reagent quality (18).

At a first glance, all high throughput scRNAseq methods offer similar data outputs and they can interchangeably be used to interrogate any biological system. However, each method is technically different, with its own intrinsic capabilities, biases and limitations; thus, platform performance could vary with particular cell types and tissues, and specific platforms may be better suited to answer particular biological questions. There is limited information on a systematic comparison of the performance of different high throughput scRNAseq platforms

and the available scRNAseq platform comparisons are based on relatively homogenous cell cultures (19,20), or artificial mixed pools (21-23). Thus, these comparisons are unable to assess each platform's ability to distinguish cellular heterogeneity in a complex tissue scenario, failing to compare performance between cells of different lineages. For example, a recent study comparing five single cells/nucleus RNAseq high-throughput methods concluded that the commercial platform 10x Chromium was the top performing technology compared with the home brew methods (24). This study, however, did not included other commercial high-throughput platforms, such as the BD Rhapsody; and for the single-cell comparison, it only used cultured cells or PBMC samples which do not require tissue dissociation. Finally, performance comparisons need to be constantly updated as each commercial platform evolves with new and improved versions.

Here, we present a technical and biological comparison of three well-established and widely used 3'-scRNAseq platforms, including the two most popular proprietary platforms - 10x Chromium (10x Genomics) and BD Rhapsody (Becton Dickinson); which we compare with the representative home-brew platform Drop-seq in biologically complex but reproducible samples, mammary gland tumours from the MMTV-PyMT mouse model (25-28). Drop-seq is a low-cost but heavily hands-on method, while BD Rhapsody and 10x Chromium are commercially available platforms that require less technical expertise from the user but are more costly. Drop-seq and 10x Chromium are droplet-based microfluidic platforms while BD Rhapsody uses microwell-based technology where cells are randomly deposited by gravity into an array of picoliter-size wells. All systems track the cell of origin with a cell barcode and count individual molecules using unique molecular identifiers (UMIs). Although these platforms are designed to produce similar readouts, essentially a digital count of genes in each cell, the device design (microfluidic or microwell), the nature of the capture beads, the molecular design of the barcodes and UMIs, and the essential nature of the molecular workflow for RNA copy and amplification strongly differs.

In this study, we evaluate how these platforms manage the challenges of complex tissues to produce meaningful data from scRNAseq and discover the strengths and weaknesses of each platform.

Material and Methods

Mouse model

The MMTV-Polyoma Middle T antigen (PyMT) was a gift from Dr. William J. Muller (McGill University) and its generation has been previously described (26,27). At ethical endpoint, (10% ± 3% tumour/body weight, which approximately corresponds to 14-week-old animals) the mice were euthanized, and size-matched tumours were harvested and processed for single cell digestions. Genotyping was performed at the Garvan Molecular Genetics facility (NATA accredited, ISO 17025) by PCR of DNA extracted from the mouse tail tip using the following primers: CGGCGGAGCGAGGAACTGAGGAGAG and TCAGAA GACTCGGCAGTCTTAGGCG. The touchdown PCR conditions were 94ºC for 10 s of initial denaturation, followed by 10 cycles of 94ºC 10 s, 65-55ºC for 30 s and 72ºC for 1 min and 10 s and then 31 cycles of 94ºC 10 s, 55ºC for 30 s and 72ºC for 1 min and 10 s; the final extension is 72ºC for 3 minutes. All animals used in this study are heterozygous for the PyMT gene.

Tissue digestion and single-cell isolation

PyMT tumours were digested as described in (29,30). Briefly, tumours were manually dissected into 3-5 mm pieces using surgical blades and further chopped to 100 µm using a tissue chopper. Samples were enzymatically digested with 15,000 U of collagenase (Sigma Aldrich Cat# C9891) and 5,000 U of hyaluronidase (Sigma Aldrich Cat# H3506) for 30 min at 37ºC. The samples were then further digested by pipetting up and down with 0.25% trypsin (GIBCO Cat# 15090-046), in 1mM EGTA and 0.1 mg/ml of Polyvinyl alcohol dissolved in PBS for 1 min at 37ºC in a waterbath. Red blood cells were then lysed with 0.8% ammonium chloride (Sigma Aldrich Cat# A9434) dissolved in water for 5 min at 37ºC. Single cell suspensions were washed with PBS containing 2% FBS and spun at 200 x g for 5 min at 4ºC between each step. The supernatant was aspirated and 1mg/mL DNase I (Roche Cat# 10104159001) was mixed with the sample before incubation with each step. Finally, cells were filtered through a 40 mM nylon mesh (Corning) and resuspended in PBS with 2% FBS. For the generation of low-quality-like samples, digested tumours were left overnight at 4ºC, and check that viability was reduced by flow cytometry by at least 20% before processing the sample in auto macs.

All tumours were labelled with Annexin specific MACS beads using the Dead Cell Removal Kit (Miltenyi Biotec Cat# 130-090-101) following the manufacturers' instructions and dead cells were removed by passing the labelled cells through the autoMACS® Pro (Miltenyi) (see more details at Salomon et al., 2019). High viable cells (> 85% viability assessed by DAPI in

flow cytometry) were loaded into 10x Chromium, BD Rhapsody or the Drop-seq microfluidic device.

Flow cytometry

The viability and cellular content of the main cell compartments of the tumours was assessed by flow cytometry. Digested tissue samples were washed with PBS supplemented with 2% FBS and centrifuged at 200 x g for 5 min at 4°C. The pellet was then resuspended with 2% FBS in PBS for use in flow cytometry analysis. For flow cytometry analysis, single cell suspensions were incubated with anti-human antibodies for EpCAM (Clone 30-F11, BD Bioscience, Cat# 561868) and CD45 (Clone HI30, BD Bioscience, Cat# 560566) on ice in the dark for 30 min before they were washed, centrifuged, and resuspended with 2% FBS in PBS for analysis using the BD FACSymphony^{TM} Cell Analyser. To check viability, cells were stained with DAPI in a 0.5µg/mL concentration (Invitrogen, D1306) at 4°C for 3 min immediately before running the samples in the flow cytometer. Flow cytometry data were analysed using the software package FlowJo (version 10.4.2).

Drop-seq

Single cell suspensions were loaded into the microfluidic Drop-seq workflow, as previously described (30). The Online Drop-seq Laboratory Protocol version 3.1 (www.mccarrolllab.com/dropseq) was followed for the subsequent molecular steps. Briefly, single cells were captured into droplets using the microfluidic devices described for Drop-seq (6). In the droplets, single cells were gently lysed to release the cytoplasm content to allow the hybridisation of the polyA-mRNA to the poly-T microbeads, which allows simultaneous cell and molecule barcoding. Droplets were then broken and pooled and reverse transcription, based on template switching, was performed to generate STAMPs, “single-cell transcriptomes attached to microparticles”. The STAMPs were then amplified and multiplexed by Nextera tagmentation and sequenced on the Illumina Nextseq 500 using Nextseq 500 High Output v.2 kit (75 cycles, Illumina Cat# FC-404-2005). The libraries were spiked with 10% of PhiX Control v3 (Illumina Cat# FC-110-3001) at 1.8 pM and Drop-seq denaturated libraries were loaded at 1.3 pM. A total of ~4,000 cells/ per run were sequenced to obtain ~10,000 reads per cell. The sequencing output was processed using the McCaroll lab core computational protocol and mapped to mm10 genome reference.

10x Chromium

Cell viability was verified by microscopy with 0.4% of Trypan blue solution (Sigma-Aldrich, T8154) immediately before cells were loaded into the 10x Chromium Controller. We aimed to capture 8,000 cells with > 90% cell viability from each condition. Libraries were prepared using the Single Cell 3' Library Kit V3 from 10x Genomics. Sequencing was performed on an Illumina NovaSeq 5000 using a NovaSeq S4 200 cycles kit (Illumina, as follows: 28bp for Read 1, 91bp for Read 2 and 8bp for Index) to an estimated depth of 20,000-30,000 reads per cell. Cell Ranger pipeline v3.0.1 was used for Fastq file generation, alignment to the mm10 transcriptome reference and UMI counting. Barcodes corresponding to empty droplets were excluded using cell calling algorithm from Cell Ranger based on EmptyDrops (31).

BD Rhapsody

We adopted the MULTI-seq protocol (32) based on lipid-modified oligos (LMOs) for multiplexing different samples in the same BD Rhapsody cartridge. Single-cell suspensions from four different tumours were washed twice with PBS and incubated in a 200nM solution containing equal amounts of anchor LMO and sample barcode oligonucleotides (generously gifted by Prof. Gartner laboratory) on ice for 5min. After LMO-barcode labelling, we incubated the co-anchor LMO (final concentration of 200nM) in each sample for 5min on ice. 1mL of 1% BSA was added in each sample to quench the LMO binding and wash once with 1% BSA. The tumour samples were pooled in equal proportions, washed twice with 1% BSA and resuspended in the cold sample buffer (BD Rhapsody Cat. No. 650000062).

Cell viability was verified (>90%) by microscopy with 0.4% of Trypan blue solution (Sigma-Aldrich, T8154). Then, immediately before cells were loaded into a BD Rhapsody cartridge, viability was corroborated by Calcein AM (Thermo Fisher Scientific, Cat# C1430) and Draq7TM (Cat# 564904) in the BD Rhapsody™ Scanner. Cell capture, cDNA synthesis and exonuclease I treatment were performed using the BD Rhapsody Express Single-Cell Analysis System as per manufacturer's instructions. Sequencing libraries of LMO tags and mRNA were prepared using a custom protocol combining BD Rhapsody's mRNA whole transcriptome analysis (WTA) library preparation protocol and the MULTI-seq protocol (32). First, we performed 2 sequential random priming and extension (RPE) reactions of the capture beads with mRNA and LMO cDNA to increase assay sensitivity using half of the reagents for each round. Then, we used the RPE product from the supernatant to continue the

WTA library preparation following BD's workflow and the post-RPE beads were resuspended in a cold bead resuspension buffer to prepare for LMOs cDNA amplification. After three washes with the resuspension buffer, the beads were resuspended in 80µL of the MULTI-seq cDNA amplification mix containing 1µL of 2.5µM Multi-seq primer 5'-CTTGGCACCCGAGAATTCC-3' and PCR was performed as follows: (95 °C, 3 min; 95 °C, 30 sec, 60 °C, 1 min and 72 °C, 1 min for 14 cycles; 72 °C, 2 min; 4 °C hold). Next, we purified the PCR products of the LMOs using a 0.6x ratio of the SPRI beads and keeping the supernatant, followed by a left side selection using 1.8X ratio. Eluted DNA was then quantified using Qubit and the expected DNA size (~95 to 115 bp) was confirmed with the Agilent 4200 TapeStation system before proceeding with the library preparation PCR (95 °C, 5 min; 98 °C, 15 min; 60 °C, 30 min; 72 °C, 30 min; 8 cycles; 72 °C, 1 min; 4 °C hold). Each PCR reaction contained 26.25 µl 2× KAPA HiFi HotStart master mix (Roche), 2.5 µl of 10 µM TruSeq RPIX primer (Illumina), 2.5 µl of 10 µM TruSeq Universal Adapter primer (Illumina), 3.5 ng barcode cDNA and nuclease-free water to a total volume of 50 µl. A final cleanup using 1.6x SPRI bead ratio was performed before sequencing. WTA libraries and LMO libraries were pooled and sequenced on an Illumina NovaSeq 6000 using a NovaSeq S1 2x100 cycles kit (Illumina, as follows: 100bp for Read 1, 100bp for Read 2 and 8bp for Index) to an estimated depth of 25,000 reads per cell for the WTA library and 2,500 reads per cell for the LMOs.

Fastq files from BD Rhapsody samples were processed using BD Rhapsody™ WTA Analysis Pipeline, v1.9.1. WTA reads were aligned to the mm10 reference genome and LMO sequences (8bp) were included as sample tags, allowing a maximum of 1 mismatch. The sample multiplexing option of the WTA Analysis pipeline was used to determine the sample of origin of each cell and singlets were used for downstream analysis.

Single-cell data integration, clustering, and annotation

Single-cell clustering and annotation were performed using Seurat v4.2.1. A Seurat object of each sample was created using *CreateSeuratObject()* function without a minimum gene or counts filtering. First, samples from the same platform were merged and normalized and variable genes were detected using the SCTransform method (33). Cell clustering was determined by shared nearest neighbour (SNN) modularity using 15 dimensions and a resolution of 0.3 and dimensional reduction was performed using *RunUMAP()* using 15

dimensions in all three platforms. Secondly, data from all three platforms were integrated utilizing either canonical correlation analysis (CCA) or reciprocal PCA (RPCA) to identify anchors. RPCA integrated object with 5 anchors and normalized using SCTransform were used for further analysis. Cell identities were annotated using SingleR signatures based on the mouse cell type reference generated by the Immunologic Genome Project (34) and based on single cell reference mapping using the *TransferData()* function and the previously characterized PyMT tumour dataset as reference (30). Bar plots for comparing cluster abundance across platforms were generated using *dittoSeq* package and dotplots and boxplots using *ggplot2*. Gene markers of specific cell types or clusters were identified using *FindMarkers()* or *FindAllMarkers()* functions from Seurat using logfc.threshold = 1, min.pct = 0.3 and logfc.threshold = 0.25, min.pct = 0.25, respectively. To determine gene sets enriched in the upregulated genes of a specific cluster we used the *enrichGO()* function from the *clusterProfiler* package, p-value was adjusted using Benjamini-Hochberg formula and the q value cutoff set at 0.2. Gene expression correlation between different samples was calculated using the *DESeq2* package and plotted with *plotPCA()* function and the Pearson correlation was plotted with *pheatmap()*.

Public data processing and analysis

10x Chromium and BD Rhapsody from bone marrow and whole blood gene expression and antibody matrices (BioProject ID: PRJNA734283) were downloaded and re-processed using Seurat v4.2.1. Filtered 10x cells based on the default settings of Cell Ranger or based on a custom threshold of Protein UMI counts per cell (>10) as Qi et al. (35) were first demultiplexed using the Hashtag oligo (HTO) libraries and the *HTODemux()* function from Seurat. Barcodes with more than one hashtag were discarded. Cells from the bone marrow and whole blood in each platform were merged, clustered, and visualised as explained above.

Ambient noise detection

The single-cell Ambient Remover (scAR) model was used to determine the gene expression profile of the empty droplets or wells representing the ambient noise, the ratio of ambient noise in each cell and to create a denoised matrix (36). For the mouse datasets, for the setup_anndata() function, the whole list of barcodes selected were used as the unfiltered matrix and the putative cells based on either Cell Ranger of the BD Rhapsody pipeline for the filtered matrix, using a probability threshold of 0.995. Empty barcodes were used to calculate

the ambient profile and train the model. For the human dataset from BD Rhapsody the noise ratio was calculated the same way than the mouse dataset for both the Gene expression and protein matrices. For the human datasets generated in 10x Chromium, the putative cells used in the *setup_anndata()* function were based on the manually filtering the UMI Protein counts (>10).

Results

Experimental design and data processing

Our comparison is achieved by using PyMT tumours that originated by the oncogenic PyMT protein in a genetically controlled congenic mouse strain (FVBn). These tumours present multiple lineages and high cell diversity while retaining reproducibility due to the limited variability (30,37,38). In addition, tumours are biologically challenging samples, as they are not homeostatic tissues, and present tissue damage, hypoxia or cell death; features that will put to the test any scRNAseq method. Two independent tumour biopsies from the MMTV-PyMT mouse mammary tumour model were digested and enriched for viable cells using Magnetic Associated Cell Separation (MACS) as previously described (29) (**Figure 1A**). Single cell suspensions containing viable cells (>90%) were loaded into a 10x Chromium, Drop-seq or BD Rhapsody chip (**Figure 1B**) and run following the corresponding standard molecular workflow.

Each platform has a different read structure for barcodes and UMIs and each system has a different processing bioinformatic pipeline. Therefore, to replicate the standard user experience, we have used the intended pipeline for each platform to generate a Digital Gene Expression (DGE) matrix per cell: *Drop-seq core computational protocol* (McCarroll lab), *cellranger* for 10x Chromium (10x Genomics) and the whole transcriptome analysis (WTA) pipeline in Seven Bridges Genomics cloud platform for BD Rhapsody. In all cases, data was subsequently integrated and visualised using the Seurat R package (39). A critical step for single-cell data processing is to determine how many real cell barcodes were detected. This is especially challenging in Drop-seq data as the barcodes are randomly generated, while 10x Chromium and BD Rhapsody manufacturers provide a whitelist of barcodes. The latest version of the 10x Chromium pipeline, Cell Ranger 3.0, and BD's pipeline in the Seven Bridges Genomics platform, now include a step that detects putative cells. Therefore, we filtered out empty barcodes based on the default parameters of each platform. For Drop-seq, we used a user-defined strategy, based on the visual guide provided during capture and a cut-

off of 200 genes per cell. Doublets were excluded using *DoubletFinder* (40) in the case of 10x Chromium and Drop-seq, while for BD Rhapsody we used the information from the sample tags used to demultiplex tumour samples. After removing the empty droplets and doublets, we obtained 1,780 and 1,900 cells per replicate in Drop-seq, 2,075 and 1,925 cells for 10x Chromium and 1,359 and 1,703 cells for BD Rhapsody.

Technical comparison: assessing sensitivity and reproducibility

An important quality metric for scRNA-seq is the sensitivity of gene and UMI detection per cell. This sensitivity metric was different between the platforms (**Figure 2A**); we detected a median of 634.5, 3,026.5 and 3,269.5 genes, and 1,326, 10,710 and 12,861 UMIs per cell in Drop-seq, 10x Chromium and BD Rhapsody, respectively. Based on this analysis Drop-seq had the least gene and UMI sensitivity, while 10x Chromium and BD Rhapsody had a similar UMI and genes counts per cell.

The proportion of mitochondrial genes is another metric commonly used to assess cell quality. Here we found that cells processed with 10x Chromium have significantly less mitochondrial content than Drop-seq or BD Rhapsody samples (**Figure 2A, right plot**). Interestingly, when we examined the reference used in each pipeline, we observed that the number of mitochondrial genes included in the annotation pipeline was very different (**Figure S1A**). The Drop-seq pipeline included almost 3 times more mitochondrial genes, including the highly expressed gene mt-Rnr2. In fact, when we removed mt-Rnr2 from the data, the Drop-seq's percentage of mitochondrial genes drops to similar levels to 10x Chromium (**Figure S1B**). Excluding mt-Rnr2, the percentage of mitochondrial content in Drop-seq and 10x Chromium was more similar with a median of less than 6%, while in BD Rhapsody was 15%.

An essential application for scRNAseq is to define single-cell phenotypes by their gene expression signature through dimensional reduction, principal component analysis (PCA), and clustering algorithms (41). We used the top 3,000 variable genes for dimensional reduction analysis in all three platforms; however, the variable genes detected in each platform were slightly different, with 1,332 common genes and the Drop-seq platform having the least overlap and 10x Chromium and BD Rhapsody having the highest overlap with 2,113 common genes **(Figure S1C**). The total variability identified in the system, measured as the standard deviation at increasing principal components (PCs) was lower on the Drop-seq

platform, while 10x Chromium and BD Rhapsody have similar standard deviations per PC (**Figure 2B**). When we examined the clustering capabilities at increasing resolutions, we found that at the lowest resolution, only BD Rhapsody was able to split the cells into two clusters suggesting that their detected transcriptome was more distinct, however, at increasing resolutions, we found that 10x Chromium had a greater number of clusters and more consistent cluster composition than BD Rhapsody or Drop-seq (**Figure 2C**). This consistency was measured by the capability of a cluster to further split into smaller clusters at higher resolutions without cells crossing into other cluster branches (**Figure 2C, crossing arrows**). However, when resolutions were set higher than 0.1, BD Rhapsody had a consistently smaller number of clusters detected. In conclusion, the gene variability detected in Drop-seq was lower than BD Rhapsody and 10x Chromium, and cluster assignment was less consistent. In contrast, BD Rhapsody and 10x Chromium had similar variability but 10x Chromium outperformed BD Rhapsody at consistently assigning cells to cluster at high resolutions.

Finally, to assess the reproducibility of each platform, we looked at the correlation of gene expression between replicates (**Figure 2D-F**). We found that the Drop-seq replicates were slightly more variable with a correlation of 0.985, while both 10x Chromium and BD Rhapsody had over 0.99 correlation values. The correlation between 10x Chromium and BD Rhapsody was higher than between Drop-seq and any of the other two platforms (**Figure 2E and F**).

Cell type identification and lineage biases among platforms

To directly compare the performance of each platform to resolve tumour heterogeneity, we merged (**Figure S2A**) and integrated (**Figure 3A**) the data from the three platforms into one Seurat object. For the integration, we tested both canonical correlation analysis (CCA) and Reciprocal PCA integration (RPCA) (**Figure S2B**). We found that RPCA outperformed CCA overlaying the three platforms, and it has been reported to be less prone to overcorrection (42). Therefore, we used the RPCA approach for the downstream analyses. For cell type identification, first, we divided the integrated object in the main cell lineages in PyMT tumours (epithelial, immune and stroma) **(Figure 3B)** and we compared their ratio in each platform to the observed percentage by flow cytometry, as it is the current gold standard method for cell annotation (**Figure 3C**). Overall, we found similar percentages between

scRNAseq and flow cytometry, but the samples analysed through 10x Chromium had a slightly higher number of stroma cells and fewer epithelial cells.

Next, we assigned each cell cluster to the main cell types based on SingleR (34) (**Figure S3A and B**) and also based on label transfer of the cell types from our previously published PyMT tumour reference (30) (**Figure S3C and D**). We found that all major cell types: B cells, T cells, myeloid cells, endothelial cells, epithelial cells, fibroblasts and myofibroblasts were detected in all platforms (**Figure 3D**), with similar percentages of cell types between replicates and platforms, except for endothelial cells and myofibroblasts (**Figure 3E**). In BD Rhapsody, endothelial cells made 1% of the total cells while in 10x Chromium and Drop-seq it was between 9-6%, and myofibroblast where 3 times or 10 times less detected in BD Rhapsody compared to Drop-seq and 10x Chromium, respectively (**Figure 3E**). In contrast, as previously reported (35), BD Rhapsody performed slightly better at detecting myeloid cells (**Figure 3E** and **3F**). Consistently, the ratio of endothelial, myofibroblast and myeloid cells in BD Rhapsody deviated from the expected inter-tumour variability in the PyMT tumour reference atlas across eight tumours (30), suggesting that these differences were not a consequence of tumour heterogeneity (**Figure 3F)**. We further confirm the consistency of cell annotation across platforms by looking at the top markers of each cell type and confirmed that these markers were similarly expressed in all cells captured in all three platforms (**Figure 3G**). However, we found that the Drop-seq platform, followed by BD Rhapsody, showed a higher percentage of cells expressing non-specific markers, again potentially suggesting higher noise compared to 10x Chromium.

To determine how the different cell population ratios, gene marker dropout or ambient noise can affect downstream analysis, we performed cell-to-cell communication predictions using the CellChat software on all three platforms independently. We found that the highest number of interactions was between fibroblasts or myofibroblasts and other cell types (**Figure S4A**), with fibroblasts and epithelial cells being the strongest interaction (**Figure S4B**). However, Drop-seq showed the smallest number of overall interactions, especially with T cells and myeloid cells (**Figure S4A**). Interestingly and in line with the number of cells detected, the interaction between endothelial or myofibroblasts with epithelial cells was weaker in BD Rhapsody, while between myeloid cells and epithelial cells was stronger (**Figure S4B**).

In summary, 10x Chromium detected more stroma cells, endothelial and myofibroblasts, while BD Rhapsody detected a slightly higher percentage of myeloid cells, which resulted in different levels of cell-cell interactions detected with those cell types in each platform. Cell type ratios in Drop-seq were in between the other two platforms but cells had more non-specific markers and there were less cell-to-cell interactions detected.

Cell subtype proportions across platforms

To further evaluate how each platform can distinguish among cellular subtypes within the main cell types in tumours, we performed subclustering of some of the main cell types. First, we selected the most abundant cell type in PyMT tumours, the epithelial compartment; unsupervised clustering revealed differential distribution in some cell clusters including clusters 1, 4, 5, 7 and 10 out of the 13 clusters (**Figure 4A and S5A**). To infer the identity of each cell subcluster, we again used the PyMT tumour atlas (30) of the epithelial cell subtypes in the integrated epithelial cluster (**Figures S5B and S5C**), that resulted in the annotation of 9 epithelial subtypes manually (**Figure 4B**). As expected in this model, most cells were classified as luminal progenitors (LP), including a subset of hormone-sensitive luminal progenitor (LP-HS) cells where no major differences were observed between platforms (**Figure 4C**). Interestingly, the cell clusters where we visually observed differences (**Figure 4A**) were classified as myoepithelial (cluster 1), Multi/Stem (clusters 4 and 10), basal (cluster 5) and luminal progenitor/ luminal hormone-sensing (LP/LP-HS) (cluster 7) (**Figure 4B and 4C**). Strikingly, cluster 7, or LP/LP-HS cells, was missing from the PyMT reference and was heavily comprised of cells from the 10x Chromium platform. Gene set enrichment analysis identified that this Chromium-specific luminal progenitor cluster was enriched with genes involved in immune responses, including antiviral response and interferon (**Figure 4D)**. This suggests that the 10x Chromium platform detects more epithelial cells that are actively interacting with the immune system. Another cluster we found major differences between platforms was the Multi/Stem **(Figure 4B** and **C**), with the highest proportion found in the Drop-seq platform. Even though, the gene expression profile of clusters 4 and 10 correlated with the multipotent/stem cell type and hormone-sensitive progenitors (**Figure S5B** and **C**), we found that these clusters also had the highest percentage of mitochondrial gene content and the lowest number of genes detected in all platforms (**Figure 4E**), potentially suggesting damaged cells. Considering cells with high mitochondrial content but low number of genes detected as the definition of damaged cells, the 10x Chromium platform had the lowest ratio

of damaged cells, followed by BD Rhapsody. The cluster that correlated with cycling cells had a similar number of cells in BD Rhapsody and 10x Chromium but two times less in Drop-seq (**Figure 4C**). However, when inferring the cell cycle phase based on gene signatures, we found a similar ratio of cells in G1, S and G2 across platforms but most of Drop-seq cells in S or G2 phase were found outside the cycling cluster (**Figure S5D** and **S5E**), suggesting that were miscalled, likely due to low gene sensitivity and technical noise.

Next, we did the same analysis in a less abundant population, the fibroblasts. We again assigned cell types based on the fibroblast subsets in PyMT tumours (30) (**Figure 4F**) and found that within the fibroblast partitions, there were four clusters that correlated with either myofibroblasts or secretory cancer-associated fibroblasts (CAFs) which were subdivided into extracellular-matrix synthesis CAFs (ECM-CAFs) and inflammatory CAFs (iCAFs) (**Figure 4F**). The 10x Chromium platform detected a lower percentage of inflammatory CAFs, while the BD Rhapsody platform detected fewer ECM-CAFs and myofibroblasts (**Figure 4G**). In contrast, in Drop-seq the ratio of each subtype of fibroblast was between the one found in 10x Chromium and BD Rhapsody (**Figure 4G**).

Together these data suggest that, even though all major cell types were represented in all platforms, there are still cell type detection biases intrinsic to each platform which should be considered when analysing the tissue heterogeneity with one platform exclusively.

Ambient noise comparison between 10x Chromium and BD Rhapsody in high and low-quality samples

One of the limitations of single-cell data is the technical noise caused by the ambient RNA contamination (43), amplification bias during library preparation and index swapping during sequencing (44). The nature of the cell partition method (nanowell or oil droplet) and the differences in the RNA amplification steps in each platform (**Figure 1B**) could result in a differential origin of the technical and/or ambient noise. Ambient RNA is defined as the mRNA molecules that have been released from dead or stressed cells and are part of the cell suspension. The ambient noise is especially present in samples that required tissue digestion and in low-quality samples, such as tissues stored for an extended period before processing, or tissues with active cell dead or hypoxia regions like tumour tissues.

Our data suggest that noise might be handled differently by the commercial platforms (**Figure 2C**, **3G** and **4E**). Thus, we studied how the two commercial platforms, 10x Chromium and BD Rhapsody performed with the technical noise from challenging samples. To recreate this "low quality sample", we incubated PyMT digested tumours for 24h at 4°C. After 24h, cell viability dropped 20 per cent compared with the freshly digested sample as measured by DAPI positive cells in flow cytometry, indicating a substantial cell decay (**Figure 5A** and **Figure S6**). As done before with fresh PyMT tumour samples, we removed dead cells using MACS and processed the samples for scRNAseq using the 10x Chromium or BD Rhapsody platforms.

First, we compared the tissue heterogeneity of the damaged samples to the fresh cells. As for this analysis we increased the number of cells to ~4000 of cells per condition, we were able to further split the myeloid cell compartment into neutrophils and macrophages. We found that endothelial cells, fibroblasts, stroma cells and neutrophils were consistently lost on the damaged sample regardless of the platform (**Figure 5B)**. However, it is important to note that the cell types with a low number of cells such as neutrophils and stroma cells are more susceptible to technical variability, as shown in **Figure 5C** for fresh samples.

Next, to assess the technical noise found in each platform, we used the single-cell Ambient Remover (scAR) method, a universal model to detect noise across single-cell platforms based on probabilistic deep learning of the ambient signal (36). Interestingly, around 5% of cells in both platforms and in any condition (fresh or damaged) were considered empty cells as their transcriptome was indistinguishable from the ambient signal of the empty barcodes (**Figure S7A**). Those so-called "empty cells" also had a high percentage of mitochondrial content (**Figure 5D**) and low gene content and therefore confirming that these assigned "cells" were instead damaged cells that thus needed to be removed from downstream analysis. When we compared the noise ratio of the remaining cells, we found that BD Rhapsody has significantly more ambient noise per cell than 10x Chromium (**Figure 5E**). Surprisingly, the ambient contamination was significantly increased in the damaged sample in the 10x Chromium but not in BD Rhapsody. We confirmed our previous observation of a higher percentage of mitochondrial content in Rhapsody (**Figure 2A**), however as expected in the damaged samples, the mitochondrial content increased in

both platforms (**Figure 5E)**, while the number of genes found per cell slightly decreased in the damaged samples (**Figure S7B**).

To evaluate which cell lineages are most affected by the technical noise in each platform, we compared the noise ratio per cell type and platform in each sample condition. We found that in BD Rhapsody all cell types had a similar level of noise, regardless of their damaged condition, while 10x Chromium's noise was notably different between cell types and there were significant differences between fresh and damaged conditions (**Figure 5F top**). Particularly, in 10x Chromium, neutrophils had the highest percentage of noise which increased to an even higher ratio in the damaged samples. As expected, neutrophils also had the lowest number of genes detected (**Figure 5F bottom**). To understand which cell types and genes are driving the ambient signal, we looked at expression across cell types of the top genes contributing to the ambient pool based on scAR analysis (**Figure 5G**). Interestingly, in the samples processed with 10x Chromium, 3 and 5 out of the top 6 most abundant genes in the ambient RNA pool in the fresh and the damaged sample, respectively, were neutrophils markers, despite the low number of neutrophils identified in the sample. In contrast, we did not find any strong expression specificity among the top genes in the ambient of the samples processed by BD Rhapsody, except for *Efemp2*, a fibroblast marker, in the fresh sample. Remarkably, in 10x Chromium, the ambient expression of the *S100a9* neutrophil marker was > 25-fold times higher than any other cell type marker and was even higher in the damaged samples, while in BD Rhapsody the ambient gene expression was more uniformly distributed and did not show major differences between fresh and damaged sample (**Figure S7C**). To confirm the neutrophil biases from 10x Chromium in the ambient profile, we correlated the average transcriptome of the single cells with the ambient gene expression and observed a higher presence of neutrophil marker genes in ambient RNA than in the single cells in 10x Chromium, but not in BD Rhapsody (**Figure S7D**).

The scAR algorithm can denoise the count matrix based on the ambient composition. When we used the denoised data to resolve cell type clusters, we found that clusters are more distinct using denoised data and there is less background expression of non-specific markers (**Figure S8**). As expected, the expression of the neutrophil markers was completely lost in the 10x Chromium data in this cell subtype (**Figure S8B**).

In conclusion, the differential molecular design, the microfluidic versus microwell format and the sample quality are sources of the ambient noise. In BD Rhapsody there is a generalised level of noise that is independent of the cell of origin and sample quality, while in 10x Chromium, ambient noise is cell type-specific and overrepresented by the neutrophil population.

Cell type bias noise validation on public datasets

To determine if the cell type biases and technical noise differences between 10x Chromium and BD Rhapsody were universal, we assessed another two public data sets, human whole blood and bone marrow, where more neutrophils are expected, and both platforms were used (35). This data was processed using CITEseq (10x Chromium) and Ab-Seq + whole transcriptome analysis (BD Rhapsody) and therefore it also included 30 antibody-derived barcodes in both datasets. Additionally, hashtag oligos (HTO) were used in 10x Chromium platform to demultiplex samples.

We first compared the abundance of each cell type detected based on the antibody markers in both platforms using the default parameters for filtering established by Cell Ranger and the Rhapsody pipelines (**Figure 6A**). After merging the whole blood and bone marrow samples, we detected virtually no granulocytes (neutrophils, basophils or eosinophils) in 10x Chromium (**Figure 6B**). This is in line with previous reports of a low number of neutrophils in human datasets detected using this platform (45,46). Interestingly, if we filtered the 10x Chromium cell barcodes using the custom threshold based on the Protein UMI counts per cell as Qi et al. (35), we recovered 4,477 and 4,186 cells, in the bone marrow and whole blood samples, respectively, from which the majority were neutrophils (**Figure 6C**). In addition, using this custom filtering for all samples, resulted in very similar percentages of each cell type in the bone marrow and whole blood between the 10x and BD platforms (**Figure 6D**). However, even though the UMAP using the protein matrix was able to distinguish all major cell types in 10x Chromium (**Figure 6E**), the low gene sensitivity in the granulocyte population, with a median of fewer than 100 counts per cell (**Figure 6F**), did not manage to resolve those clusters at the transcriptome level (**Figure 6F, and Figure 6E right panel**). In fact, when we run the scAR algorithm using the threshold for filtering cells based on Protein UMI counts, we found that in 10x Chromium some granulocytes, especially eosinophils, still had a higher RNA noise ratio, although not as pronounced as in the mouse

dataset. Overall, the noise ratio from both the RNA and the protein data was more similar across cell types in BD Rhapsody than in 10x Chromium, confirming that in BD Rhapsody the ambient noise is not cell type specific. In contrast, in 10x Chromium the noise ratio was highest in granulocytes, which were indistinguishable from the ambient RNA found in "empty droplets" (**Figure 6G**). These results are in line with our tumour mouse results. In summary, 10x Chromium was unable to detect granulocytes in human samples based exclusively on their RNA content, and their low gene detection hinders the denoising algorithms to distinguish empty droplets from real cells with low UMI counts.

Discussion

Single-cell RNA-seq allows the classification of different cell types based on the gene expression profile of individual cells. The rise of commercial instruments and kits has enabled the use of this technique by the broad scientific community. However, there are still technical challenges during sample preparation, cell capture and library preparation (47), that can affect the bioinformatic readout and consequently, the biological insights driven from the data. Therefore, it is crucial to understand the advantages and limitations of each platform to best tailor the selection of a platform to the sample type and the biological question that wants to be answered, in addition to the costs and expertise level of the user. Here, we have compared the most popular commercial platforms, the microfluidic-based system, 10x Chromium, and the microwell plate-based system, BD Rhapsody, as well as Drop-seq, a low-cost but hands-on and home-made microfluidic method using mammary gland tumours from the PyMT breast cancer mouse model (25-28). Analysing samples that require tissue digestion prior to the single-cell experiments is more challenging as the damaged cells can introduce additional ambient noise and different cell-type vulnerabilities can change the observed tissue heterogeneity. An advantage of our design is the comparison of the same tumours digested using the same protocol and in the same laboratory, which allows us to discard sample preparation as the source of variability. In addition, we have created matching low-quality samples with reduced viability to assess how the quality of the sample affects the readout from 10x Chromium and BD Rhapsody. Frequently, the samples processed for scRNAseq have high levels of damaged cells, for example, tumour samples have necrotic regions or increased cell death due to their high cell turnover or exposure to treatments, such as radiotherapy or chemotherapy. Transportation of the biological material to the laboratory

where cell capture is performed may also increase the cellular stress of the sample. In this study, we have used a biologically relevant tissue and replicated the standard user experience allowing us a realistic assessment of the different technologies in challenging samples.

Our comprehensive data analysis identified, first, at the technical level, that 10x Chromium and BD Rhapsody had comparable overall gene sensitivity and reproducibility at similar sequencing depths, while Drop-seq was far inferior. Next, we measured tissue heterogeneity and found a lower percentage of endothelial and myofibroblast cells in BD Rhapsody compared to 10x Chromium and Drop-seq. However, comparing their ratios with the flow cytometry data, we found that 10x Chromium tends to detect more cells from the stroma and fewer epithelial cells than BD Rhapsody and flow cytometry. This suggests that 10x Chromium may enrich stroma cells to the detriment of cancer epithelial cells which could be beneficial if you are interested in those populations. Interestingly, endothelial, and myofibroblast cells are reduced in low-quality samples. We also found differences within the epithelial and fibroblast compartments, including an immune-responsive cluster within the epithelial cells that were enriched in the 10x Chromium samples. Orthogonal validation of this cell population is needed to understand its origin.

In the immune compartment, the biggest differences were found within the myeloid cells, especially in granulocytes, which 10x Chromium was not able to detect granulocytes due to the low gene detection, which misclassifies them as empty droplets. It is known that granules in neutrophils, eosinophils and basophils have high levels of nucleases including RNases (48). This has been suggested to be the reason why granulocytes have lower genes and counts detected than other cell types (35,49). We have also confirmed in our PyMT dataset and human blood and bone marrow public datasets that granulocytes, including neutrophils, have the lowest number of genes detected. Interestingly, BD Rhapsody's neutrophils had an average of 1,240 or 602 genes while in 10x Chromium neutrophils only had 551 or 10 in the mouse and human datasets, respectively. We hypothesise that there are two factors that may contribute to this phenomenon; firstly, neutrophils may be more sensitive to the pressure from the microfluidic devices which could cause their breakage and loss of RNA to the ambient pool in the 10x Chromium system; secondly, we hypothesise that the 10x platform is more sensitive to the presence of RNases released by neutrophils than BD Rhapsody. Supporting this theory, Qi et al found that cell doublets of neutrophils and T cells from 10x Chromium had less number of UMIs detected than T cell singlets, suggesting that

the RNases from the neutrophils degraded the RNA of the T cells (35). The exact recipes of the lysis buffers used in 10x Chromium and BD Rhapsody are not disclosed, but it is possible that BD Rhapsody has stronger detergents or RNase inhibitors that degrade RNases. BD Rhapsody workflow also includes washes of the beads before cDNA generation which could remove any remaining RNases or strong detergent from the lysis buffer to not interfere with the reverse transcription reaction. Moreover, a recent report describing TAS-seq, a BD Rhapsody-based technique with improved cDNA amplification step, resulted in better cellular composition fidelity, especially in neutrophils, using their system compared to 10x Chromium and Smart-seq (50). Together, this demonstrates that BD Rhapsody is better suited for granulocytes research, and new strategies for cDNA amplification that increase the number of genes detected per cell will further improve the transcriptional resolution of this challenging cell population.

Mitochondrial content has been commonly used as a sign of broken cells, where their cytoplasmatic RNA has been lost but the RNA contained in the mitochondria remains (51). In fact, we see an increase in the mitochondrial fraction in the damaged samples. Surprisingly, all cells processed with the BD Rhapsody have a higher percentage of mitochondrial genes, even though they are not considered damaged. A possible explanation for this is that the lysis buffer of BD Rhapsody is more effective at digesting the organelles including mitochondria and therefore releasing the mitochondrial RNA into the cytosol to bind to the beads. For this reason, different thresholds of mitochondrial content should be used to filter out damaged cells in BD Rhapsody or 10x Chromium.

Ambient RNA release during the sample preparation and cell capture introduces noise to the data affecting the downstream analysis. Here, we have compared the level of ambient noise present in each barcode corresponding to putative cells and the gene expression of the ambient RNA from empty barcodes to identify vulnerabilities across platforms. We found that BD Rhapsody has a higher ratio of ambient RNA per cell than 10x Chromium, but the source of that RNA was coming from all cell types, suggesting a uniform contribution to the ambient pool. Ambient noise in BD Rhapsody could be created during the bead retrieval step where all beads are pooled and unbound RNAs could bind to free poly T tails from beads of different wells, and thus the likelihood of this occurring would be very similar across all cell types, in contrast to 10x Chromium where the RT occurs inside the droplet. Alternatively, the perceived high level of noise in BD Rhapsody could be due to the housekeeping genes'

contribution to the ambient RNA which are constitutively expressed in all cells at high levels. In contrast, the 10x Chromium ambient pool was biased towards certain cell types, in particular granulocyte, likely caused by damaged cells or RNases reducing the number of UMIs making it hard to distinguish between real or empty barcodes, as previously discussed. This suggests that BD Rhapsody may benefit from data-cleaning steps where the ambient fractions are removed, while 10x Chromium decontamination of the data may result in cell-type dropouts.

A limitation of our comparison analysis is the fact that we have used mouse mammary gland tumours, human whole blood and bone marrow only. Other tissues where different cell types are found may show other cell type biases, especially for lowly represented cell types. However, this study highlights the importance of using different technologies when assessing the population ratios of a sample and avoiding comparing cell type ratios across biological conditions performed in different platforms. Based on our analysis, we also found that the expected quality of the sample and origin of the tissue should also be considered when choosing a platform to perform single-cell RNAseq, as damaged tissues or tissues with high levels of RNases, such as bone marrow or spleen (49), would be more susceptible to RNA loss in 10x Chromium.

Complementary strategies, such as performing multi-omic studies where two or more molecular layers of information are investigated per cell (52), or spatial transcriptomics (53), as well as alternative methods for tissue preservation such as the ALTEN system (11) or parafilm fixation in combination with fixed RNA profiling (49,54), may also help dissect the true heterogeneity of complex tissues and overcome the current limitations of single-cell RNAseq.

Data Availability

Data will be made available through GEO upon publication acceptance. A reviewer token will be supplied upon request. Human data sets were downloaded from PRJNA734283

Funding

This work is supported by Cancer Institute NSW Fellowship (2019/CDF002-CDF181218) to FVM and Cancer Institute NSW Fellowship (DG00625) and Cancer Council NSW project

grant (RG18-03), NHMRC (2012941), NBCF Elaine Henry Fellowship (IIRS-21-096) to DGO; YCS is supported by Cure Brain Cancer Foundation (20210917)

Acknowledgements

We thank the Garvan-Weizmann Centre for Cellular Genomics, the Molecular Genetics Facility, the Australian BioResources Pty Ltd facility and the KCCG Sequencing Laboratory from the Garvan Institute. We would also like to thank Dr Antoine de Weck for his guidance on the use of scAR.

Author contributions

Conceptualisation: DGO, FVM and YCS; Methodology: LRF, YCS, AMKL, KH; Formal analysis: YCS, BG; Investigation and validation: YCS, LRF, AMKL, KH; Resources: RS, MP, DGO, FVM; Data curation: YCS, DGO, FVM; Visualisation: YCS; Funding acquisition: FVM, DGO, MP, RS. Writing-Original, Reviewing & Editing: YCS, FVM, DGO. Supervision: FVM and DGO.

References

1. Tanay, A. and Regev, A. (2017) Scaling single-cell genomics from phenomenology to mechanism. *Nature*, **541**, 331.
2. Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D. *et al.* (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, **498**, 236-240.
3. Wills, Q.F., Livak, K.J., Tipping, A.J., Enver, T., Goldson, A.J., Sexton, D.W. and Holmes, C. (2013) Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nat Biotechnol*, **31**, 748-752.
4. Liu, S. and Trapnell, C. (2016) Single-cell transcriptome sequencing: recent advances and remaining challenges. *F1000Res*, **5**.
5. Pal, B., Chen, Y., Vaillant, F., Jamieson, P., Gordon, L., Rios, A.C., Wilcox, S., Fu, N., Liu, K.H., Jackling, F.C. *et al.* (2017) Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nat Commun*, **8**, 1627.
6. 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.
7. 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.

8. Tirosh, I., Venteicher, A.S., Hebert, C., Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M., Rodman, C., Mount, C., Filbin, M.G. *et al.* (2016) Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature*, **539**, 309-313.
9. Farbehi, N., Patrick, R., Dorison, A., Xaymardan, M., Janbandhu, V., Wystub-Lis, K., Ho, J.W., Nordon, R.E. and Harvey, R.P. (2019) Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. *Elife*, **8**.
10. Herring, C.A., Simmons, R.K., Freytag, S., Poppe, D., Moffet, J.J.D., Pflueger, J., Buckberry, S., Vargas-Landin, D.B., Clement, O., Echeverria, E.G. *et al.* (2022) Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell*, **185**, 4428-4447 e4428.
11. Law, A.M.K., Chen, J., Colino-Sanguino, Y., Fuente, L.R., Fang, G., Grimes, S.M., Lu, H., Huang, R.J., Boyle, S.T., Venhuizen, J. *et al.* (2022) ALTEN: A High-Fidelity Primary Tissue-Engineering Platform to Assess Cellular Responses Ex Vivo. *Adv Sci (Weinh)*, **9**, e2103332.
12. Twigger, A.J., Engelbrecht, L.K., Bach, K., Schultz-Pernice, I., Pensa, S., Stenning, J., Petricca, S., Scheel, C.H. and Khaled, W.T. (2022) Transcriptional changes in the mammary gland during lactation revealed by single cell sequencing of cells from human milk. *Nat Commun*, **13**, 562.
13. 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.
14. Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A. *et al.* (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science*, **343**, 776-779.
15. Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F. *et al.* (2018) Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*, **172**, 1091-1107 e1017.
16. Chen, H., Liao, Y., Zhang, G., Sun, Z., Yang, L., Fang, X., Sun, H., Ma, L., Fu, Y., Li, J. *et al.* (2021) High-throughput Microwell-seq 2.0 profiles massively multiplexed chemical perturbation. *Cell Discov*, **7**, 107.
17. Salomon, R., Kaczorowski, D., Valdes-Mora, F., Nordon, R.E., Neild, A., Farbehi, N., Bartonicek, N. and Gallego-Ortega, D. (2019) Droplet-based single cell RNAseq tools: a practical guide. *Lab Chip*, **19**, 1706-1727.
18. Svensson, V., Vento-Tormo, R. and Teichmann, S.A. (2018) Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc*, **13**, 599-604.
19. Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I. and Enard, W. (2017) Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell*, **65**, 631-643 e634.
20. Zhang, X., Li, T., Liu, F., Chen, Y., Yao, J., Li, Z., Huang, Y. and Wang, J. (2019) Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol Cell*, **73**, 130-142 e135.
21. Dueck, H.R., Ai, R., Camarena, A., Ding, B., Dominguez, R., Evgrafov, O.V., Fan, J.B., Fisher, S.A., Herstein, J.S., Kim, T.K. *et al.* (2016) Assessing characteristics of RNA amplification methods for single cell RNA sequencing. *BMC Genomics*, **17**, 966.
22. Svensson, V., Natarajan, K.N., Ly, L.H., Miragaia, R.J., Labalette, C., Macaulay, I.C., Cvejic, A. and Teichmann, S.A. (2017) Power analysis of single-cell RNA-sequencing experiments. *Nat Methods*, **14**, 381-387.

23. Tian, L., Dong, X., Freytag, S., Le Cao, K.A., Su, S., JalalAbadi, A., Amann-Zalcenstein, D., Weber, T.S., Seidi, A., Jabbari, J.S. *et al.* (2019) Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. *Nat Methods*, **16**, 479-487.
24. Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T. *et al.* (2020) Author Correction: Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol*, **38**, 756.
25. Gallego-Ortega, D., Ledger, A., Roden, D.L., Law, A.M., Magenau, A., Kikhtyak, Z., Cho, C., Allerdice, S.L., Lee, H.J., Valdes-Mora, F. *et al.* (2015) ELF5 Drives Lung Metastasis in Luminal Breast Cancer through Recruitment of Gr1+ CD11b+ Myeloid-Derived Suppressor Cells. *PLoS Biol*, **13**, e1002330.
26. Guy, C.T., Cardiff, R.D. and Muller, W.J. (1992) Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol*, **12**, 954-961.
27. Lin, E.Y., Jones, J.G., Li, P., Zhu, L., Whitney, K.D., Muller, W.J. and Pollard, J.W. (2003) Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *Am J Pathol*, **163**, 2113-2126.
28. Maglione, J.E., Moghanaki, D., Young, L.J., Manner, C.K., Ellies, L.G., Joseph, S.O., Nicholson, B., Cardiff, R.D. and MacLeod, C.L. (2001) Transgenic Polyoma middle-T mice model premalignant mammary disease. *Cancer Res*, **61**, 8298-8305.
29. Rodriguez de la Fuente, L., Law, A.M.K., Gallego-Ortega, D. and Valdes-Mora, F. (2021) Tumor dissociation of highly viable cell suspensions for single-cell omic analyses in mouse models of breast cancer. *STAR Protoc*, **2**, 100841.
30. Valdes-Mora, F., Salomon, R., Gloss, B.S., Law, A.M.K., Venhuizen, J., Castillo, L., Murphy, K.J., Magenau, A., Papanicolaou, M., Rodriguez de la Fuente, L. *et al.* (2021) Single-cell transcriptomics reveals involution mimicry during the specification of the basal breast cancer subtype. *Cell Rep*, **35**, 108945.
31. Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., participants in the 1st Human Cell Atlas, J. and Marioni, J.C. (2019) EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol*, **20**, 63.
32. McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu, J.L., Murrow, L.M., Weissman, J.S., Werb, Z. *et al.* (2019) MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods*, **16**, 619-626.
33. Hafemeister, C. and Satija, R. (2019) Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*, **20**, 296.
34. Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R. *et al.* (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*, **20**, 163-172.
35. Qi, J., D'Souza, D., Dawson, T., Geanon, D., Stefanos, H., Marvin, R., Walker, L. and Rahman, A.H. (2021) Multimodal Single-Cell Characterization of the Human Granulocyte Lineage. *bioRxiv*, 2021.2006.2012.448210.
36. Sheng, C., Lopes, R., Li, G., Schuierer, S., Waldt, A., Cuttat, R., Dimitrieva, S., Kauffmann, A., Durand, E., Galli, G.G. *et al.* (2022) Probabilistic modeling of ambient noise in single-cell omics data. *bioRxiv*, 2022.2001.2014.476312.

37. Araujo, A.M., Abaurrea, A., Azcoaga, P., Lopez-Velazco, J.I., Manzano, S., Rodriguez, J., Rezola, R., Egia-Mendikute, L., Valdes-Mora, F., Flores, J.M. *et al.* (2022) Stromal oncostatin M cytokine promotes breast cancer progression by reprogramming the tumor microenvironment. *J Clin Invest*, **132**.
38. Papanicolaou, M., Parker, A.L., Yam, M., Filipe, E.C., Wu, S.Z., Chitty, J.L., Wyllie, K., Tran, E., Mok, E., Nadalini, A. *et al.* (2022) Temporal profiling of the breast tumour microenvironment reveals collagen XII as a driver of metastasis. *Nat Commun*, **13**, 4587.
39. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019) Comprehensive Integration of Single-Cell Data. *Cell*, **177**, 1888-1902 e1821.
40. McGinnis, C.S., Murrow, L.M. and Gartner, Z.J. (2019) DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst*, **8**, 329-337 e324.
41. Pezzotti, N., Lelieveldt, B.P.F., Van Der Maaten, L., Hollt, T., Eisemann, E. and Vilanova, A. (2017) Approximated and User Steerable tSNE for Progressive Visual Analytics. *IEEE Trans Vis Comput Graph*, **23**, 1739-1752.
42. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M. *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**, 3573-3587 e3529.
43. Yang, S., Corbett, S.E., Koga, Y., Wang, Z., Johnson, W.E., Yajima, M. and Campbell, J.D. (2020) Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol*, **21**, 57.
44. Costello, M., Fleharty, M., Abreu, J., Farjoun, Y., Ferriera, S., Holmes, L., Granger, B., Green, L., Howd, T., Mason, T. *et al.* (2018) Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics*, **19**, 332.
45. Travaglini, K.J., Nabhan, A.N., Penland, L., Sinha, R., Gillich, A., Sit, R.V., Chang, S., Conley, S.D., Mori, Y., Seita, J. *et al.* (2020) A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature*, **587**, 619-625.
46. Martin, J.C., Chang, C., Boschetti, G., Ungaro, R., Giri, M., Grout, J.A., Gettler, K., Chuang, L.S., Nayar, S., Greenstein, A.J. *et al.* (2019) Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell*, **178**, 1493-1508 e1420.
47. Ke, M., Elshenawy, B., Sheldon, H., Arora, A. and Buffa, F.M. (2022) Single cell RNA-sequencing: A powerful yet still challenging technology to study cellular heterogeneity. *Bioessays*, **44**, e2200084.
48. Lu, L., Li, J., Moussaoui, M. and Boix, E. (2018) Immune Modulation by Human Secreted RNases at the Extracellular Space. *Front Immunol*, **9**, 1012.
49. Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y.Y., Carre, C., Burdin, N., Visan, L., Ceccarelli, M., Poidinger, M. *et al.* (2019) RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep*, **26**, 1627-1640 e1627.
50. Shichino, S., Ueha, S., Hashimoto, S., Ogawa, T., Aoki, H., Wu, B., Chen, C.Y., Kitabatake, M., Ouji-Sageshima, N., Sawabata, N. *et al.* (2022) TAS-Seq is a robust and sensitive amplification method for bead-based scRNA-seq. *Commun Biol*, **5**, 602.
51. Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C. and Teichmann, S.A. (2016) Classification of low quality cells from single-cell RNA-seq data. *Genome Biol*, **17**, 29.

52. Lee, J., Hyeon, D.Y. and Hwang, D. (2020) Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med*, **52**, 1428-1442.
53. Moses, L. and Pachter, L. (2022) Museum of spatial transcriptomics. *Nat Methods*, **19**, 534-546.
54. Janesick, A., Shelansky, R., Gottscho, A.D., Wagner, F., Rouault, M., Beliakoff, G., de Oliveira, M.F., Kohlway, A., Abousoud, J., Morrison, C.A. *et al.* (2022) High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue. *bioRxiv*, 2022.2010.2006.510405.

Figure legends

Figure 1: Experimental design and key differences between platforms. A. Experimental workflow of this study: PyMT tumours were digested as per (32) and after isolation of viable cells by MACS, cell suspensions were processed using Drop-seq, BD Rhapsody (whole transcriptome analysis) or 10x Chromium v3 platforms. To replicate damaged samples, tumours were also left overnight at 4ºC which reduced cell viability pre-clean up (Simulated low-quality). **B.** Table comparing the main difference among the three scRNAseq platforms used in this study. RPE: random priming and extension

Figure 2: Measurements of the technical quality, sensitivity and reproducibility among scRNAseq platforms. A. Violin plots showing the number of UMIs (nCounts) and genes (nGenes) detected per cell, as well as the percentage of mitochondrial content per cell in 10x Chromium (blue), Dropseq (red) and BD Rhapsody (green) each using two biological replicates (r1, r2). **B.** Principal component (PC) plot sowing the total variability identified in each system measured as the standard deviation at increasing PCs. **C.** Cluster tree showing the phylogenic relationship of clusters in each platform at increasing resolutions. **D.** UMAP plots showing the two replicates in each platform. **E.** Heatmap showing the Pearson correlation coefficient of gene expression between the platforms and replicates**. F.** Principal Component Analysis (PCA) plot of the pseudobulk RNAseq in each replicate and platform.

Figure 3: Identification of the main cell lineages in tumours by each scRNAseq platform. A UMAP plot of all integrated samples using the RPCA method with 5 anchors parameter. **B.** UMAP plot showing the three main cell lineages: epithelial, immune and stroma cells identified in all cells from all replicates and platforms after integration. **C**. Percentage of cells from each main cell lineage. The box plot represents the percentage of each cell lineage identified by flow cytometry and the colour dots represent each of our

samples from the single cell dataset. **D.** UMAP plots showing the cell types assigned using label transfer analysis of the PyMT reference previously published (29), and split based on the platform. **E**. Bar plot of the percentage of cells from each cell type per sample (left) and excluding the epithelial cells (right). **F**. Cell type percentages of each cell type per sample. The box plot represents the percentages found in the PyMT reference (29) and the colour dots represent each of our samples from our single cell dataset. **G**. Dot plot showing the gene expression comparison of the top cell lineage marker genes between platforms.

Figure 4: Cancer-epithelial and fibroblast subtype detection across platforms. A. UMAP projections of epithelial cells showing the unsupervised KNN clusters split by platform**. B** UMAP plot showing the epithelial cell types based on the annotated PyMT reference (29). (LP: luminal progenitor; HS: Hormone sensing) **C.** Bar plot of the percentage of cells from each epithelial cell type per sample. **D.** Gene set Enrichment analysis of the marker genes from cluster 5 versus the rest of the luminal progenitor cells (GO: Gene Ontology. BP: Biological Processes). **E.** UMAP visualization of the number of genes (n_Feature_RNA) detected and the percentages of mitochondrial genes (percent.mt) per cell in each platform. **F.** UMAP projections showing the fibroblast subtypes divided per platform. **G.** Bar plot representation of the percentage of each fibroblast subtype per sample**.** (iCAF: immune cancer-associated fibroblasts; ECM_CAFs: extracellular remodelling cancer-associated fibroblasts

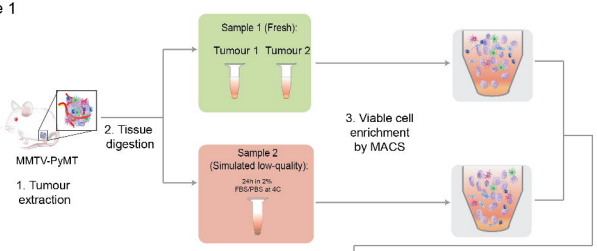
Figure 5: Analysis of the ambient noise in challenging samples in 10x Chromium and BD Rhapsody. A. Cell viability measured by flow cytometry as the percentage of DAPI negative cells in digested fresh tumours and simulated low-quality tumour samples (24h). **B.** Log fold change of the cell number of each cell type between the low-quality tumour samples and fresh samples. **C.** Line plot showing the changes in percentage of cells for each cell type between fresh tumours and low-quality tumours (24h). **D.** UMAP visualizations of the noise ratio, percentages of mitochondrial genes (percent.mt) and the number of genes (nGene) detected. Cells marked in red are considered empty barcodes based on the scAR model. **E.** Violin plots showing noise ratio (left) and percentage of mitochondrial content (right) per cell in each condition. Statistical analysis of the comparisons between fresh (F) and low-quality (24h) samples in each platform per cell type was performed using paired t-test adjusted with the Bonferroni method, ****adjusted p-value<0.001; ns = not significant. **F.** Violin plots split by cell type of the noise ratio (top) and the number of genes (bottom) detected in each cell.

Statistical analysis of the noise ratio comparisons per cell type between fresh and 24h samples in each platform was performed using paired t-test adjusted with the Bonferroni method, **** adjusted p-value <0.001; ns = not significant. One-way ANOVA was also performed to compare the noise ratio distribution across all cell types in the fresh samples in 10x Chromium (p-value=0.4903), or BD Rhapsody (p-value=2.2e-16). **G.** Violin plots for the gene expression levels of the top 6 most abundant genes in the ambient profile per cell type in each condition.

Figure 6: Performance evaluation for the detection of human granulocytes between 10x Chromium and BD Rhapsody. A. UMAPs of cell types found in human whole blood and bone marrow samples processed with 10x Chromium or BD Rhapsody using the transcriptome for dimension reduction. Cell barcodes were filtered using either Cell Ranger or BD Rhapsody pipelines. **B.** Bar plot showing the number of cells found in each cell type per condition. **C.** Bar plot showing the number of cells found in each cell type in 10x Chromium after custom filtering based on a minimum of 10 Protein UMI counts. **D.** Bar plot of the percentage of cells from each cell type per condition using the same custom threshold than panel C for 10x Chromium. **E.** UMAP plots of the human datasets processed with 10x Chromium and filtered based on custom threshold where the dimensional reduction was done using either the protein data (left) or the RNA data (right) **F.** Violin plots of the number of RNA and Protein UMI counts in each cell split by cell type. The red dotted line highlights a 100 counts threshold. **G.** Noise ratio in each cell using either from the RNA or the protein data matrix and split by cell type. (BM: bone marrow; WB: whole blood)

Figure 1

A



B

4. Single cell capture and sequencing

scRNAseq Platform	10X Chromium	Dropseq	BD Rhapsody
Cell Isolation	Microfluidics	Microfluidics	Nanowells
Capture Bead Type	Gel bead	Solid bead	Magnetic bead
2nd Strand cDNA	TSO	TSO	RPE
Full length cDNA	Yes	Yes	No
In droplet/well Reactions	1. Cell Lysis 2. RNA capture 3. RT (TSO)	1. Cell Lysis 2. RNA capture	1. Cell Lysis 2. RNA capture
Outside droplet/well Reactions	1. Clean up 2. cDNA amplific. 3. Clean up 4. Fragmentation 5. Library prep	1. RT and TSO 2. Clean up 3. cDNA amplific. 4. Clean up 5. Fragmentation 6. Library prep	1. Washes 2. RT (RP) 3. Exo 4. RPE 5. Clean up 6. cDNA amplific. 7. Library prep
Computational Analysis	Cell Ranger 3.0	Dropseq core computation protocol	Seven Bridges Genomics

Figure 2

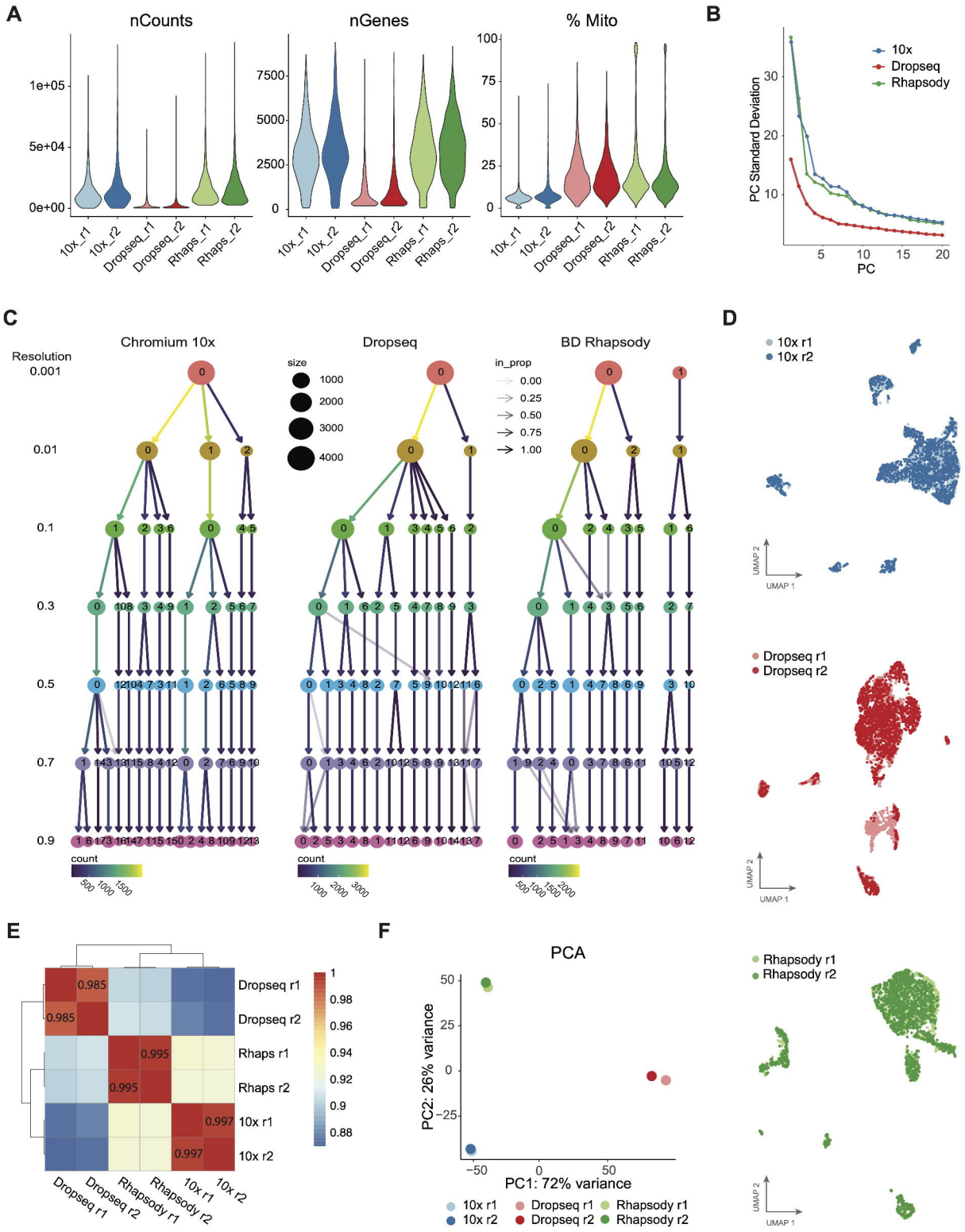

A
nCounts
nGenes
% Mito
10x_r1
10x_r2
Dropseq_r1
Dropseq_r2
Rhaps_r1
Rhaps_r2
B
10x
Dropseq
Rhapsody
PC Standard Deviation
PC
C
Chromium 10x
Dropseq
BD Rhapsody
Resolution
size
in_prop
count
D
10x r1
10x r2
Dropseq r1
Dropseq r2
Rhapsody r1
Rhapsody r2
UMAP 1
UMAP 2
E
Dropseq r1
Dropseq r2
Rhaps r1
Rhaps r2
10x r1
10x r2
Rhapsody r1
Rhapsody r2
F
PCA
PC2: 26% variance
PC1: 72% variance

Figure 3

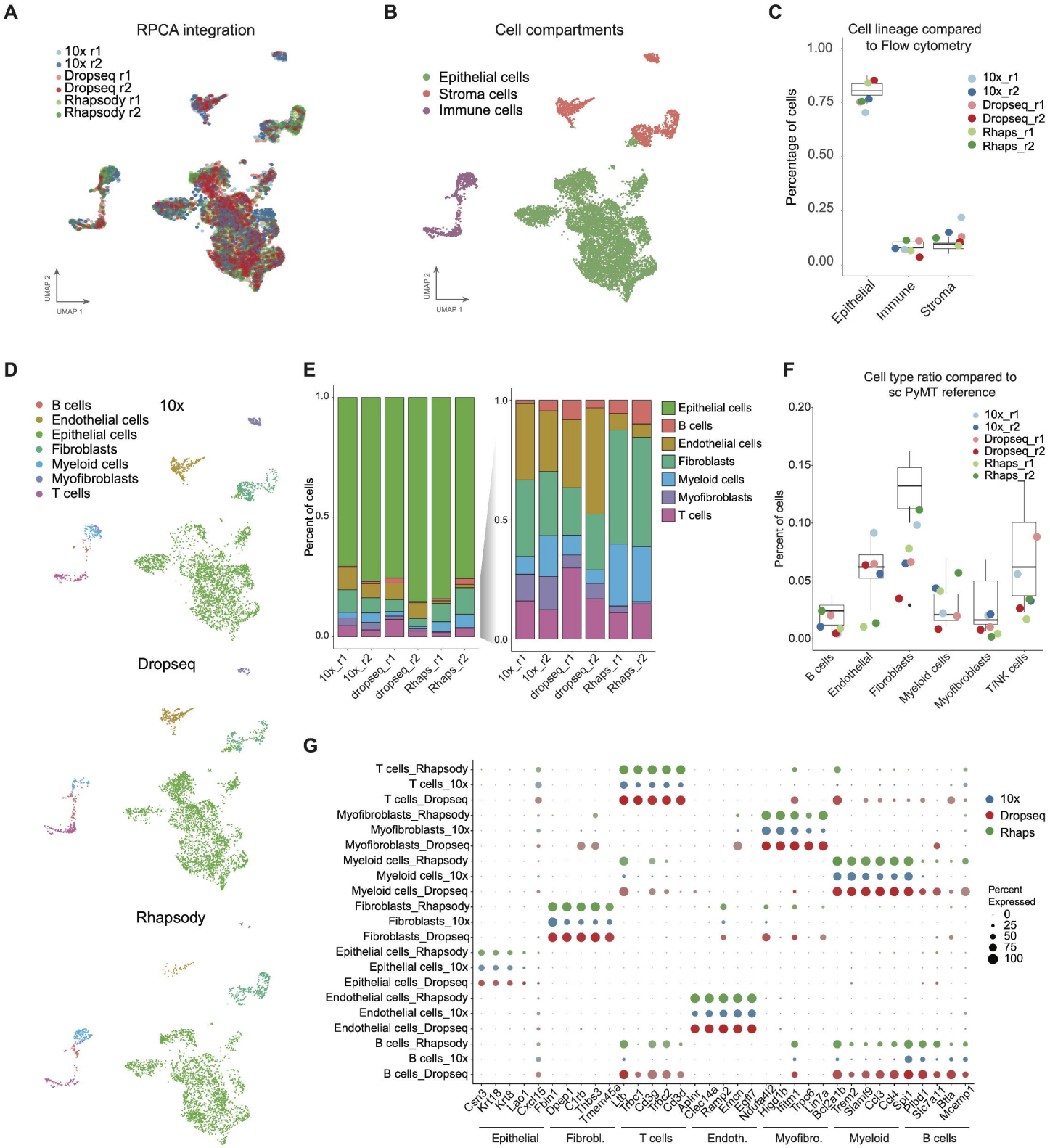

A
RPCA integration
10x r1
10x r2
Dropseq r1
Dropseq r2
Rhapsody r1
Rhapsody r2
UMAP 2
UMAP 1
B
Cell compartments
Epithelial cells
Stroma cells
Immune cells
UMAP 2
UMAP 1
C
Cell lineage compared to Flow cytometry
Percentage of cells
1.00
0.75
0.50
0.25
0.00
10x_r1
10x_r2
Dropseq_r1
Dropseq_r2
Rhaps_r1
Rhaps_r2
Epithelial
Immune
Stroma
D
B cells
Endothelial cells
Epithelial cells
Fibroblasts
Myeloid cells
Myofibroblasts
T cells
10x
Dropseq
Rhapsody
E
Percent of cells
1.0
0.5
0.0
10x_r1
10x_r2
dropseq_r1
dropseq_r2
Rhaps_r1
Rhaps_r2
Epithelial cells
B cells
Endothelial cells
Fibroblasts
Myeloid cells
Myofibroblasts
T cells
F
Cell type ratio compared to sc PyMT reference
Percent of cells
0.20
0.15
0.10
0.05
0.00
10x_r1
10x_r2
Dropseq_r1
Dropseq_r2
Rhaps_r1
Rhaps_r2
B cells
Endothelial
Fibroblasts
Myeloid cells
Myofibroblasts
T/NK cells
G
T cells_Rhapsody
T cells_10x
T cells_Dropseq
Myofibroblasts_Rhapsody
Myofibroblasts_10x
Myofibroblasts_Dropseq
Myeloid cells_Rhapsody
Myeloid cells_10x
Myeloid cells_Dropseq
Fibroblasts_Rhapsody
Fibroblasts_10x
Fibroblasts_Dropseq
Epithelial cells_Rhapsody
Epithelial cells_10x
Epithelial cells_Dropseq
Endothelial cells_Rhapsody
Endothelial cells_10x
Endothelial cells_Dropseq
B cells_Rhapsody
B cells_10x
B cells_Dropseq
Csn3
Krt16
Krt8
Lao1
Cxcl15
Fbln1
Dpep1
C1rb
Thbs3
Tmem45a
Ltb
Tnfrc1
Cd3g
Tnfrc2
Cd3d
Aplnr
Clec14a
Ramp2
Emcn
Egfl7
Ndufa4l2
Higd1b
Ifitm1
Trpc6
Lin7a
Bcl2a1b
Trem2
Slamf9
Ccl3
Ccl4
Spp1
Plbd1
Slc7a11
Bta
Mcemp1
Epithelial
Fibrobl.
T cells
Endoth.
Myofibro.
Myeloid
B cells
10x
Dropseq
Rhaps
Percent Expressed
0
25
50
75
100

Figure 4

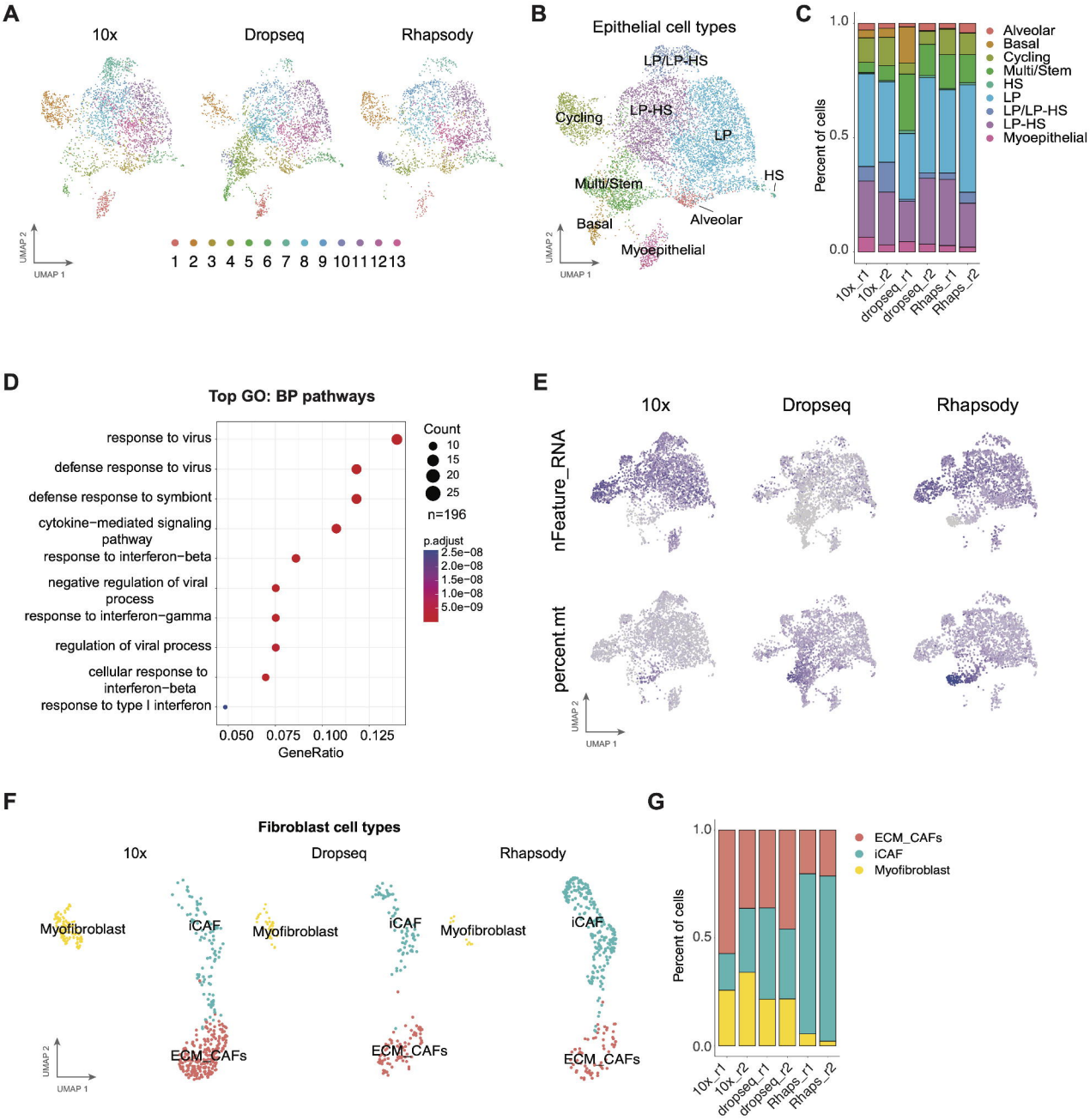

A
10x
Dropseq
Rhapsody
UMAP 2
UMAP 1
1 2 3 4 5 6 7 8 9 10 11 12 13
B
Epithelial cell types
LP/LP-HS
Cycling
LP-HS
LP
HS
Multi/Stem
Basal
Alveolar
Myoepithelial
UMAP 2
UMAP 1
C
Percent of cells
1.0
0.5
0.0
Alveolar
Basal
Cycling
Multi/Stem
HS
LP
LP/LP-HS
LP-HS
Myoepithelial
10x_r1
10x_r2
dropseq_r1
dropseq_r2
Rhaps_r1
Rhaps_r2
D
Top GO: BP pathways
response to virus
defense response to virus
defense response to symbiont
cytokine−mediated signaling pathway
response to interferon−beta
negative regulation of viral process
response to interferon−gamma
regulation of viral process
cellular response to interferon−beta
response to type I interferon
0.050
0.075
0.100
0.125
GeneRatio
Count
10
15
20
25
n=196
p.adjust
2.5e−08
2.0e−08
1.5e−08
1.0e−08
5.0e−09
E
10x
Dropseq
Rhapsody
nFeature_RNA
percent.mt
UMAP 2
UMAP 1
F
Fibroblast cell types
10x
Dropseq
Rhapsody
Myofibroblast
iCAF
ECM_CAFs
UMAP 2
UMAP 1
G
Percent of cells
1.0
0.5
0.0
ECM_CAFs
iCAF
Myofibroblast
10x_r1
10x_r2
dropseq_r1
dropseq_r2
Rhaps_r1
Rhaps_r2

Figure 5

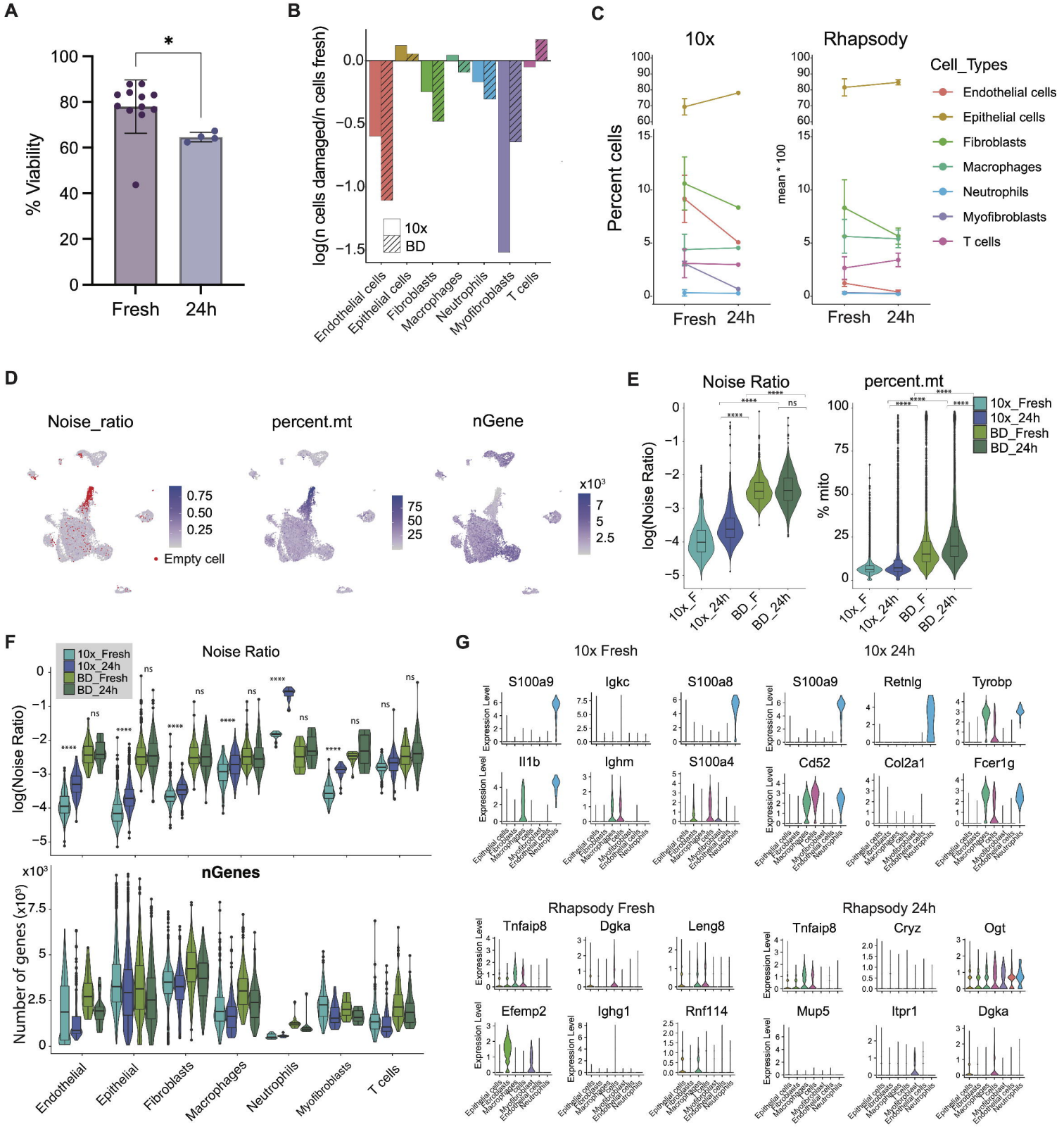
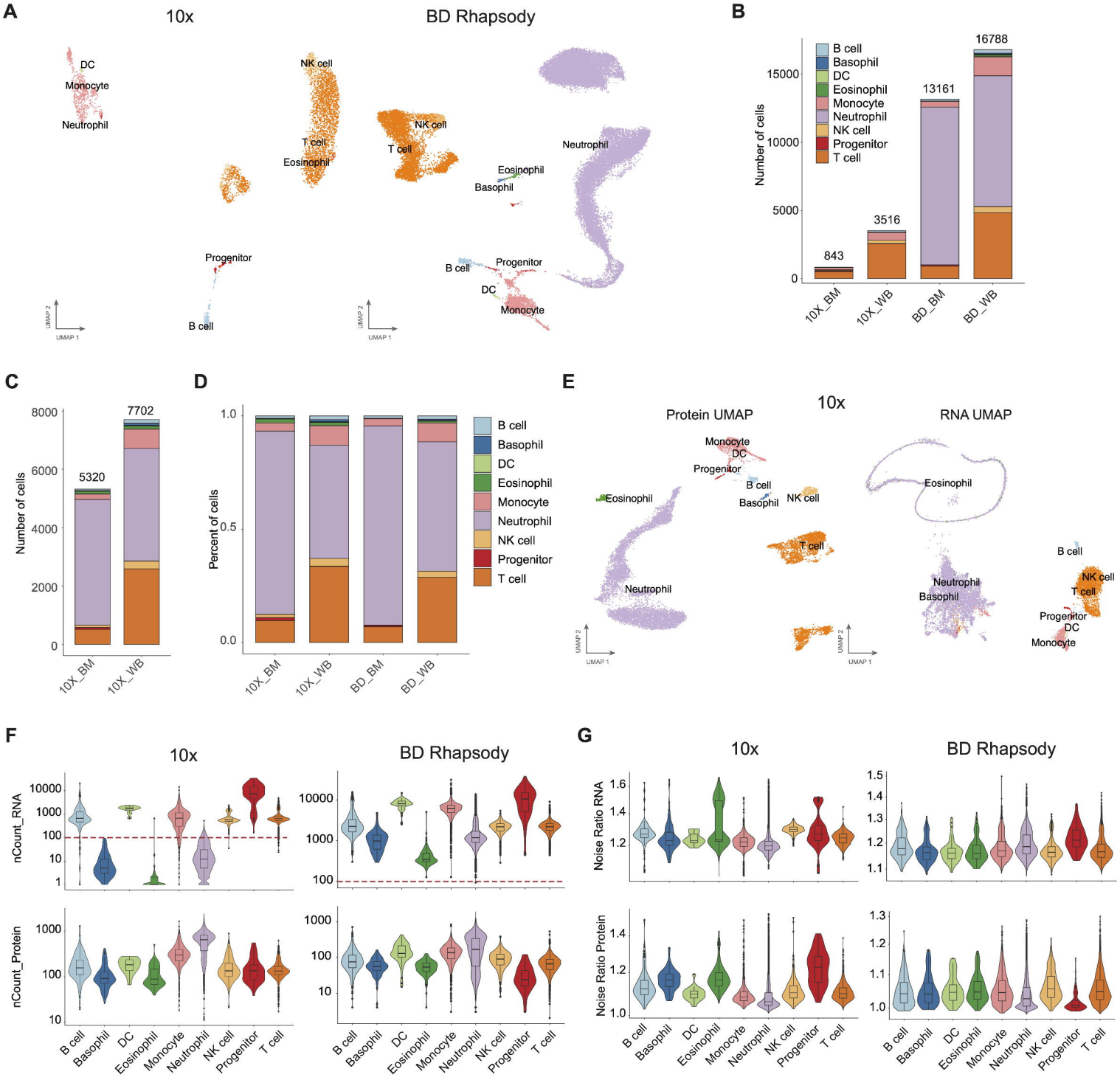
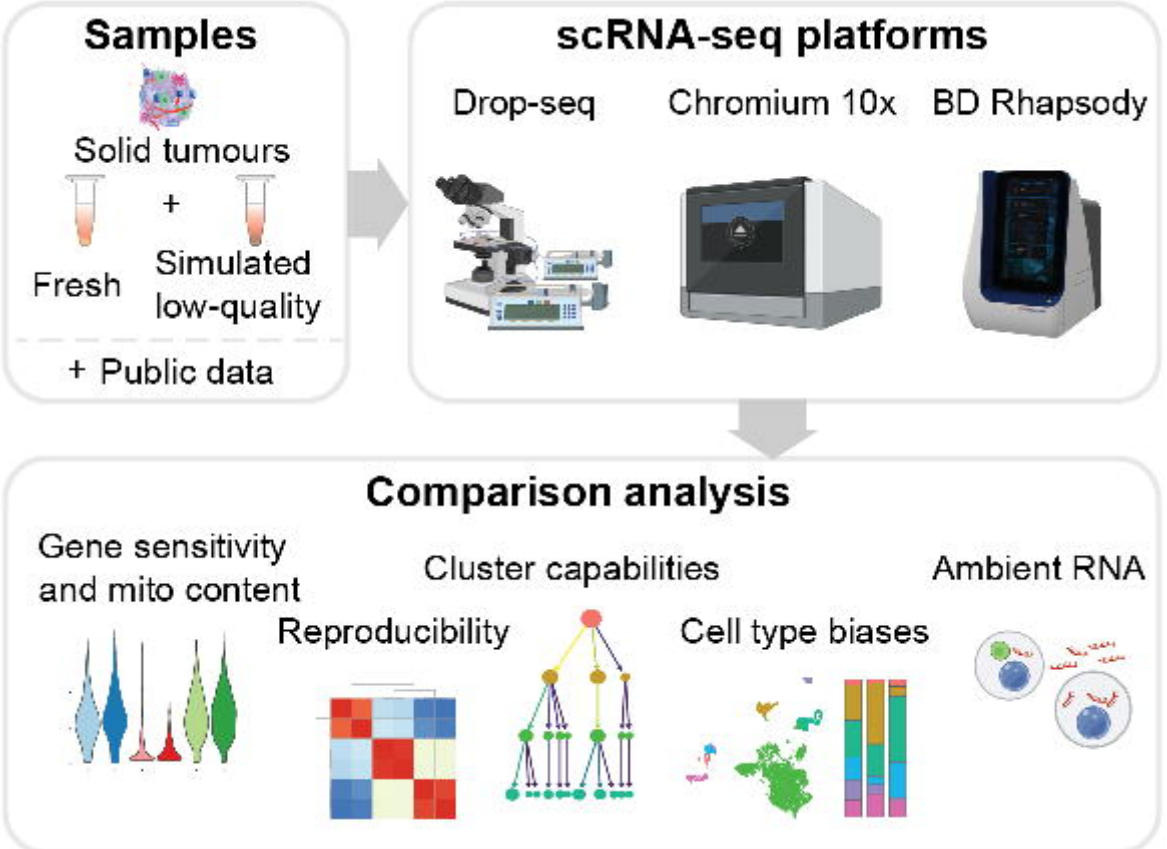

A
% Viability
Fresh
24h
B
log(n cells damaged/n cells fresh)
10x
BD
Endothelial cells
Epithelial cells
Fibroblasts
Macrophages
Neutrophils
Myofibroblasts
T cells
C
10x
Rhapsody
Percent cells
mean * 100
Cell_Types
Endothelial cells
Epithelial cells
Fibroblasts
Macrophages
Neutrophils
Myofibroblasts
T cells
D
Noise_ratio
percent.mt
nGene
Empty cell
E
Noise Ratio
percent.mt
log(Noise Ratio)
% mito
10x_Fresh
10x_24h
BD_Fresh
BD_24h
F
Noise Ratio
log(Noise Ratio)
nGenes
Number of genes (x10³)
Endothelial
Epithelial
Fibroblasts
Macrophages
Neutrophils
Myofibroblasts
T cells
G
10x Fresh
10x 24h
S100a9
Igkc
S100a8
S100a9
Retnlg
Tyrobp
Il1b
Ighm
S100a4
Cd52
Col2a1
Fcer1g
Rhapsody Fresh
Rhapsody 24h
Tnfaip8
Dgka
Leng8
Tnfaip8
Cryz
Ogt
Efemp2
Ighg1
Rnf114
Mup5
Itpr1
Dgka
Expression Level

Figure 6


A
10x
BD Rhapsody
DC
Monocyte
Neutrophil
NK cell
T cell
Eosinophil
Progenitor
B cell
Basophil
UMAP 1
UMAP 2
B
Number of cells
843
3516
13161
16788
10X_BM
10X_WB
BD_BM
BD_WB
C
Number of cells
5320
7702
D
Percent of cells
E
Protein UMAP
RNA UMAP
F
nCount_RNA
nCount_Protein
G
Noise Ratio RNA
Noise Ratio Protein


Samples
Solid tumours
+
Fresh
Simulated low-quality
+ Public data
scRNA-seq platforms
Drop-seq
Chromium 10x
BD Rhapsody
Comparison analysis
Gene sensitivity and mito content
Cluster capabilities
Ambient RNA
Reproducibility
Cell type biases