Research article

Comparative analysis of 10X Chromium vs. BD Rhapsody whole transcriptome single-cell sequencing technologies in complex human tissues

Stefan Salcher [a,1], Isabel Heidegger [b,1], Gerold Untergasser [a,1], Georgios Fotakis [c], Alexandra Scheiber [a], Agnieszka Martowicz [a], Asma Noureen [c], Anne Krogsdam [c], Christoph Schatz [d], Georg Schäfer [d], Zlatko Trajanoski [c], Dominik Wolf [a], Sieghart Sopper [a,**,2], Andreas Pircher [a,*,2]

[a] *Department of Internal Medicine V, Haematology & Oncology, Comprehensive Cancer Center Innsbruck (CCCI) and Tyrolean Cancer Research Institute (TKFI), Medical University of Innsbruck (MUI), Innsbruck, Austria*
[b] *Department of Urology, Medical University of Innsbruck, Innsbruck, Austria*
[c] *Biocenter, Institute of Bioinformatics, Medical University of Innsbruck, Austria*
[d] *Department of Pathology, Medical University Innsbruck, Innsbruck, Austria*

ARTICLE INFO

ABSTRACT

The development of single-cell omics tools has enabled scientists to study the tumor microenvironment (TME) in unprecedented detail. However, each of the different techniques may have its unique strengths and limitations. Here we directly compared two commercially available high-throughput single-cell RNA sequencing (scRNA-seq) technologies - droplet-based 10X Chromium *vs.* microwell-based BD Rhapsody - using paired samples from patients with localized prostate cancer (PCa) undergoing a radical prostatectomy.

Although high technical consistency was observed in unraveling the whole transcriptome, the relative abundance of cell populations differed. Cells with low mRNA content such as T cells were underrepresented in the droplet-based system, at least partly due to lower RNA capture rates. In contrast, microwell-based scRNA-seq recovered less cells of epithelial origin. Moreover, we discovered platform-dependent variabilities in mRNA quantification and cell-type marker annotation. Overall, our study provides important information for selection of the appropriate scRNA-seq platform and for the interpretation of published results.

1. Introduction

In the past decade a variety of high-dimensional single-cell omics tools have been developed and optimized at exponential pace, providing unprecedented opportunities to deconvolute tissue composition in various research fields including the tumor

---

microenvironment (TME). This enabled the discovery of novel cell types with previously unknown functions across different diseases including various malignancies [1–6]. In contrast to other single-cell technologies such as fluorescence-based or mass spectral flow cytometry, this enabled the unbiased discovery of novel cell types and distinct functional states across multicellular tissues. Since the first description in 2009 [7], several scRNA-seq platforms have been introduced to refine scRNA-seq throughput, sensitivity, precision, and costs. These advances paved the way to broadly apply the technology to map cellular and molecular complexity of the TME and to characterize diverse cell types in cancer [5,6,8,9]. ScRNA-seq is used to quantify the individual cell-specific transcriptome at single-cell level. Gene expression levels are denoted by the sequenced reads and the unique transcriptomes individual cell types express, which are then presented as a data matrix. Hence, a typical scRNA-seq workflow would encompass single-cell isolation and capture, sample barcoding, mRNA reverse transcription, cDNA amplification and cDNA library preparation, next-generation sequencing (NGS), and computational data processing.

Different scRNA-seq strategies have been developed to generate either full-length cDNA (for full-length sequencing) or cDNA with a unique molecular identifier (UMI) at the 3′ or 5′ end (for 3′/5′ sequencing). Full-length sequencing protocols like SMART-seq [10], as well as the improved version SMART-seq2 [11], generate reads across the entire length of genes by template switching-based PCR. The high sensitivity to detect gene expression enables various downstream analyses of specific cell types, tissue composition, spliced transcript variants, isoforms, or allele-specific gene expression patterns [12]. Full-length sequencing is low-throughput, relatively expensive, and struggles with a gene length bias akin to bulk RNA-seq data. Genes with shorter lengths generally exhibit lower read counts and a higher drop-out rate [13]. "Molecular tagging" with UMIs [14] used in 3′/5′ sequencing mitigated biases from downstream PCR amplification and enables digital counting of absolute numbers of each transcript per individual cell. In addition, UMIs allow sample multiplexing to improve gene expression quantification and throughput, and thus reduce overall costs [12]. Although the sensitivity is inferior to full-length sequencing, tag-based sequencing methods are more suitable for quantification purposes (e.g. cell typing), this is particularly true when considering the simultaneous sampling of tens of thousands of cells [12,15,16].

The technology used to capture a single cell for RNA sequencing determines the number of isolated cells, whether there is a biased or unbiased selection of cells, the integrity and purity of the cells, and lastly the costs of the experiment. For biological samples containing only few cells, manual low throughput methods, such as laser capture microdissection (LCM), limiting dilution of cell suspensions, or manual cell picking with micromanipulators are applicable. To increase throughput, FACS is suitable to analyze, sort, and enrich single cells. Still, cell numbers remain limited and downstream library preparation is laborious. Thus, distinctive small-volume microfluidic droplet-based 3′/5′ sequencing technologies are primarily applied as an ultra-high-throughput, unbiased solution [17]. Currently, the prevalent droplet-based scRNA-seq systems are inDrop [18], Drop-seq [19], and the commercially available platform 10X Chromium (10X Genomics, USA) [20]. Each of these three technologies generates microfluidic droplets to capture single cells together utilizing on-bead primers with unique barcodes. Besides the droplet-based methods, Seq-well [21], Microwell-seq [22], Cyto-seq [23], as well as the commercialized platforms Singleron GEXSCOPE (Singleron Biotechnologies, China) and BD Rhapsody (Becton Dickinson, USA) [24] apply microwell arrays into which individual cells are loaded together with barcoded magnetic beads. Upon cell lysis, the mRNA content of each cell is captured, reverse transcribed into cDNA, and amplified for sequencing-library generation through oligo-dT priming procedures.

These tools are helping scientists to sequence and study RNA in unprecedented detail, but each technique has its own strengths and inherent limitations as described in several comparative studies [17,25–32]. In addition, our recent scRNA-seq study on lung cancer, in which we integrated scRNA-seq data from 19 studies and 21 datasets, comprising 505 samples from 298 patients [33] revealed that low-mRNA content cells are frequently dismissed due to technical issues. The integrated datasets were generated using various scRNA-seq platforms, including 10X Chromium and BD Rhapsody. Lung cancer tissue resident neutrophils, characterized by exceptionally low mRNA content, were predominantly detected in the dataset generated with BD Rhapsody, while they were underrepresented or even absent in datasets generated with 10X Chromium or other platforms [33]. Using this substantial advantage of microwell-based scRNA-seq, we recently also characterized previously unrecognized tissue resident low-mRNA content neutrophils in human livers [34].

These performance differences have been described using either rather artificial cell systems [25] or comparisons of data from different studies [33] but a rigorous side by side comparison of primary samples covering the complete cellular heterogeneity within the tissue has not been performed. Thus, we performed a systematic comparison of two commercially available high-throughput scRNA-seq technology platforms, i.e. the droplet-based 10X Chromium and the microwell-based BD Rhapsody platform, using paired surgically resected prostate cancer (PCa) and corresponding benign tissues, in order to derive the necessary information for the planning of scRNA-seq experiments and the interpretation of the results.

2. Results

2.1. Droplet- vs. microwell-based single-cell RNA sequencing of PCa and benign prostate tissues

To directly compare the capability to deconvolute the TME and benign tissue composition at single-cell resolution of the most abundant cancer in men, we dissociated six freshly isolated tissue samples from treatment-naïve localized PCa patients undergoing a radical prostatectomy (3 tumor tissues samples and 3 corresponding benign prostate tissues samples) and subjected the obtained single-cell suspensions to 3′WTA scRNA-seq analysis using the 10X Chromium as well as the BD Rhapsody platforms, respectively (the experimental setup is illustrated in Fig. 1A; selection criteria of tissue samples obtained from fresh radical prostatectomy (RP) specimens to perform scRNA-seq analysis have been described previously [35]; the detailed scRNAseq workflow of both platforms is shown in Fig. S1).

To investigate the capability of the 10X Chromium protocol, we generated and sequenced six WTA index libraries from a total of 48,000 cells, with 8000 cells per sample. After pre-processing the raw sequencing data using the 10X Chromium Cell Ranger software, we detected 29,484 cells (cells with >100 genes expressed in ≥3 cells), corresponding to a capture rate of ~60% (including multiplets; Fig. 1B).

In the BD Rhapsody workflow, we used the BD Rhapsody Single Cell Multiplexing kit (SMK) to process multiple samples simultaneously. Following the process of labeling single cells derived from both benign and tumor tissue using sample-tag antibody staining, we generated a separate sequencing library for each of the three individual patients. In contrast to 10X Chromium, the microwell-based BD Rhapsody workflow enables the quantification of single living cells captured together with a bead within the system. Hence, this detailed information about the quality and number of captured single cells, as well as the number of captured multiplets, enables a more precise calculation of the desired sequencing depth per cell, especially when multiple samples are combined in one sequencing run. Thus, it is possible to use a predefined number of cells for WTA index library preparation and to store remaining beads for later use. The capability of the BD Rhapsody platform for sample-multiplexing and the flexibility to adjust cell numbers significantly contribute to cost reduction in library preparation and optimization of subsequent sequencing expenses. From the 65,000 deployed single-cells, 32,000 living cells including 5.5% multiplets could be captured together with a bead. Sequencing resulted in 25,841 detected cell barcodes. Still, just 21,196 cell barcodes with sample-tag information could be recovered during sample demultiplexing after pre-processing of the raw sequencing data using the BD Rhapsody WTA Analysis Pipeline app in the cloud-based Seven Bridges Genomics environment. Approximately 10–15% of "undetermined" cell barcodes with no sample-tag information had to be excluded from downstream analyses. Overall, we could depict 18,360 cells (cells with >100 genes expressed in ≥3 cells), resulting in an effective cell capture rate of ~30% (excluding multiplets, Fig. 1B).
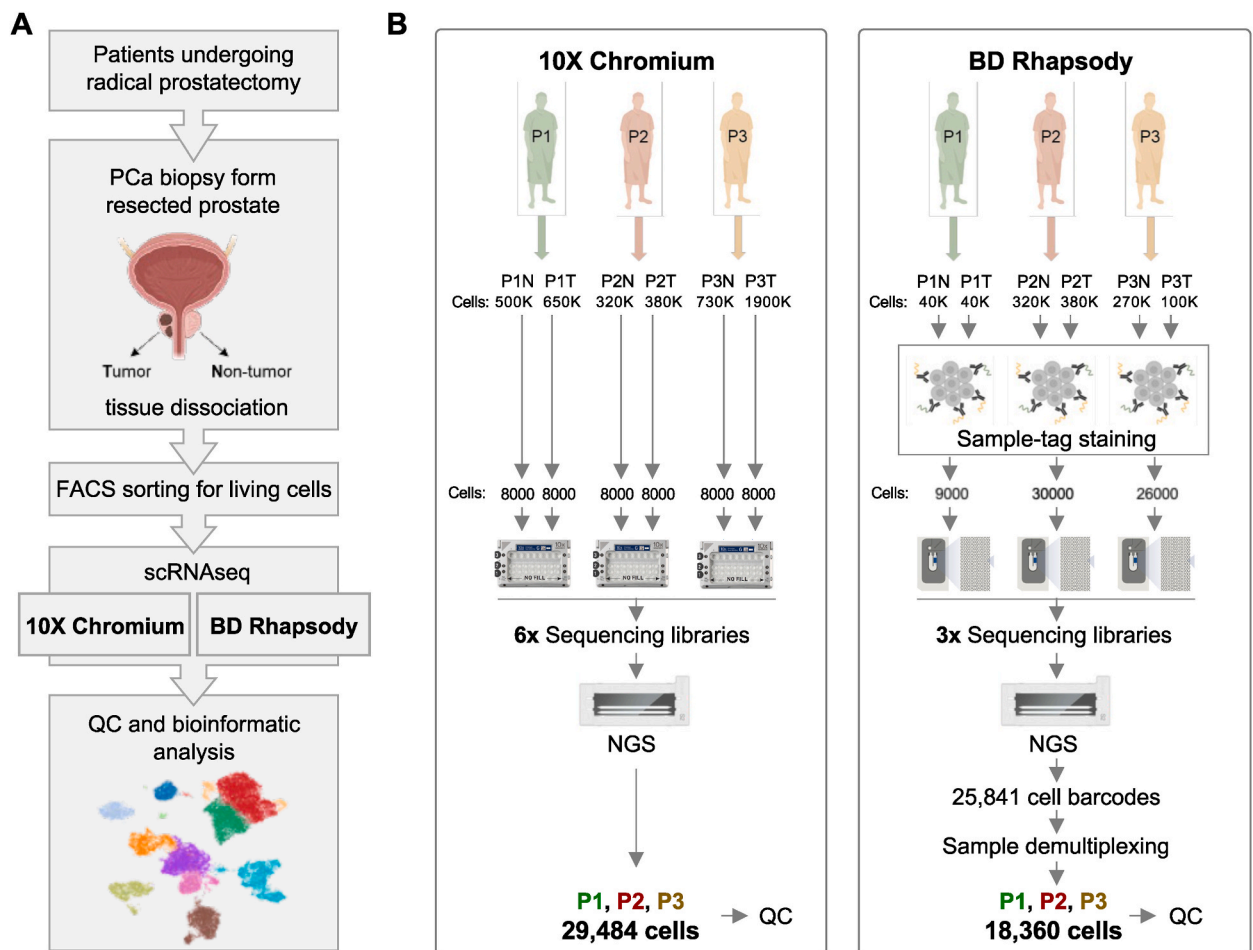

Fig. 1. Experimental setup. (A) Summary of the analysis workflow. (B) scRNA-seq datasets were generated from freshly isolated benign prostate (n = 3) and PCa (n = 3) tissues using the 10X Chromium and BD Rhapsody platforms, respectively. Cell numbers are shown for each step starting with number of isolated cells from each sample used for the two platforms. In the 10X Chromium dataset 29,484 cells (cells with >100 genes expressed in ≥3 cells) were detected and subjected to quality control processing. In the BD Rhapsody dataset 25,841 cell barcodes were identified and 21,196 cell barcodes with sample-tag information could be recovered during sample demultiplexing. Thereof, 18,360 cells (cells with >100 genes expressed in ≥3 cells) were subjected to quality control processing.
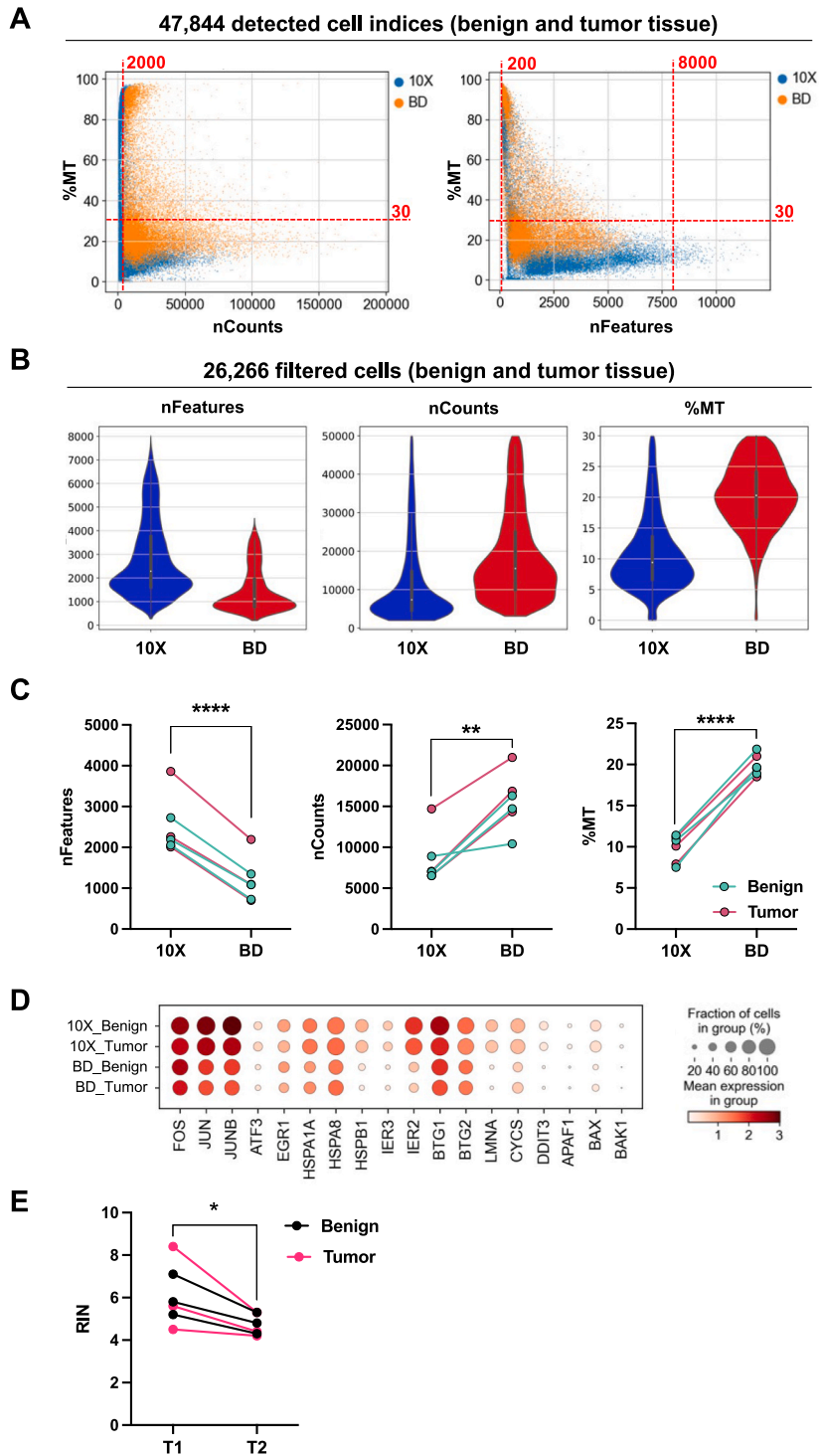
Fig. 2. QC metrics in datasets generated with 10X Chromium and BD Rhapsody. (A) Correlation of %MT with nCounts and nFeatures quality metrics in data generated with 10X Chromium and BD Rhapsody using cells expressing >100 genes (features). Applied cut-off values to filter for high quality cells are indicated (nCounts >2000, nFeatures >200 and < 8000, %MT < 30%). (B) nFeature, nCount, and %MT quality metrics in filtered cells derived from benign and PCa tumor tissues. (C) nFeature, nCount, and %MT levels in individual samples processed with 10X Chromium and BD Rhapsody (n = 6; benign n = 3, tumor n = 3). Paired *t*-test, **p ≤ 0.01, ****p ≤ 0.0001. (D) Expression of stress-related transcripts in 10X Chromium and BD Rhapsody data generated from benign prostate and PCa samples. (E) RNA quality (RIN) before (T1) and after (T2) the sample-tag staining procedure in freshly isolated lung cancer (NSCLC) tumor tissues (n = 6). Paired *t*-test, *p ≤ 0.05.

In summary, 29,484 cells (10,557 benign (~36%) and 18,927 tumor (~64%)) acquired with 10X Chromium as well as 18,360 cells (7145 benign (~39%) and 11,215 tumor (~61%)) obtained with BD Rhapsody (detected cells in each individual sample are detailed in Table S1), were subjected to quality control (QC) processing, a critical step during scRNA-seq data processing [36].

2.2. Microwell-based single-cell RNA sequencing results in elevated levels of mitochondrial transcripts

During QC user-defined thresholds for different metrics computed for each individual cell are applied to filter for "high-quality" cells to yield biologically meaningful results in subsequent down-stream analyses [37]. The three commonly used QC thresholds are the number of different genes detected in each cell, the total transcript count per cell (also known as the library size), as well as the ratio of reads mapped to mitochondrial DNA-encoded genes to the total number of reads mapped. These metrics have to be adjusted individually depending on the analyzed cell or tissue type, respectively [38].

Here, we filtered for cells with >200 and < 8000 detected genes (nFeatures), total transcript counts >2000 (nCounts), and <30% of mitochondrial transcripts (%MT) (Fig. 2A). 16,111 cells sequenced with 10X Chromium and 10,155 cells sequenced with BD Rhapsody passed the quality control step (26,266 cells in total; Table S1). Although the overall cell recovery rates (cells passing QC) of the processed PCa and benign prostate samples were well comparable between both scRNA-seq platforms (BD Rhapsody: 55.3%, 10X Chromium: 54.6%; metrics for each individual sample are detailed in Table S1), we found substantial differences in individual QC metrics between the two platforms (Fig. 2B, shown are 26,266 filtered cells). In all six individual samples the microwell-based BD Rhapsody protocol resulted in significantly less genes (nFeatures), a higher number of transcript counts per cell (nCounts), as well as a higher proportion of mitochondrial transcripts post quality control (%MT; Fig. 2C, Fig. S2).

The fraction of reads mapped to mitochondrial DNA-encoded genes (%MT) represents a general indicator for cell stress, and thus a high level of mitochondrial transcripts has been associated with stressed, apoptotic, low-quality cells [39–41]. Although the proportion of mitochondrial counts per cell was significantly higher in the dataset generated with BD Rhapsody (mean %MT: 20.2) compared to 10X Chromium (mean %MT: 10.8), the expression of diverse stress-related transcripts did not markedly differ (Fig. 2D). Cell stress-related transcripts, such as *IRE2*, *HSPB1*, *LMNA*, and *BAX*, were even expressed to a lesser extent in the BD Rhapsody dataset, in both benign prostate and PCa tissues, compared to 10X Chromium, respectively (Fig. 2D).

## 2.3. Sample-multiplexing by sample-tag antibody staining impairs RNA quality

In contrast to the droplet-based protocol, where each sample is processed separately, we utilized sample-tag antibodies in the BD Rhapsody workflow to enable multiplexing of multiple samples, including benign and tumor samples. A graphical representation of the detailed workflow from cell preparation to library construction for both platforms is outlined in Fig. S1.

Next, we investigated whether the overall lengthier protocol of the microwell-based platform including the additional sample-tagging procedure may affect RNA quality in freshly isolated cell suspensions. For RNA quality benchmarking we used an additional cohort of freshly resected human lung tumors (n = 3) as well as benign lung tissues (n = 3). Following tissue dissociation, we subjected the obtained single-cell suspensions to the sample-tag staining procedure (20 min RT, 3x washing by 5 min centrifugation at 400 rpm). The BD Rhapsody sample-tag protocol significantly impaired RNA quality (RNA Integrity Number (RIN)) in multiple single-cell suspensions (n = 6; Fig. 2E). Importantly, transcripts associated with RNA decay did not exhibit elevated expression levels in BD Rhapsody data (Fig. S3A), indicating that the RNA-degrading machinery is not actively induced during the scRNA-seq procedure. However, as the quality of RNA is a critical determinant of the reliability and accuracy of scRNA-seq data, any factors that compromise RNA quality may lead to biased results. Overall, these findings underscore the critical need for thorough evaluation and optimization of scRNA-seq protocols to ensure the reliability and validity of experimental results.

2.4. The microwell-based platform captures significantly more mRNA molecules per cell

The number of genes detected in each cell (nFeatures) is linked to the applied sequencing depth (mean reads per cell per gene) - a measure of the available sequencing capacity spent for a single sample. The three sequencing libraries generated with BD Rhapsody were sequenced on the NovaSeq S1 flowcell (1.6 billion single reads; Illumuna) resulting in a mean coverage of ~45,000 reads/cell. In contrast, the sequencing libraries prepared with the 10X Chromium system were sequenced on the NovaSeq S2 flowcell (4.1 billion single reads; Illumuna) resulting in a 1.6-fold higher mean coverage of ~72,000 reads/cell (samples obtained from four PCa patients (~64,000 cells) [35]; three out of these four patients (~48,000 cells) were processed in parallel with the BD Rhapsody platform and used for the comparative analysis). Consequently, the number of individual genes detected per cell was higher in the dataset obtained with 10X Chromium (Median nFeatures: BD Rhapsody ~1,100, 10X Chromium ~2,300, Fig. 2B and Table S1). The elevated number of transcripts detected per cell was consistent over individual samples (Fig. 2C, Fig. S2A and Table S1) but did not markedly affect the number of quality cells after filtering (nFeatures >200 and < 8000; Fig. 2A and Fig. S2B).

Remarkably, despite the lower sequencing depth, the detected number of mRNA molecules per cell (nCount RNA) was markedly higher in BD Rhapsody data compared to 10X Chromium data (Median nCounts: BD Rhapsody ~15,500, 10X Chromium: ~7350), indicating a subtantially better mRNA capture performance of the microwell-based platform (Fig. 2B). We observed this siginficant difference in detected mRNA molecules per cell between the two platforms in individual samples derived from PCa and benign tissues (Fig. 2C, Fig. S2, and Table S1). Thus, filtering for cells with a defined number of mRNA molecules, markedly reduced the number of high-quality cells predominantly in the 10X Chromium dataset (nCounts >2000; Fig. 2A), indicating superiority of the BD Rhapsody platform in depicting cells with low-mRNA content.
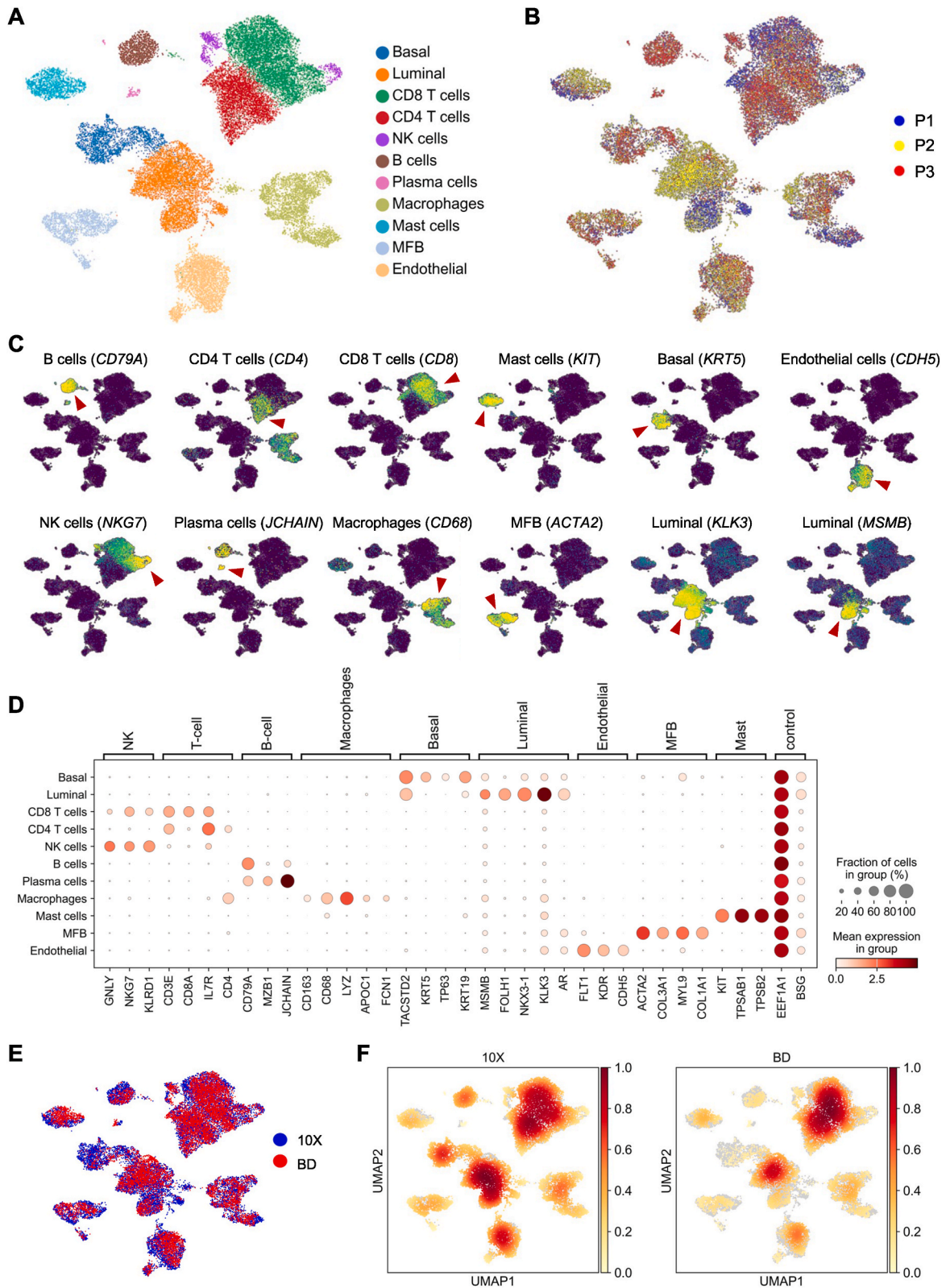
(caption on next page)

Fig. 3. Prostate cancer tumor microenvironment revealed by 10X Chromium and BD Rhapsody. (A) Uniform manifold approximation and projection (UMAP) plot of 26,266 high-quality cells, color-coded by cell type. (B) UMAP plot colored by cells derived from individual patients. (C) UMAP plots colored for the expression of indicated celltype-specific marker genes. (D) Gene expression levels of cell type-specific markers. (E) UMAP plot colored by cells derived from datasets generated with 10X Chromium or BD Rhapsody. (F) UMAP plots showing the cell-density in datasets generated with 10X Chromium or BD Rhapsody. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2.5. Prostate cancer tumor microenvironment mapping by microwell- and droplet-based single-cell RNA sequencing

Next, we performed an extensive assessment of the degree of technical biases and the efficacy of the protocols in accurately characterizing distinct cellular populations. Following quality control and filtering procedures, a total of 26,266 PCa and corresponding benign cells (10X Chromium: 16,111 cells; BD Rhapsody: 10,155 cells) were analyzed in terms of their transcriptomes. Using characteristic canonical cell markers, 11 cell clusters were annotated according to the respective cell type (Fig. 3A) in the 3 analyzed patients (Fig. 3B, individual samples are shown in Fig. S4A). We could delineate B cells (*CD79A*), CD4$^+$ T cells (*CD4*), CD8$^+$ T cells (*CD8*), NK cells (*NKG7*), plasma cells (*JCHAIN*), macrophages & monocytes (*CD68*), mast cells (*KIT*), myofibroblasts (MFB; *ACTA2*), endothelial cells (*CDH5*), basal/intermediate epithelial cells (*KRT5* as well as *KRT19*, *TP63* and low *AR* expression as defined previously in PCa [42]), as well as luminal epithelial cells (*KLK3*, *MSMB*) (Fig. 3C and D). The cells were derived from both benign prostate and PCa tumor tissues (Fig. S4B). Both platforms yielded high-resolution tissue profiles of the analyzed tissues, as illustrated in Fig. 3E (see also Figs. S4C–F). However, we observed a cell type-specific bias in cell recovery between the two platforms (Fig. 3F), which we investigated in more detail as described further below.

2.6. Molecule capture efficiency and sequencing library complexity

The overall higher number of different transcripts detected per cell in the 10X Chromium dataset compared to BD Rhapsody (Fig. 2B and C) was consistent across individual cell types (Fig. 4A). In addition to the previously described differences in sequencing depth, we reasoned that impaired RNA quality/RNA degradation (Fig. 2E) might result in less diversity in gene expression due to the degradation of more unstable mRNAs and, subsequently, in lower complexity of transcriptome libraries generated with the BD Rhapsody platform compared to 10X Chromium, respectively. Despite the lower sequencing depth, the median expression level of housekeeping genes (*EEF1A1*, *B2M*, *ACTB*) was notably higher in the BD Rhapsody dataset both in normalized (Fig. 4B) and in raw sequencing data (Fig. S5A) in all individual samples (Figs. S5B and S5C).

Using a depth-adjusted negative binomial model (DANB) we were able to correlate the frequency of dropout events (as a measure of dropout ratios) to the gene expression levels. Consistent with prior studies [25,29], genes exhibiting lower expression levels demonstrated higher dropout ratios (Fig. 4C). Due to variable library complexity, the dropout probability was found to be higher in the BD Rhapsody dataset (AUC = 5.19) when compared to the 10X Chromium dataset (AUC = 5.05).

Contrary, in the 10X Chromium dataset we found significantly higher expression levels of the Metastasis Associated Lung Adenocarcinoma Transcript 1 (*MALAT1*, also known as *NEAT2*) and the Nuclear-Enriched Abundant Transcript 1 (*NEAT1*; Fig. 4D). This marked difference in *MALAT1* and *NEAT1* expression between the two platforms was visible in all individual cell types of the prostate TME (Fig. S6A). Concordantly, our recently published Non-Small Cell Lung Cancer (NSCLC) scRNA-seq atlas [33] revealed that *MALAT1* and *NEAT1* are highly abundant in various cells of the lung cancer TME (Fig. S6B) and noticeable higher expressed in datasets generated with 10X Chromium compared to BD Rhapsody (Fig. S6C), respectively. *MALAT* and *NEAT1* are both nuclear retained long non-coding RNAs (lncRNAs). *MALAT* lacks a poly(A) tail [43], whereas NEAT1 does undergo some degree of poly-adenylation, which primarily applies to the NEAT1_1 isoform [44].

To further investigate the difference in the expression of *MALAT1* and *NEAT1* between the two platforms, we determined the gene body coverage using the IGV viewer (version 2.16.1) (Fig. 4E). BD Rhapsody exhibited an increased number of reads mapping to the gene body of both genes, which was exponentially higher than 10X Chromium. In the BD Rhapsody data, the reads mapped unevenly on the gene body of *MALAT1*, more specifically there appears to be a bias towards the genomic regions 65,498,957 - 65,506,516 with very few reads mapping to exon 1 of *MALAT1* (RefSeq gene: NR_144568.1). In contrast, in the 10X Chromium data, the bias appears to be towards exon 1 with the majority of reads mapping on the genomic region 65,497,738 - 65,502,629 of Chromosome 11. These marked differences were visible in all individual patients. The distribution of reads towards exon 1 seems to capture the expression of *MALAT1* with higher sensitivity, which may explain the observed difference in *MALAT1* expression between the two platforms. In the case of *NEAT1*, the gene coverage followed a similar pattern in both platforms. More specifically most of the reads mapped on the genomic regions 65,422,798 - 65,426,543 and 65438091 - 65,445,540 of Chromosome 11. It is important to note that the genomic region 65,422,798 - 65,426,543 of Chromosome 11 overlaps with exon 13 of the FRMD8 gene (RefSeq gene: XR_007062512.1), which may lead to the reads appearing as "multi-mappers" in the downstream analysis. Of note, in the case of BD Rhapsody, there appears to be an increased percentage of mismatches in the nucleotide sequences of the reads mapping to the reference genome of *NEAT1*, which may possibly affect the gene counts in the downstream analysis.

We next conducted a comparative analysis of the top 200 differentially expressed genes (DEGs) exemplarily in luminal epithelial cells and in endothelial cells, as identified in PCa and benign prostate tissues by 10X Chromium and BD Rhapsody (Fig. 5A). We observed a partially overlapping expression pattern of DEGs in both, luminal epithelial (139/200 overlapping DEGs) and endothelial cells (145/200 overlapping DEGs), between both platforms (Fig. 5B). To further assess the suitability of the protocols for characterizing
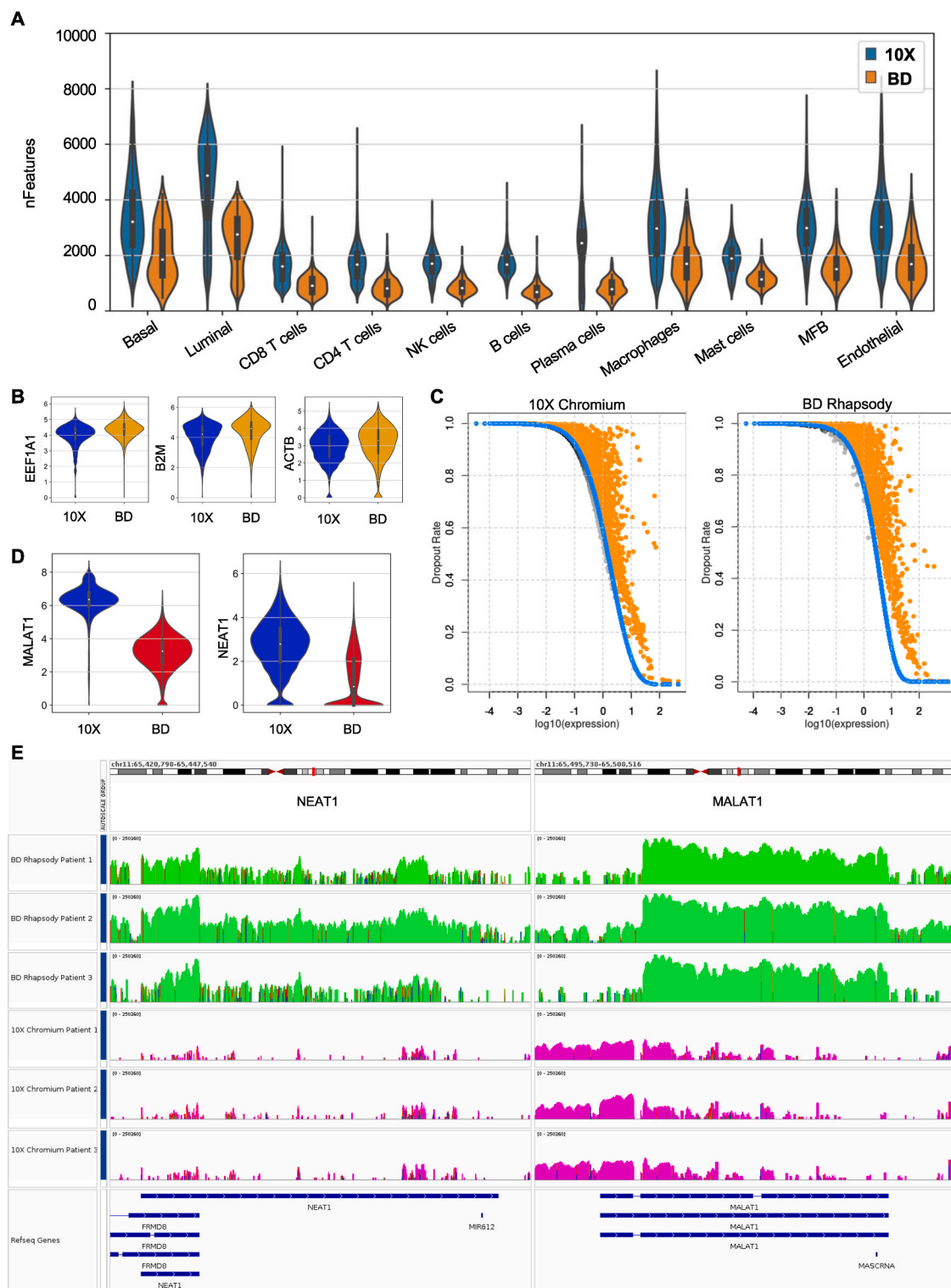
Fig. 4. Molecule detection efficiency and sequencing library complexity. (A) Number of genes detected per cell in individual cell types depicted by 10X Chromium and BD Rhapsody. (B) Gene expression levels of indicated house-keeping genes in datasets generated with 10X Chromium and BD Rhapsody. (C) Dropout ratios as a function of log10 expression for 10X Chromium (left) and BD Rhapsody (right). Orange dots represent the significant features under the DANB model (at 1% FDR) while gray dots represent the non-significant features. Blue dots represent the expected dropout probabilities as returned from the DANB model. (D) Gene expression levels of *MALAT1* and *NEAT1* in datasets generated with 10X Chromium and BD Rhapsody. (E) Gene body read coverage of *NEAT1* (left) and *MALAT1* (right). BD Rhapsody samples are represented with green

and 10X Chromium samples with purple colours. Data range is set to 0–250360 (the min - max read output of both platforms for the specific genomic regions) and a log scale is used to visualize the data and to allow comparisons. The blue and red bands indicate mismatches between the reads and the reference genome's nucleotide sequences. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

cell types, we evaluated their sensitivity in detecting population-specific expression markers and uncovered a significant disparity in their ability to discern marker genes. The prostate-specific-antigen (PSA, *KLK3*) as well as other prostate epithelial markers such as *KLK2*, *KLK4*, or *ACPP* were highly expressed in luminal epithelial cells in data generated with 10X Chromium as well as BD Rhapsody (Fig. 5C). However, particularly endothelial cell population markers were detected with varying sensitivity (Fig. 5C–Table S2). Both, the interferon alpha-inducible protein 27 (*IFI27*), previously described as marker for capillary endothelial cells of the human lung [45], as well as the key endothelial cell marker von-Willebrand factor (*VWF*), exhibited markedly higher expression in the 10X Chromium dataset compared to the BD Rhapsody dataset, respectively (Fig. 5C). One explanation for this discrepancy is that specific RNAs that are more sensitive to RNA degradation are lost during the BD Rhapsody protocol. However, stability of mRNAs is influenced by various factors, such as the presence of stabilizing or destabilizing elements within the mRNA sequence, interactions with RNA-binding proteins, and the cellular context. The current literature does not provide clear evidence to support the notion that these particular mRNAs are inherently unstable or rapidly degraded. Conversely, we observed a trend towards elevated expression levels of *CD34* (a well-known marker of progenitor cells of blood vessels) in endothelial cells depicted with the BD Rhapsody platform (Fig. 5C). The discrepancies in depicting endothelial cell population markers were visible in the raw sequencing data (Fig. S7A) as well as in the filtered and normalized data (Fig. 5D) in all individual samples (Fig. 5E). In concordance with the overall elevated expression levels of lncRNAs (Fig. 4D), the expression of *MALAT1* and *NEAT1* was also significantly higher in endothelial cells depicted with the 10X Chromium protocol (Fig. S7B, Table S2).

In summary, the comparison of the 10X Chromium and BD Rhapsody platforms revealed similarities in detecting differentially expressed genes (DEGs), while also exhibiting platform-dependent variability in detecting specific RNAs, including lncRNAs and cell type-specific population markers.

2.7. Platform-dependent cellular composition in single-cell RNA sequencing data

The overall lower mRNA count in the 10X Chromium dataset (Fig. 2B and C) may particularly affect detection and characterization of cells with low-mRNA content, such as various immune cells. Concordantly, we observed that CD4[+] T cells, CD8[+] T cells, and to a lesser extend NK cells were primarily affected by filtering for cells above a minimum mRNA amount in the 10X Chromium dataset (Fig. 6A).
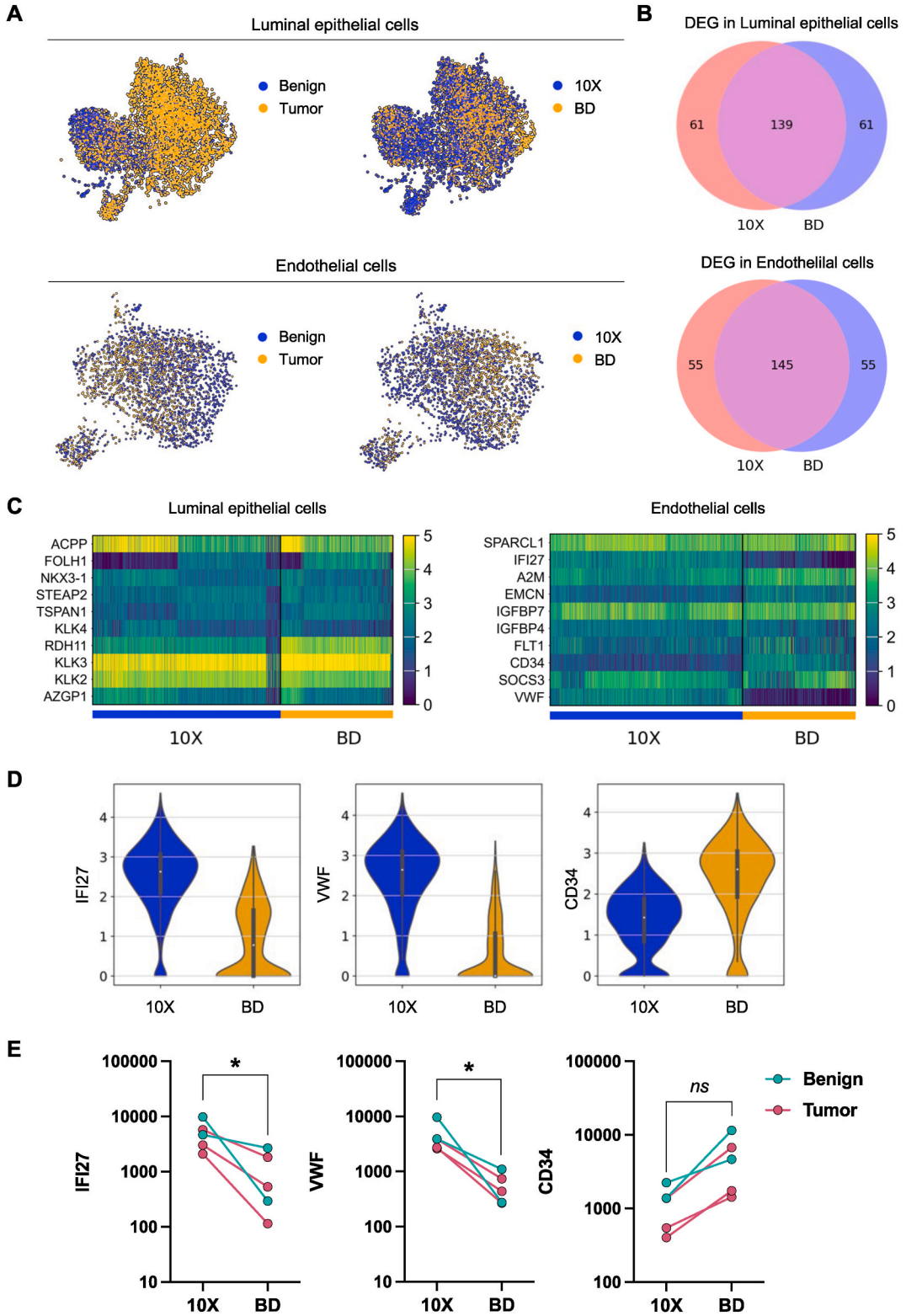
Consequently, CD4[+] T cells and CD8[+] T cells were detected at significantly smaller proportions by 10X Chromium compared to BD Rhapsody (Fig. 6B and C), in all individual samples (Fig. 6D). Additionally, we noted a trend towards higher NK cell numbers in the BD Rhapsody dataset (Fig. S8A). Conversely, epithelial cells (Fig. 6E) and myofibroblasts (Fig. 6F), were substantially better represented in the 10X Chromium dataset. Particularly basal epithelial cells were significantly better depicted by the 10X Chromium platform in all six analyzed samples (Fig. 6E). Primary prostate epithelial cells are highly sensitive to anoikis and cell stress, which may lead to their loss. Consequently, we observe a substantial increase in %MT expression in basal epithelial cells compared to other cell types in the BD Rhapsody dataset, which might at least in part reduce the cell number after filtering (Fig. S8B). Overall, the leukocyte fraction was particularly represented in BD Rhapsody data, whereas non-leukocytes, including epithelial cells and myofibroblasts were depicted at higher proportions by 10X Chromium (Fig. 6G).

PCa is generally deemed as immunologically 'cold' tumor with relatively low immune cell infiltration, as also demonstrated by IHC analysis showing a higher abundance of epithelial and stromal cells in all analyzed benign prostate and PCa tumor tissues compared to CD45[+] leukocytes (Fig. 6H). Contrary, both scRNA-seq protocols resulted in an overrepresentation of immune cells and an under-representation of the epithelial/stromal compartment. This phenomenon may be mostly attributable to the process of tissue disso-ciation, nonetheless, the composite data propose that the microwell-based scRNA-seq technique preferentially enriches for immune cells, while the droplet-based protocol recovers epithelial/stromal cells at comparatively elevated proportions (Fig. 6H).

3. Discussion

We systematically compared two of the most broadly used high-throughput scRNA-seq technologies (i.e., 10X Chromium and BD Rhapsody) using paired samples from surgically resected PCa and the respective healthy prostate tissues. Contrary to previous comparative analysis [17,25–32], we simultaneously processed a sufficient number of samples derived from complex tissues for a statistically sound comparison. Our analysis revealed differences between the two platforms in converting RNA molecules into sequencing libraries, impacting the reliability and accuracy of the resulting scRNA-seq data (identified key characteristics and major differences of both platforms are listed in Table S3). This information is essential for scientists asking for the right technology platform for their experimental research question: which platform is indeed appropriate to reliably detect expression profiles of single cells for the posed question.

First, our comparative analysis of QC metrics revealed that the droplet-based 10X Chromium platform yields a higher effective cell capture rate and allows a more sensitive detection of genes per cell, while the microwell-based BD Rhapsody protocol exhibits a substantially higher number of mRNA molecules detected per cell. We demonstrate that the BD Rhapsody platform captures a higher

A Luminal epithelial cells

B DEG in Luminal epithelial cells

Endothelial cells

DEG in Endothelial cells

C Luminal epithelial cells   Endothelial cells

D

E

(caption on next page)

Fig. 5. Platform-specific gene quantification and cell-type marker identification. (A) UMAP blots of luminal epithelial cells and endothelial cells colored by cells derived from benign and PCa tumor tissues as well as by cells derived from datasets generated with 10X Chromium and BD Rhapsody. (B) VENN diagram of top 200 DEG detected by 10X Chromium or BD Rhapsody in luminal epithelial cells (upper panel) and endothelial cells (lower panel). (C) Top expressed genes in luminal epithelial cells and endothelial cells in datasets generated with 10X Chromium and BD Rhapsody. (D) Gene expression levels of *IFI27*, *VWF*, and *CD34* detected in endothelial cells by 10X Chromium and BD Rhapsody. (E) Gene expression levels of *IFI27*, *VWF*, and *CD34* in individual samples. Each dot refers to a sample (benign or tumor tissue) with at least 40 endothelial cells in both 10X Chromium and BD Rhapsody groups. Paired *t*-test, *$p \leq 0.05$.

proportion of mitochondrial transcripts, typically associated with increased cell stress. However, we found no evidence of procedural stress as a contributing factor. A recent comparative analysis consistently reported elevated proportions of mitochondrial transcripts in BD Rhapsody, with no indications of cell damage [31]. One possible explanation for the elevated proportion of mitochondrial transcripts in the BD Rhapsody dataset could be a more efficient disruption of organelle membranes during the cell lysis step compared to the 10X Chromium protocol, which employs a markedly lengthier yet potentially weaker cell lysis procedure. Of note, each platform utilizes distinct bioinformatic approaches for pre-processing of the raw output files (i.e., barcode correction and UMI counting) that might influence QC metrics. Consequently, these observed variances emphasize the necessity for researchers to adapt QC filter metrics specifically to the datasets produced by different scRNA-seq platforms, ensuring accurate interpretation and data quality.

Second, we demonstrate that the sample-multiplexing capability in the BD Rhapsody workflow using sample-tags, markedly impairs RNA quality in freshly isolated single-cell suspensions derived from complex tissues. On the plus side, sample-multiplexing reduces technical bias caused by batch effects and substantially lowers sequencing library preparation costs. However, as RNA quality is a critical determinant of scRNA-seq data reliability and accuracy, the BD Rhapsody sample-tag staining procedure might compromise specific mRNAs, impair sequencing library complexity, and thus potentially affect the outcome of the scRNA-seq experiment.

Third, we report that varying library complexities and the observed transcript-specific mRNA capture efficacy of both protocols affect their ability to quantify gene expression levels and identify cell-type markers. Overall, the BD Rhapsody protocol demonstrated a higher gene dropout probability. The relative high abundance of mitochondrial and house-keeping transcripts in the BD Rhapsody data may indirectly contribute to elevated dropouts of genes exhibiting lower expression levels.

Although the vast majority of transcripts shows congruent expression pattern in different cell types with both platforms, our analysis also revealed substantial differences for certain genes, exemplarily shown for *IFI27*, *VWF* or *CD34*, putative marker genes of prostate endothelial cells. In turn, this discrepancy in amplifying certain types of RNA molecules, might significantly impact the underlying composition of depicted tissue profiles and affect the exploratory value of generated scRNA-seq datasets. Of note, we observed remarkably high expression levels of the lncRNAs *MALAT1* and *NEAT1* exclusively in 10X Chromium data. We could confirm this platform-specific differences in depicting lncRNAs in our recently published NSCLC scRNA-seq atlas. Gene coverage analysis revealed a pronounced 3′ end bias in BD Rhapsody read distribution for *MALAT1*, whereas the 10X Chromium platform exhibited a distinct bias towards exon 1 at the 5′ end of the gene. Thus, we speculate that a higher sensitivity of exon 1 read capture may account for the observed variation in *MALAT1* expression between the two platforms.

Overall, we advocate to consider potential platform-mediated gene expression bias when interpreting data or comparing results derived from different scRNA-seq protocols.

Finally, our direct comparative analysis corroborates that the BD Rhapsody protocol excels in characterizing cells with low-mRNA content, capturing leukocytes more effectively. This important benefit of the microwell-based platform is based on a significantly better mRNA capture efficiency. In line with our recent observation in lung cancer [33], the detected number of mRNA molecules per cell was markedly higher in BD Rhapsody data compared to 10X Chromium, respectively. Conversely, the droplet-based protocol performed noticeably better in depicting epithelial cells, particularly prostate basal epithelial cells, as well as myofibroblasts. Considering that the microwell-based BD Rhapsody platform is validated for cells smaller than 20 μm, we hypothesize that the loss of epithelial cells or myofibroblasts may be attributable to their larger size relative to leukocytes. Hence, processing single-cell suspensions derived from complex tissues using microwell-based platforms might result in a bead-exclusion phenomenon that results in loss of larger cell types. In accordance, our recent scRNA-seq studies on liver tissues processed with the BD Rhapsody platform, resulted in an underrepresentation of hepatocytes (typical size >20 μM), the primary liver cells [34]. Overall, our findings propose that the choice of the scRNA-seq protocol might substantially influence the composition of captured cell types.

In conclusion, our study emphasizes the importance of considering the distinct characteristics of scRNA-seq technologies. Researchers must carefully evaluate their research questions and select the most appropriate platform accordingly. Furthermore, we advocate to be cautious when comparing data generated by different platforms, as platform-dependent variability in detecting population-specific markers and distinctive cell subpopulations may lead to discrepancies in the interpretation of the results. In that direction, our results might lead to informed algorithms for better data integration of different platforms. Future research is necessary to elucidate the underlying reasons for these differences and to optimize scRNA-seq protocols for obtaining reliable and accurate data in various biological contexts.

4. Limitations of the study

While our results provide valuable insights into the performance of these scRNA-seq protocols, it is important to acknowledge several potential limitations. First, while our study is the first to use sufficient independent samples for a proper statistical analysis, the number is still not enough to validate smaller differences between both platforms such as the proportion of NK cells. Concomitant with decreasing sequencing costs, in the future, we may see comparative studies with more specimens. Second, as the field is advancing at a
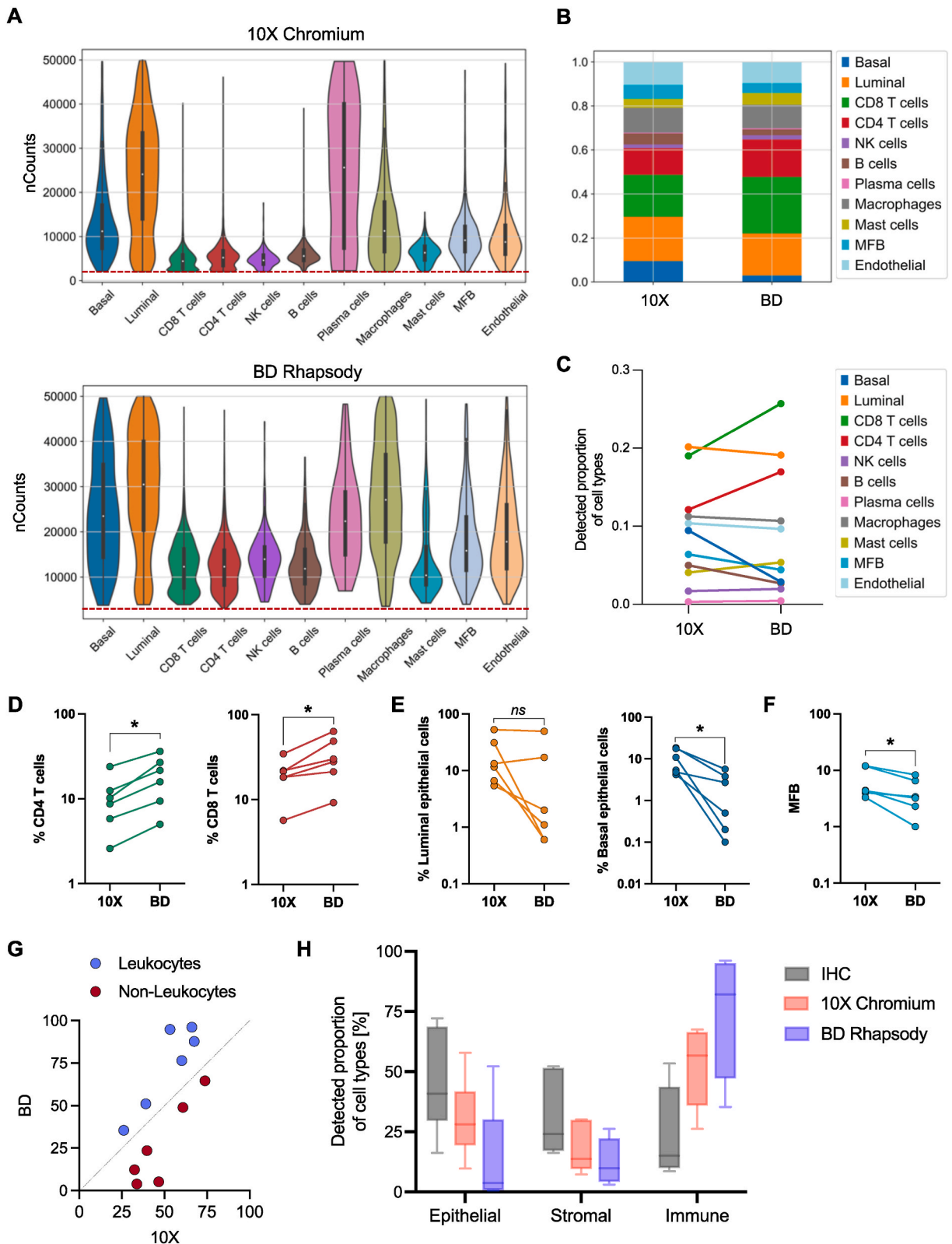
Fig. 6. Platform-dependent cellular composition in single-cell RNA sequencing data. (A) Number of captured transcripts (total counts) in individual cell types recovered with 10X Chromium (upper panel) and BD Rhapsody (lower panel). (B) Relative cell-type composition in benign prostate and PCa tissues from three individual patients in data generated with 10X Chromium and BD Rhapsody. (C) Detected proportion of cell-types in benign prostate and PCa tissues from three individual patients in data generated with 10X Chromium vs. BD Rhapsody. (D–F) The proportion of CD4 T cells and CD8 T cells (D), luminal and basal epithelial cells (E), and myofibroblasts (MFB) (F), depicted by 10X Chromium vs. BD Rhapsody in individual samples. Paired *t*-test, *$p \leq 0.05$. (G) The proportion of leukocytes and non-leukocytes depicted by 10X Chromium vs BD Rhapsody. (H) Proportions of epithelial, stromal, and CD45$^+$ immune cells in benign prostate and PCa tissues from three individual patients determined by IHC and by scRNA-seq using 10X Chromium or BD Rhapsody.

tremendous pace, newer versions of the two platforms with improved performance are already available. Still, we think that differences intrinsic to the technologies persist. In addition, our work might lay the ground for updated comparisons and future studies with more diverse samples could further enhance our understanding of scRNA-seq platform performance. Lastly, researchers should account for potential biases during data analysis and ensure their methods suit the specific platform and research question. Despite these potential limitations, our findings offer valuable insights into the performance characteristics of the 10X Chromium and BD Rhapsody platforms in the context of scRNA-seq analysis of prostate cancer tissues.

Ethics Declarations

- This study was reviewed and approved by the Institutional Review Board at the Medical University Innsbruck, Austria, with the approval numbers: 1017/2018; 1072/2018; AN214-0293,342/4.5
- All participants/patients (or their proxies/legal guardians) provided informed consent to participate in the study.

Data availability statement

Processed scRNA-seq data from this study has been deposited on Zenodo (https://zenodo.org/record/8063560) as listed in the key resources table. Raw data is not made available due to privacy concerns.

Star methods

Resource availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by corresponding authors, Sieghart Sopper (sieghart.sopper@i-med.ac.at) and Andreas Pircher (andreas.pircher@i-med.ac.at).

Experimental model and subject details

Human subjects
The local ethics committee (study code:1017/2018; 1072/2018) approved the use of tissue samples obtained from fresh radical prostatectomy (RP) specimens. Written informed consent is available from all patients. As described by our group, the malignancy or benignity of the tissue was confirmed within 1 h after surgery. Subsequently, tissue dissociation and FACS sorting of freshly isolated cells was performed to obtain a FACS-sorted PCa and corresponding benign prostate single-cell suspension [35].

Samples of NSCLC tumor tissues and matched benign lung tissues were obtained from surgical specimens of patients undergoing resection at the Department of Visceral, Transplant and Thoracic Surgery (VTT), Medical University Innsbruck, Austria, and in collaboration with the INNPATH GmbH, Innsbruck, Austria, after obtaining informed consent in accordance with a protocol reviewed and approved by the Institutional Review Board at the Medical University Innsbruck, Austria (study code: AN214-0293,342/4.5).

Method details

10X Chromium library preparation and sequencing
The droplet-based 10X Chromium system employs the GemCode technology, which utilizes gel beads in emulsion to achieve barcoding [46]. This involves combining a suspension of single cells with gel beads containing barcoded oligonucleotides and reagents for reverse transcriptase (RT) in an oil environment, resulting in the formation of nanoliter-scale droplets that facilitate cDNA synthesis. Subsequently, the droplets are pooled, dissolved, and used to create a sequencing library containing unique molecular identifiers (UMIs). The microfluidic chips used in this system can accommodate up to eight samples simultaneously, each containing up to 10,000 cells. Samples obtained from four PCa patients were converted to scRNA-seq libraries using the Chromium Next GEM Single-Cell 3′Kit v3.1 (10X Genomics) as described previously [35]. Three out of these four patients (~48,000 cells) were processed in parallel with the BD Rhapsody platform.

BD Rhapsody library preparation and sequencing
The BD Rhapsody platform uses a single cartridge with more than 200,000 microwells, wherein up to 30,000 individual cells are isolated together with UMI-barcoded magnetic mRNA capture-beads. Upon cell lysis the mRNAs are captured and retrieved together

with the beads out of the microwells. All beads are pooled, and subsequently the RT step is performed within a single tube. Freshly isolated, FACS-sorted single-cells were immediately processed with the BD Rhapsody scRNA-seq platform (BD Biosciences). The BD Single-Cell Multiplexing Kit (BD Biosciences) was used to combine and load two samples (tumor tissue and corresponding benign tissue) onto a single BD Rhapsody cartridge (BD Biosciences). Sample-tag staining was performed according to the manufacturer's protocol (sample-tag staining at room temperature for 20 min and washing by centrifugation at 400*g* for 5 min). Single-cell isolation in microwells (cell load: 20 min incubation at room temperature) with subsequent cell-lysis and capturing of poly-adenylated mRNA molecules with barcoded, magnetic captured-beads was performed according to the manufacturer's instructions. Beads were magnetically retrieved from the microwells, pooled into a single tube before reverse transcription. Unique molecular identifiers (UMIs) were added to the cDNA molecules during cDNA synthesis. Whole transcriptome amplification (WTA) and sample-tag sequencing libraries were generated according to the BD Rhapsody single-cell whole-transcriptome amplification workflow. The quantity and quality of the sequencing libraries was analyzed with the Qubit dsDNA HS (High Sensitivity) assay kit (Invitrogen) and the 4200 TapeStation (Agilent) system. Libraries were sequenced on the Novaseq 6000 system (Illumina) targeting a sequencing depth of 45.000 reads/cell.

Preprocessing and quality control of scRNA-seq data

Bioinformatic pre-processing of the obtained FastQ sequencing files was performed via the cloud-based Seven Bridges Platform environment (Seven Bridges Genomics) using the BD Rhapsody WTA Analysis Pipeline app making use of the sample-tag workflow to achieve csv-files with cell codes and gene list for each individual sample. Fastq sequencing data derived from 10X Chromium were mapped to the human genome (build GRCh38) using the CellRanger software (10X Genomics, v3.1) to achieve mtx, barcodes and genes files for each sample as described previously by our group [35]. Sample files were imported in Scanpy version 1.9.1 running with Python version 3.8 [47] and loaded into AnnData [48] for further processing with scverse tools to perform removal of empty cell beats and cell barcodes without sufficient captured transcripts (>100 genes, gene occurrence in >3 cells). Thereafter, all samples were imported into a single anndata object (h5ad) by the concatenation "outer" join function to maintain all platform/sample specific genes since objects had differing variables. Empty genes (variables) for each sample were filled up with 0 values.

Integration of scRNA-seq datasets

Quality control was performed using scanpy, only retaining cells with (I) between 200 and 8000 detected genes, (II) between 1000 and 50,000 transcripts, and (III) < 30% mitochondrial transcripts. The 2000 most highly variable genes (HVGs) were selected using scanpy's highly_variable_genes function with the options flavor = "seurat_v3″ and batch_key = "sample". Cell transcriptomes were embedded into a batch-corrected low-dimensional latent space using scVI [49,50] treating each sample as a separate batch. A neighborhood graph and uniform manifold approximation and projection (UMAP) embedding [51] were computed based on the scVI latent space. Cell types were annotated based on unsupervised clustering with the Leiden algorithm [52] and known marker genes specific for epithelial cells, endothelial cells, myofibroblasts and immune cell types.

Differential gene expression testing

For each cell type and patient, we summed up transcript counts for each gene that is expressed in at least 25% of cells and at least in 3 samples using decoupler-py [53]. Pseudo-bulk samples consisting of fewer than 1000 counts or 40 cells were discarded. Pseudo-bulked data for endothelial and luminal epithelial cells were used for further differential gene expression testing between the two different platforms using DESeq2 [54] which has been demonstrated to perform well [55].

Dropout ratio analysis

The dropout ratios were correlated to gene expression levels by using the depth-adjusted negative binomial model (DANB) from the R package M3Drop [56] in order to account for zeros resulting from insufficient sequencing depth. DANB assumes that each observation follows a negative binomial distribution, where the mean is proportional to both the mean expression of the specific gene and the relative sequencing depth of the corresponding cell.

Gene body coverage analysis

In order to visualize the gene body coverage, the BAM files produced in the pre-processing step were directly loaded in the IGV viewer [57] (version 2.16.1) as different tracks. The viewing panel was centered around the *NEAT1* and *MALAT1* genomic regions. To account for the difference in sequencing depth the maximum data threshold was set to 250,360 (which is the maximum number of reads found to map on the genomic regions of *NEAT1* and *MALAT1*) and the data were log transformed for comparison reasons.

RNA quality of single-cell suspensions derived from NSCLC and normal lung tissue

Surgically resected NSCLC tumor tissues and corresponding benign lung tissues were minced into small pieces (<1 mm) on ice and enzymatically digested with agitation for 30 min at 37 °C using the BD TuDoR™ dissociation reagent (BD Biosciences). The obtained single-cell solution was sieved through a 70 μM cell strainer (Corning) and red blood cells were removed using the BD Pharm Lyse™ lysing solution (BD Biosciences). Cells were counted and viability assessed with the BD Rhapsody scRNA-seq platform (BD Biosciences) using Calcein-AM (Invitrogen) and Draq7 (BD Biosciences). Immediately, $>1 \times 10^6$ cells of the obtained single-cell suspensions were subjected to the sample-tag staining procedure (20 min RT, 3x washing by 5 min centrifugation at 400 rpm). Total RNA was isolated before (T1) and after (T2) the sample-tag staining procedure using the RNeasy Mini kit (Qiagen) and RNA quality (RNA integrity

number, RIN) was assessed with the High Sensitivity RNA ScreenTape assay (Agilent) and the 4200 TapeStation (Agilent) system according to the manufacturer's instructions.

Immunohistochemistry

Immunohistochemical analysis with slides from the investigated 3 PCa patients was executed. Locations of interest on the slides were chosen based on slides from the same blocks stained with the markers P504s (clone SP116, Roche, #8035130001) and p63 (clone 4A4, Roche, #5867061001) to identify tumor regions or benign prostate regions. CD45 was stained (clone 2B11&PD7/26, Roche, #5269423001) to identify immune cells. Antibody staining was detected with the ultraView Universal DAB Detection Kit (Roche, #5269806001). The slides were scanned with Hamamatsu NanoZoomer S210 (Hamamatsu, Shizuoka, Japan) with a 40x magnification. Representative regions were selected by an expert prostate pathologist (G.S.) and the proportions of stromal, epithelial, and CD45$^+$ cells were calculated using the QuPath software for digital pathology image analysis [58].

Quantification and statistical analysis

Statistical analysis was performed GraphPad Prism. Single cell-data were aggregated into pseudo-bulk samples by biological replicates. Significance levels on the statistical tests are indicated in the figure captions.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Fresh resections of tumor tissue and corresponding benig tissue from PCa patients	Heidegger et al., *Molecular Cancer* 2022	https://doi.org/10.1186/s12943-022-01597-7
Fresh resections of tumor tissue and corresponding benig tissue from NSCLC patients	This paper.	N/A
Critical commercial assays and reagents		
BD TuDoR™ dissociation reagent	BD Biosciences	Cat#: 661563
BD Pharm Lyse™	BD Biosciences	Cat#: 555899
BD Rhapsody™ Cartridge Kit	BD Biosciences	Cat#: 633733
BD Rhapsody™ Cartridge Reagent Kit	BD Biosciences	Cat#: 633731
BD™ Human Single-Cell Multiplexing Kit	BD Biosciences	Cat#: 633781
BD Rhapsody™ WTA Amplification Kit	BD Biosciences	Cat#: 633801
BD Rhapsody™ cDNA Kit	BD Biosciences	Cat#: 633773
AMPure XP	Beckman Coulter	Cat#: A63880
Qubit™ dsDNA HS Assay Kit	Invitrogen	Cat#: Q32854
High Sensitivity D1000 Reagents	Agilent	Cat#: 5067-5585
High Sensitivity D5000 Reagents	Agilent	Cat#: 5067-5593
High Sensitivity D1000 ScreenTape	Agilent	Cat#: 5067-5584
High Sensitivity D5000 ScreenTape	Agilent	Cat#: 5067-5588
High Sensitivity RNA ScreenTape	Agilent	Cat#: 5067-5579
High Sensitivity RNA ScreenTape Sample Buffer	Agilent	Cat#: 5067-5580
High Sensitivity RNA ScreenTape Ladder	Agilent	Cat#: 5067-5581
RNeasy Mini Kit	Qiagen	Cat#: 74104
BD Pharmingen™ 7-AAD	BD Biosciences	Cat#: 559925
Calcein AM	Invitrogen	Cat#: C1430
Draq7	BD Biosciences	Cat#: 564904
ultraView Universal DAB Detection Kit	Roche	Cat#: 5269806001
Antibodies used for immunohistochemistry		
CD45 (clone 2B11&PD7/26)	Roche	Cat#: 5269423001
P504s (clone SP116)	Roche	Cat#: 8035130001
p63 (clone 4A4)	Roche	Cat#: 5867061001
Deposited data		
BD Rhapsody dataset (in H5AD format)	This study	https://zenodo.org/record/8063560
10X Chromium dataset	Heidegger et al., *Molecular Cancer* 2022	GSE193337
Software and algorithms		
Seven Bridges - BD Rhapsody™ WTA Analysis Pipeline	Seven Bridges Genomics	v1.7.1
Cellranger v5.0.0	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger
QuPath	Bankhead et al., *Scientific Reports* 2017	https://qupath.github.io
M3Drop	Andrews et al., *Bioinformatics* 2019	https://github.com/tallulandrews/M3Drop
Integrative Genomics Viewer (IGV)	Robinson JT et al., *Nature Biotechnology* 2011	https://software.broadinstitute.org/software/igv/
GraphPad Prism	Graphpad	v9
scanpy	Wolf et al., *Genome Biology* 2018	https://github.com/scverse/scanpy
scVI	Gayoso et al., *Nature Biotechnology* 2022	https://github.com/scverse/scvi-tools
decoupler-py	Badia-i-Mompel P et al., *Bioinformatics Advances* 2022	https://github.com/saezlab/decoupler-py
DESeq2	Love et al., *Genome Biology* 2014	https://github.com/thelovelab/DESeq2

CRediT authorship contribution statement

Stefan Salcher: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Isabel Heidegger:** Writing – review & editing, Validation, Resources, Funding acquisition. **Gerold Untergasser:** Investigation, Formal analysis, Conceptualization. **Georgios Fotakis:** Formal analysis, Data curation. **Alexandra Scheiber:** Visualization. **Agnieszka Martowicz:** Formal analysis. **Asma Noureen:** Data curation. **Anne Krogsdam:** Resources. **Christoph Schatz:** Validation. **Georg Schäfer:** Validation. **Zlatko Trajanoski:** Writing – review & editing. **Dominik Wolf:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Sieghart Sopper:** Writing – review & editing, Validation, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Andreas Pircher:** Writing – review & editing, Supervision, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e28358.

References

[1] D. Lambrechts, E. Wauters, B. Boeckx, S. Aibar, D. Nittner, O. Burton, et al., Phenotype molding of stromal cells in the lung tumor microenvironment, Nature medicine 24 (2018) 1277–1289.

[2] X. Guo, Y. Zhang, L. Zheng, C. Zheng, J. Song, Q. Zhang, et al., Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing, Nature medicine 24 (2018) 978–985.

[3] Q. Zhang, Y. He, N. Luo, S.J. Patel, Y. Han, R. Gao, et al., Landscape and dynamics of single immune cells in hepatocellular carcinoma, Cell 179 (2019) 829–845 e820.

[4] J. Goveia, K. Rohlenova, F. Taverna, L. Treps, L.C. Conradi, A. Pircher, et al., An integrated gene expression landscape profiling approach to identify lung tumor endothelial cell heterogeneity and angiogenic candidates, Cancer Cell 37 (2020) 421.

[5] E. Azizi, A.J. Carr, G. Plitas, A.E. Cornish, C. Konopacki, S. Prabhakaran, et al., Single-cell map of diverse immune phenotypes in the breast tumor microenvironment, Cell 174 (2018) 1293–1308 e1236.

[6] H. Li, A.M. van der Leun, I. Yofe, Y. Lubling, D. Gelbard-Solodkin, A.C.J. van Akkooi, et al., Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma, Cell 176 (2019) 775–789 e718.

[7] F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, et al., mRNA-Seq whole-transcriptome analysis of a single cell, Nat. Methods 6 (2009) 377–382.

[8] C. Zheng, L. Zheng, J.K. Yoo, H. Guo, Y. Zhang, X. Guo, et al., Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing, Cell 169 (2017) 1342–1356 e1316.

[9] J. Goveia, K. Rohlenova, F. Taverna, L. Treps, L.C. Conradi, A. Pircher, et al., An integrated gene expression landscape profiling approach to identify lung tumor endothelial cell heterogeneity and angiogenic candidates, Cancer Cell 37 (2020) 21–36 e13.

[10] D. Ramskold, S. Luo, Y.C. Wang, R. Li, Q. Deng, O.R. Faridani, et al., Author Correction: full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells, Nat. Biotechnol. 38 (2020) 374.

[11] S. Picelli, A.K. Bjorklund, O.R. Faridani, S. Sagasser, G. Winberg, R. Sandberg, Smart-seq2 for sensitive full-length transcriptome profiling in single cells, Nat. Methods 10 (2013) 1096–1098.

[12] P. See, J. Lum, J. Chen, F. Ginhoux, Corrigendum: a single-cell sequencing guide for immunologists, Front. Immunol. 10 (2019) 278.

[13] B. Phipson, L. Zappia, A. Oshlack, Gene length and detection bias in single cell RNA sequencing protocols, F1000Res 6 (2017) 595.

[14] T. Kivioja, A. Vaharautio, K. Karlsson, M. Bonke, M. Enge, S. Linnarsson, et al., Counting absolute numbers of molecules using unique molecular identifiers, Nat. Methods 9 (2011) 72–74.

[15] E. Hedlund, Q. Deng, Single-cell RNA sequencing: technical advancements and biological applications, Mol. Aspect. Med. 59 (2018) 36–46.

[16] T. Kalisky, S. Oriel, T.H. Bar-Lev, N. Ben-Haim, A. Trink, Y. Wineberg, et al., A brief review of single-cell transcriptomic technologies, Brief Funct Genomics 17 (2018) 64–76.

[17] X. Zhang, T. Li, F. Liu, Y. Chen, J. Yao, Z. Li, et al., Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems, Molecular cell 73 (2019) 130–142 e135.

[18] A.M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, et al., Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells, Cell 161 (2015) 1187–1201.

[19] E.Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, et al., Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets, Cell 161 (2015) 1202–1214.

[20] G.X. Zheng, J.M. Terry, P. Belgrader, P. Ryvkin, Z.W. Bent, R. Wilson, et al., Massively parallel digital transcriptional profiling of single cells, Nat. Commun. 8 (2017) 14049.

[21] T.M. Gierahn, M.H. Wadsworth 2nd, T.K. Hughes, B.D. Bryson, A. Butler, R. Satija, et al., Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput, Nat. Methods 14 (2017) 395–398.

[22] X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, et al., Mapping the mouse cell atlas by microwell-seq, Cell 173 (2018) 1307.

[23] H.C. Fan, G.K. Fu, S.P. Fodor, Expression profiling. Combinatorial labeling of single cells for gene expression cytometry, Science (New York, NY) 347 (2015) 1258367.

[24] E.Y. Shum, E.M. Walczak, C. Chang, H. Christina Fan, Quantitation of mRNA transcripts and proteins using the BD Rhapsody single-cell analysis system, Advances in experimental medicine and biology 1129 (2019) 63–79.

[25] E. Mereu, A. Lafzi, C. Moutinho, C. Ziegenhain, D.J. McCarthy, A. Alvarez-Varela, et al., Benchmarking single-cell RNA-sequencing protocols for cell atlas projects, Nat. Biotechnol. 38 (2020) 747–755.

[26] C. Ziegenhain, B. Vieth, S. Parekh, B. Reinius, A. Guillaumet-Adkins, M. Smets, et al., Comparative analysis of single-cell RNA sequencing methods, Molecular cell 65 (2017) 631–643 e634.

[27] K.N. Natarajan, Z. Miao, M. Jiang, X. Huang, H. Zhou, J. Xie, et al., Comparative analysis of sequencing technologies for single-cell transcriptomics, Genome Biol. 20 (2019) 70.

[28] T.M. Yamawaki, D.R. Lu, D.C. Ellwanger, D. Bhatt, P. Manzanillo, V. Arias, et al., Systematic comparison of high-throughput single-cell RNA-seq methods for immune cell profiling, BMC Genom. 22 (2021) 66.

[29] X. Wang, Y. He, Q. Zhang, X. Ren, Z. Zhang, Direct comparative analyses of 10X genomics Chromium and smart-seq2, Dev. Reprod. Biol. 19 (2021) 253–266.

[30] W. Chen, Y. Zhao, X. Chen, Z. Yang, X. Xu, Y. Bi, et al., A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples, Nat. Biotechnol. 39 (2021) 1103–1114.

[31] Y. Colino-Sanguino, LRdl Fuente, B. Gloss, A.M.K. Law, K. Handler, M. Pajic, et al., Systematic comparison of high throughput Single-Cell RNA-Seq platforms in complex tissues, bioRxiv 2023 (2023), 2004.2004.535585.

[32] C. Gao, M. Zhang, L. Chen, The comparison of two single-cell sequencing platforms: BD Rhapsody and 10x genomics Chromium, Curr Genomics 21 (2020) 602–609.

[33] S. Salcher, G. Sturm, L. Horvath, G. Untergasser, C. Kuempers, G. Fotakis, et al., High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer, Cancer Cell 40 (2022) 1503–1520 e1508.

[34] T. Hautz, S. Salcher, M. Fodor, G. Sturm, S. Ebner, A. Mair, et al., Immune cell dynamics deconvoluted by single-cell RNA sequencing in normothermic machine perfusion of the liver, Nat. Commun. 14 (2023) 2285.

[35] I. Heidegger, G. Fotakis, A. Offermann, J. Goveia, S. Daum, S. Salcher, et al., Comprehensive characterization of the prostate tumor microenvironment identifies CXCR4/CXCL12 crosstalk as a novel antiangiogenic therapeutic target in prostate cancer, Mol. Cancer 21 (2022) 132.

[36] B. Hwang, J.H. Lee, D. Bang, Single-cell RNA sequencing technologies and bioinformatics pipelines, Exp. Mol. Med. 50 (2018) 1–14.

[37] M.D. Luecken, F.J. Theis, Current best practices in single-cell RNA-seq analysis: a tutorial, Mol. Syst. Biol. 15 (2019) e8746.

[38] F. Ji, R.I. Sadreyev, Single-cell RNA-seq: introduction to bioinformatics analysis, Curr. Protoc. Mol. Biol. 127 (2019) e92.

[39] Q. Zhao, J. Wang, I.V. Levichkin, S. Stasinopoulos, M.T. Ryan, N.J. Hoogenraad, A mitochondrial specific stress response in mammalian cells, The EMBO journal 21 (2002) 4411–4419.

[40] T. Ilicic, J.K. Kim, A.A. Kolodziejczyk, F.O. Bagger, D.J. McCarthy, J.C. Marioni, et al., Classification of low quality cells from single-cell RNA-seq data, Genome Biol. 17 (2016) 29.

[41] D. Osorio, J.J. Cai, Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control, Bioinformatics 37 (2021) 963–967.

[42] S. Chen, G. Zhu, Y. Yang, F. Wang, Y.T. Xiao, N. Zhang, et al., Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression, Nat. Cell Biol. 23 (2021) 87–98.

[43] J.E. Wilusz, C.K. JnBaptiste, L.Y. Lu, C.D. Kuhn, L. Joshua-Tor, P.A. Sharp, A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails, Genes & development 26 (2012) 2392–2407.

[44] A. Naveed, J.A. Cooper, R. Li, A. Hubbard, J. Chen, T. Liu, et al., NEAT1 polyA-modulating antisense oligonucleotides reveal opposing functions for both long non-coding RNA isoforms in neuroblastoma, Cell. Mol. Life Sci. : CMLS 78 (2021) 2213–2230.

[45] J.C. Schupp, T.S. Adams, C. Cosme Jr., M.S.B. Raredon, Y. Yuan, N. Omote, et al., Integrated single-cell atlas of endothelial cells of the human lung, Circulation 144 (2021) 286–302.

[46] M. Eisenstein, Startups use short-read data to expand long-read sequencing market, Nat. Biotechnol. 33 (2015) 433–435.

[47] F.A. Wolf, P. Angerer, F.J. Theis, SCANPY: large-scale single-cell gene expression data analysis, Genome Biol. 19 (2018) 15.

[48] I. Virshup, S. Rybakov, F.J. Theis, P. Angerer, F.A. Wolf, anndata: annotated data, bioRxiv 2021 (2021), 2012.2016.473007.

[49] C. Xu, R. Lopez, E. Mehlman, J. Regier, M.I. Jordan, N. Yosef, Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models, Mol. Syst. Biol. 17 (2021) e9620.

[50] A. Gayoso, R. Lopez, G. Xing, P. Boyeau, V. Valiollah Pour Amiri, J. Hong, et al., A Python library for probabilistic analysis of single-cell omics data, Nat. Biotechnol. 40 (2022) 163–166.

[51] E. Becht, L. McInnes, J. Healy, C.A. Dutertre, I.W.H. Kwok, L.G. Ng, et al., Dimensionality reduction for visualizing single-cell data using UMAP, Nat. Biotechnol. 37 (2019) 38–44.

[52] V.A. Traag, L. Waltman, N.J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities, Sci. Rep. 9 (2019) 5233.

[53] I.M.P. Badia, J. Velez Santiago, J. Braunger, C. Geiss, D. Dimitrov, S. Muller-Dott, et al., decoupleR: ensemble of computational methods to infer biological activities from omics data, Bioinform Adv 2 (2022) vbac016.

[54] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.

[55] J.W. Squair, M. Gautier, C. Kathe, M.A. Anderson, N.D. James, T.H. Hutson, et al., Confronting false discoveries in single-cell differential expression, Nat. Commun. 12 (2021) 5692.

[56] T.S. Andrews, M. Hemberg, M3Drop: dropout-based feature selection for scRNASeq, Bioinformatics 35 (2019) 2865–2867.

[57] J.T. Robinson, H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, et al., Integrative genomics viewer, Nat. Biotechnol. 29 (2011) 24–26.

[58] P. Bankhead, M.B. Loughrey, J.A. Fernandez, Y. Dombrowski, D.G. McArt, P.D. Dunne, et al., QuPath: open source software for digital pathology image analysis, Sci. Rep. 7 (2017) 16878.