



# Robust detection of clinically relevant features in single-cell RNA profiles of patient-matched fresh and formalin-fixed paraffin-embedded (FFPE) lung cancer tissue

Alexandra Trinks<sup>1</sup> · Miha Milek<sup>2</sup> · Dieter Beule<sup>2</sup> · Julie Kluge<sup>3</sup> · Stefan Florian<sup>3,4</sup> · Christine Sers<sup>3,5</sup> · David Horst<sup>3,5</sup> · Markus Morkel<sup>1,3,5</sup> · Philip Bischoff<sup>3,4,5</sup>

Accepted: 23 January 2024 / Published online: 1 February 2024  
© The Author(s) 2024

## Abstract

**Purpose** Single-cell transcriptional profiling reveals cell heterogeneity and clinically relevant traits in intra-operatively collected patient-derived tissue. So far, single-cell studies have been constrained by the requirement for prospectively collected fresh or cryopreserved tissue. This limitation might be overcome by recent technical developments enabling single-cell analysis of FFPE tissue.

**Methods** We benchmark single-cell profiles from patient-matched fresh, cryopreserved and archival FFPE cancer tissue.

**Results** We find that fresh tissue and FFPE routine blocks can be employed for the robust detection of clinically relevant traits on the single-cell level. Specifically, single-cell maps of fresh patient tissues and corresponding FFPE tissue blocks could be integrated into common low-dimensional representations, and cell subtype clusters showed highly correlated transcriptional strengths of signaling pathway, hallmark, and clinically useful signatures, although expression of single genes varied due to technological differences. FFPE tissue blocks revealed higher cell diversity compared to fresh tissue. In contrast, single-cell profiling of cryopreserved tissue was prone to artifacts in the clinical setting.

**Conclusion** Our analysis highlights the potential of single-cell profiling in the analysis of retrospectively and prospectively collected archival pathology cohorts and increases the applicability in translational research.

**Keywords** Single-cell RNA sequencing · Single-cell transcriptomics · Lung cancer · FFPE tissue analysis · Tumor heterogeneity

---

Markus Morkel and Philip Bischoff have equally contributed to this article.

✉ Markus Morkel  
markus.morkel@charite.de

✉ Philip Bischoff  
philip.bischoff@charite.de

<sup>1</sup> Bioportal Single Cells, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

<sup>2</sup> Core Unit Bioinformatics (CUBI), Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

<sup>3</sup> Institute of Pathology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Charitéplatz 1, 10117, Berlin, Germany

<sup>4</sup> BIH Biomedical Innovation Academy, BIH Charité Clinician Scientist Program, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

<sup>5</sup> German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK) Partner Site Berlin, Heidelberg, Germany

## 1 Introduction

High-throughput single-cell RNA sequencing (scRNA-seq) has enabled researchers to study the various aspects of cellular heterogeneity of tissues, including human clinical samples. In particular, scRNA-seq has been used extensively to characterize the complex cell compositions of tumor tissue cohorts [1–3], and can reveal features which contribute to patient prognosis [4, 5] and therapy resistance [6]. So far, single-cell RNA profiling required the prospective collection of fresh tissue samples during surgery or biopsy. Retrospective analysis of frozen or formalin-fixed and paraffin-embedded (FFPE) tissue cohorts would allow faster correlation of scRNA-seq profiles to clinical features, as many important clinical characteristics, such as tumor genetics, therapy response and patient survival, are only obtainable weeks to years after sample acquisition. In addition, this could unlock archival FFPE tissue collections for single-cell profiling.

Recently, scRNA-seq chemistry has been developed for frozen and formalin-fixed single-nucleus suspensions. Early studies have shown that this allows performing single-nucleus RNA sequencing (snRNA-seq) of nuclei isolated from frozen and FFPE tissue samples [7–10] at read depths that allow a similarly fine-grained analysis compared to fresh cell suspensions. However, to our knowledge, no side-by-side comparisons have been performed on fresh, frozen, and FFPE clinical tissue samples from the same patients. To answer the question which features of fresh tissue are preserved in archival samples, we profiled three lung adenocarcinomas using fresh, cryopreserved and FFPE tissue samples side-by-side. We found that cell-intrinsic and clinically relevant features of cancers are robustly preserved in single-cell transcriptomes of FFPE samples.

## 2 Methods

### 2.1 Sample acquisition

Tissue samples were collected from therapy-naïve lung adenocarcinoma patients undergoing primary surgery. Informed consent was obtained from all participants before sample acquisition. The ethics committee of Charité - Universitätsmedizin Berlin approved the use of tissue samples for single-cell gene expression analysis (vote EA4/164/19). Fresh tumor tissue close to the invasive front was sampled by a pathologist during intraoperative pathology consultation. From the same portion of tissue, tissue samples of approx.  $5 \times 5 \times 5$  mm each were either immediately snap-frozen in liquid nitrogen and stored at  $-80$  °C, or placed in Tissue Storage Solution (Miltenyi) on ice for max. 30 min before subsequent further processing of the fresh tissue. The remaining surgery specimen was subjected to routine histological examination, including formalin fixation and paraffin embedding. FFPE tissue was sampled from archive tissue blocks after storage under standard archive conditions (dry, room temperature) for 4–5 months. For FFPE tissue sampling, 4–6 tissue cores of 1 mm diameter and 1–3 mm length, depending on the block thickness, were punched out of FFPE tissue blocks. Tissue cores were obtained from a tumor area of approx.  $5 \times 10$  mm close to the invasive front, which had been marked by a pathologist, in order to closely resemble the intraoperatively collected fresh and cryopreserved tissue (Fig. S1a).

### 2.2 Processing of fresh tissue samples

Fresh tissue samples were cut into small pieces of 1 mm diameter and dissociated using the Tumor Dissociation Kit,

human (Miltenyi) and a gentleMACS Octo Dissociator with heaters (Miltenyi), using preinstalled program 37C\_h\_TDK\_1 for 30–45 min. All subsequent steps were performed at 4 °C or on ice. Dissociated tissue was filtered through 100 µm filters, pelleted by centrifugation at  $300 \times g$  in BSA-coated low-binding tubes, incubated with 1 ml ACK erythrocyte lysis buffer (Gibco) for 1 min, washed with DMEM, pelleted, resuspended in PBS and filtered through 20 µm filters. Finally, debris was removed using the Debris Removal Solution (Miltenyi) and cells were counted using a Neubauer chamber. Single-cell suspensions were further processed using the Chromium Single Cell 3' Reagent Kit v3 and the Chromium Controller (10x Genomics, Pleasanton, California, USA) according to the manufacturer's protocol without any adjustments.

### 2.3 Processing of cryopreserved tissue samples

Cryopreserved tissue samples were cracked using a pestle and mortar placed on dry ice and pre-chilled with liquid nitrogen into small pieces of 1 mm diameter. All subsequent steps were performed at 4 °C or on ice. Tissue pieces were homogenized in homogenization buffer [1x Nuclei EZ lysis buffer (Sigma), 0.6 U/ml RNase Inhibitor (Ambion), 0.3 U/ml Superasin (Ambion)] using a pestle and douncer by approx. 10 strokes with a loose pestle and 5 strokes with a tight pestle. Homogenized tissue was filtered through a 30 µm filter, stained with DAPI (0.1 µg/mL) for 5 min and sorted using a BD FACSAria Fusion (100 µm nozzle) to remove debris and doublets. Cell concentration was determined using a Neubauer chamber before single-nuclei suspensions were further processed using the Chromium Single Cell 3' Reagent Kit v3 and the Chromium Controller iX (10x Genomics, Pleasanton, California, USA) according to the manufacturer's protocol without any adjustments.

### 2.4 Processing of FFPE tissue samples

FFPE tissue samples were processed according to a published protocol [8] with adjustments according to another published protocol [9]. For a detailed version of the protocol see Supplementary Methods. In short, tissue cores were obtained as described above and cut into small pieces of 1 mm diameter, washed at room temperature in xylene 3 times for 10 min, in 100% ethanol 2 times for 30 s, in 70% ethanol for 30 s, in 50% ethanol for 30 s, in distilled water for 30 s, and in Buffer V (FFPE Tissue Dissociation Kit, Miltenyi) for 30 s. Tissue was dissociated using the FFPE Tissue Dissociation Kit (Miltenyi) and a gentleMACS Octo Dissociator with heaters (Miltenyi), using preinstalled program 37C\_FFPE\_1. After 20 min of dissociation and after complete dissociation, dissociated

tissue was pipetted through a 20 G needle for 10–20 times. Next, samples were filtered through 70  $\mu\text{m}$  filters, placed on ice, washed with chilled Buffer V, resuspended in Resuspension Buffer (0.5x PBS, 50 mM Tris pH8, 0.02% BSA, 0.24 U/ $\mu\text{l}$  RNasin (Ambion, AM2684) in  $\text{H}_2\text{O}$ ), and cells were counted using a Neubauer chamber. Single-nuclei suspensions were further processed using the Chromium Fixed RNA Profiling Reagent Kit and the Chromium Controller Xi (10x Genomics, Pleasanton, California, USA) according to the manufacturer's protocol. After step 2.1.m in the manufacturer's protocol, cells were stained with DAPI (0.1  $\mu\text{g}/\text{mL}$ ) for 5 min and sorted using a BD FACSAria Fusion (100  $\mu\text{m}$  nozzle) to remove debris and doublets. After cell sorting, samples were further processed according to the manufacturer's protocol.

## 2.5 Library preparation and sequencing

After isolation, 10,000 cells/nuclei were subjected to barcoding and library preparation. Libraries of fresh and cryopreserved samples were prepared using the Chromium Single Cell 3' Reagent Kit v3 (10x Genomics) according to the manufacturer's protocol. Libraries of FFPE samples were prepared using the Chromium Fixed RNA Profiling Reagent Kit (10x Genomics) according to the manufacturer's protocol. Libraries were sequenced on a NovaSeq (Illumina) at approx. 400 mio. reads per library.

## 2.6 Data analysis

Sequencing reads were aligned against reference transcriptome GRCh38 and UMIs were quantified using Cellranger, version 7.1.0 (10x Genomics). Subsequent analyses were performed using R version 4.1.1 and the Seurat package version 4.3.0 [11], if not stated otherwise. First, signal from ambient RNA was removed using the SoupX package version 1.6.2 [12], assuming a contamination fraction of up to 0.2. Gene expression data of all samples were merged and filtered for the following quality parameters: 300–10,000 genes per cell, 500–100,000 UMIs per cell, fraction of mitochondrial reads lower than 15% and fraction of hemoglobin reads lower than 5% (see Fig. S1b for cell counts before and after filtering). Next, gene expression data was log-normalized and dimensionality reduction was performed by principal component analysis (PCA). Uniform manifold approximation projection (UMAP) based on the top 10 principal components (PCs) was used for data visualization. Gene expression data was integrated by sample type (fresh, frozen or FFPE) using reciprocal PCA based

on the top 30 PCs. Clustering was performed using shared nearest neighbor graph calculation. In fresh and FFPE data, main cell types and cell subtypes were manually annotated using canonical marker genes selected from the literature. Cell type labels were transferred based on the first 30 PCs to the frozen data using the FindTransferAnchors and TransferData functions of the Seurat package. Clusters containing cell doublets were identified by discrepant marker gene expression and removed prior to further analysis. Signaling pathway activity scores were calculated using the PROGENY package version 1.17.3 [13]. Gene signature expression scores were calculated using Seurat. The inferCNV package version 1.10.1 (<https://github.com/broadinstitute/inferCNV>) was used for copy number analysis in epithelial transcriptomes.

## 2.7 Immunohistochemistry and image analysis

FFPE tissue sections of 2–3  $\mu\text{m}$  were prepared for immunohistochemistry (IHC). Staining procedures were performed on Bond MAX (Leica) and BenchMark XT (Ventana) immunostainers according to manufacturer's protocols. Tissue sections were deparaffinized, rehydrated, and antigen retrieval was performed either by protease incubation for 8 min (for Cytokeratin), or incubation for 30 min at 100  $^{\circ}\text{C}$  in Epitope Retrieval ER2 Solution (Leica) (for CD45, PD-L1), or CC1 buffer (Ventana) (for ERG), or without antigen retrieval (for smooth muscle Actin). Incubation with primary antibodies was performed for 30 min at room temperature. The following primary antibodies were used: monoclonal mouse anti-Cytokeratin (clone AE1/AE3, 1:500, Dako/Agilent, cat. no. M3515), monoclonal mouse anti-CD45 (clone X16/99, 1:100, Leica, cat. no. NCL-L-LCA), monoclonal mouse anti-smooth muscle Actin (clone 1A4, 1:400, Dako/Agilent, cat. no. M0851), monoclonal rabbit anti-ERG (clone EPR3864, ready-to-use dilution, Ventana/Roche, cat. no. 790-4576), and monoclonal mouse anti-PD-L1 (clone E1L3N, 1:200, Cell Signaling, cat. no. 13684). Subsequently, tissue sections were incubated with HRP-conjugated secondary antibody (Leica) and DAB, and nuclei were counterstained with hematoxylin.

After IHC staining was completed, tissue sections were digitized using a DP600 slide scanner (Ventana/Roche). For quantification of positive cells, the open-source software QuPath (version 0.3.2) was used. Three representative tumor regions of 1  $\text{mm}^2$  each adjacent to the sampling area were selected by a pathologist. Within these areas, cells were detected based on nuclei counterstaining and positive cells were identified via IHC intensity based on cut-off values set by a pathologist. The percentage of

positive cells was exported from QuPath for further analysis.

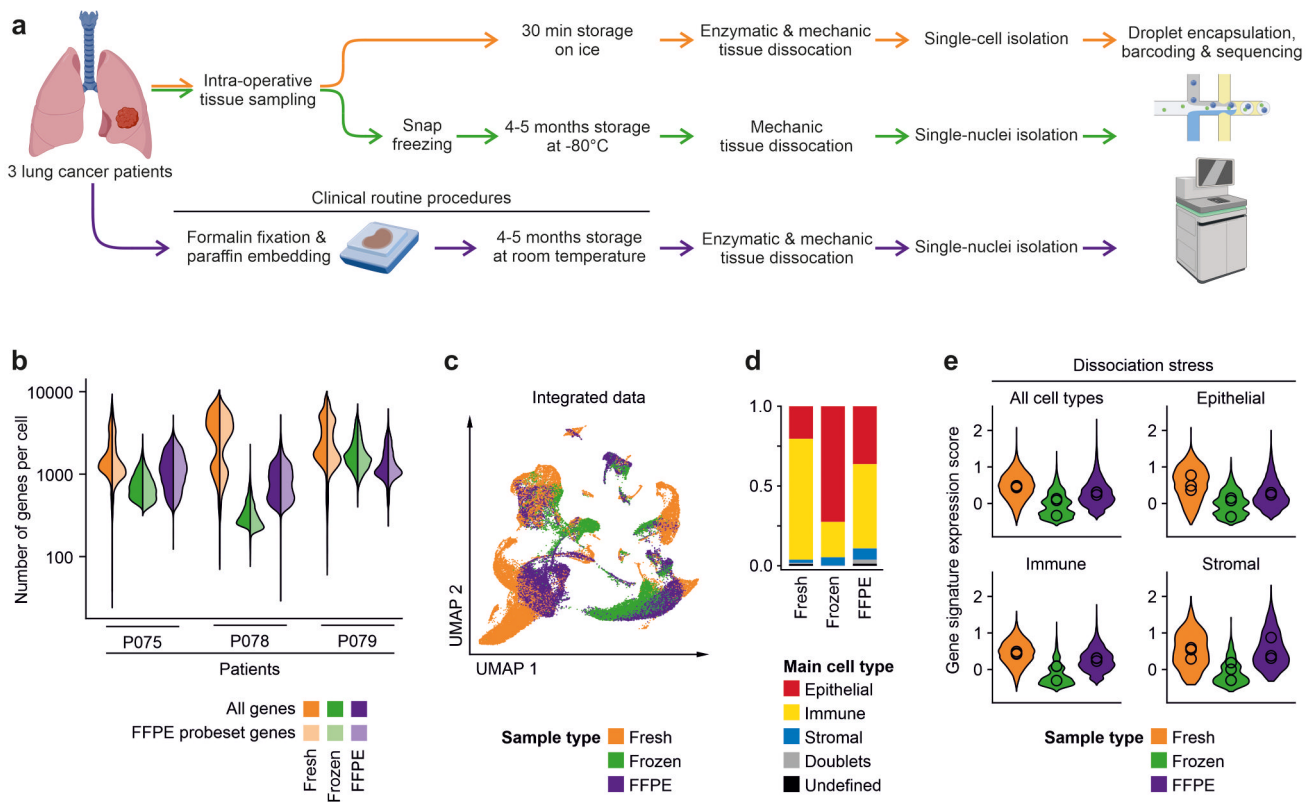
## 2.8 Statistical analysis

The quantity of main cell types in fresh and FFPE samples as well as IHC images were compared using the paired t test. The correlation of PROGENy pathway and Hallmark signature scores in different sample types (fresh, frozen, FFPE) was analyzed by calculating the Pearson correlation coefficient. Differentially expressed genes were identified using the FindAllMarkers function of the Seurat package with the following parameters: include only positive markers, proportion of expressing cells inside the cluster  $\geq 0.2$ , difference between proportions of expressing cells inside and outside the cluster  $\geq 0.2$ .

## 3 Results

### 3.1 Quality metrics of single-cell transcriptomes from fresh, frozen and archival FFPE samples

We selected cryopreserved and FFPE tissue samples of three lung adenocarcinoma patients (P075, P078, P079) that were previously analyzed by fresh tissue scRNA-seq for snRNA-seq analysis (Figs. 1a, S1a and S1b). Initial analysis of fresh and FFPE tissue-derived libraries showed expected high quality parameters (Figs. 1b and S1c); however, results gained from cryopreserved tissue showed unequal and lower quality scores in two subsequent rounds of library preparation and optimization of lab workflows (Fig. S1d). Across the patients, numbers of detected genes per sample were highest for fresh tissue, but lower for FFPE or frozen tissue (Fig. 1b), even when adjusting for



**Fig. 1** Workflow and quality metrics of fresh, frozen and FFPE tissue single-cell analysis. **a** Workflow of tissue specimens used for the study. In short, fresh tissues were procured intra-operatively, dissociated enzymatically, and cell suspensions were used for single-cell library preparation. Cryopreserved tissues were stored for 4–5 months at  $-80^{\circ}\text{C}$ , homogenized in the frozen state for nuclei isolation, and nucleus suspensions were used for library preparation. For FFPE analysis, 4–5 months old routine FFPE blocks were dissociated enzymatically and mechanically, and nucleus suspensions were used

for library preparation. **b** Numbers of genes called per cell in the various libraries. Full colors: all genes; lighter colors: genes limited by FFPE probe set. **c** UMAPs based on the top 10 principal components of all single-cell transcriptomes after filtering and data integration, color-coded by fresh, frozen or FFPE tissue origin. **d** Quantification of main cell type, by clustering and calling of cell type-specific marker genes. **e** Module score of gene signatures related to dissociation stress in the various main cell types, by fresh, frozen and FFPE origin, mean scores per patient indicated by circles



the restricted gene set used in the targeted FFPE sequencing approach. Single-cell information of the fresh, frozen and FFPE samples could be integrated into a common UMAP with clusters sharing transcriptomes of fresh, frozen and FFPE tissue origin, indicating that cell type information was stable across the methods (Figs. 1c and S1e). When calling main cell types, we found that immune cell transcriptomes were enriched, but epithelial and stromal cells transcriptomes were rather depleted from fresh tissue single-cell libraries (Fig. 1d), indicating potential negative effects of tissue dissociation on cell representation. In line with this interpretation, a recently published gene signature for dissociation stress [14] was highest in fresh tissue transcriptomes, in particular in epithelial cells (Fig. 1e).

### 3.2 Cell type abundancies in FFPE versus fresh samples

As FFPE tissue blocks are most relevant for clinical applications and frozen tissue libraries were hampered by low quality parameters, we integrated fresh and FFPE data for in-depth analysis (Fig. 2a), and again determined fractions of transcriptomes of epithelial, immune and stromal origin (Figs. 2b and S2). The proportion of main cell types as quantified by IHC was better recapitulated in FFPE data compared to fresh tissue data, although in both fresh and FFPE data, we observed a trend towards overrepresentation of immune cells and underrepresentation of epithelial and stromal cells. Assessment of epithelial marker genes showed similar expression scores across fresh and FFPE tissue samples, as *EPCAM*, *KRT5*, *SCGB3A2*, *FOXJ1*, *AGER* and *SFTPC* marked tumor epithelial, Basal, Transitional, Ciliated, AT1, and AT2 cells, respectively, across the datasets (Fig. 2c). However, we also saw that some marker genes were preferentially detected in fresh versus FFPE transcriptomes, such as *MUC5B* and *SFTPC* in fresh or FFPE-derived tumor epithelial cells, respectively. This was likely due to the different absolute numbers of epithelial cell transcriptomes across fresh versus FFPE tissues (Fig. 2b), but also due to higher epithelial cell type diversity in the FFPE tissue blocks of patients P078 and P079 (Fig. 2d). The latter result suggests that normal epithelial cells are most easily lost during fresh tissue dissociation; however, we cannot rule out compositional bias in the fresh versus FFPE tissue specimens used. Detection of immune cell marker genes and cell types was more even across fresh and FFPE tissue specimens (Fig. 2e, f), while many more fibroblasts were detected among the stromal cell types in FFPE tissue-derived sequence libraries (Fig. 2g, h). Cell type label transfer to the cryopreserved tissue samples showed often skewed representation of cell types (Fig. S3a, b), in particular for immune cells such as regulatory or CD8+ T cells that are

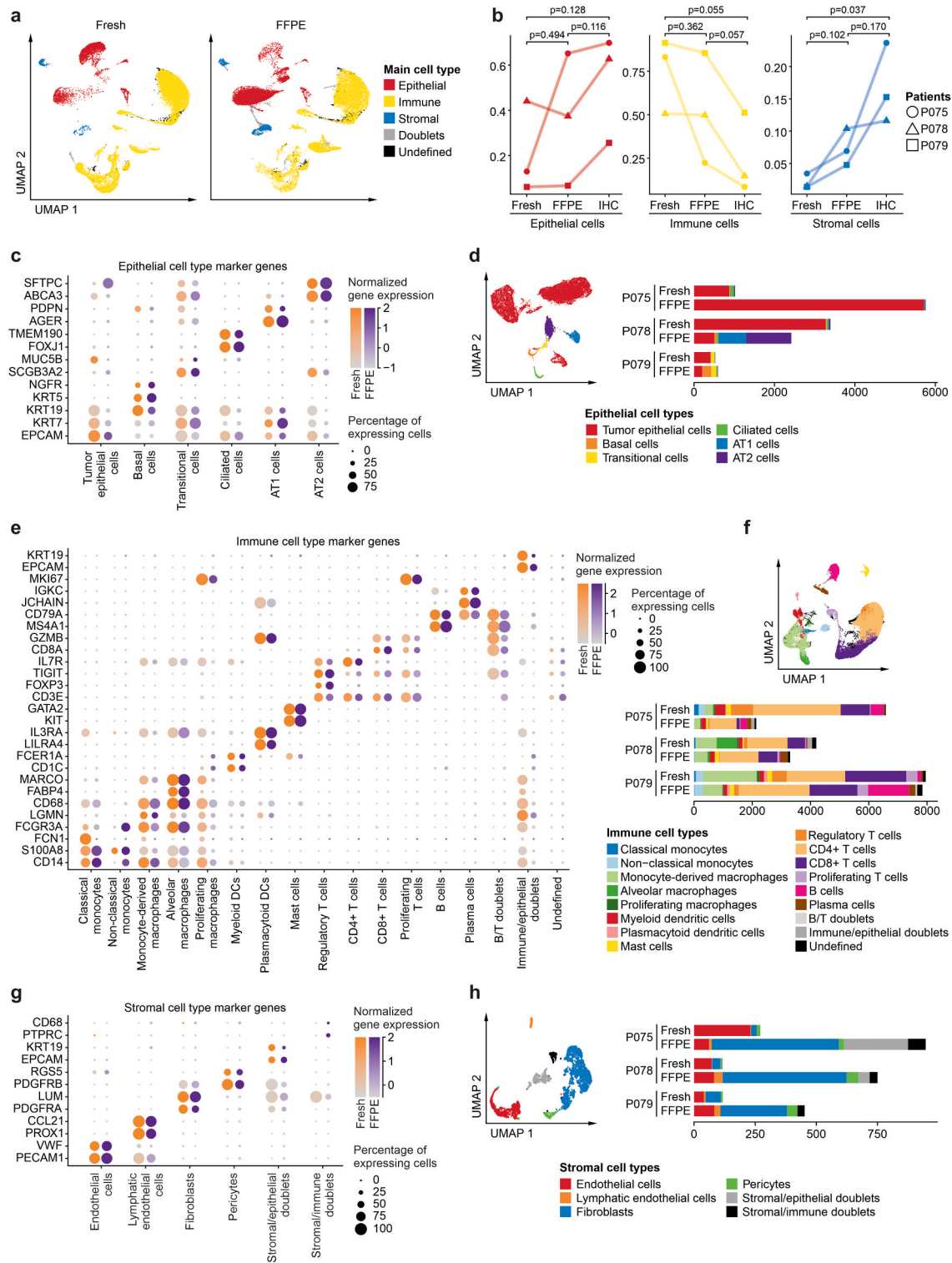
defined by few key marker genes which could not be reliably detected in the frozen samples.

### 3.3 Cell trait heterogeneity in FFPE versus fresh samples

We asked whether archival FFPE tissue analysis allows characterization of cell types to a similar extent as fresh tissue single-cell analysis. We performed this analysis on gene signature level, taking into account sets of pathway target gene signatures [13] and relevant Hallmark gene sets [15]. Technological differences in sample preparation workflows resulted in differences of normalized gene expression scores between the fresh tissue and the FFPE-derived libraries (Fig. S4a, b), where, for instance, the MAPK target gene and feedback regulator *DUSP4* was found higher expressed in fresh tissue libraries, whereas the MAPK-regulated methyltransferase *KMT5A* scored higher in FFPE-derived libraries (Fig. 3a). Despite these differences on individual gene expression levels, we found high correlations between inferred pathway activities (Fig. 3b) and tumor-related cell traits (Fig. 3c) from fresh and FFPE samples throughout. In line with prior studies, we found high TGF $\beta$  activity in fibroblasts, and this cell type also had the highest score for mesenchymal traits judged by activity of the EMT and Angiogenesis Hallmark signature. As expected, highest activity of the G2/M checkpoint and E2F targets Hallmark signatures was found in proliferating T cells and macrophages, which correlated with high MAPK activity in proliferating T cells. We detected high activity of TNF $\alpha$  and NF $\kappa$ B signaling in monocytes, and high scores of the Interferon Gamma Response Hallmark signature in monocytes and tumor-associated monocyte-derived macrophages (Mo-Macs), while tissue-resident alveolar macrophages were characterized by high scores of the Oxidative Phosphorylation Hallmark signature. In summary, correlations between signature activities derived from fresh or FFPE tissue samples were strong on the cell type level with Pearson correlation coefficients mostly larger than 0.9 and significances mostly smaller than  $p=0.001$  across all signatures analyzed, showing that biology-related cell type characteristics can be inferred from fresh and archival FFPE tissue specimens alike. Scores were lower when comparing correlations of cell type traits between fresh versus frozen libraries, in particular for pathways with smaller variation between cell types (Fig. S3c).

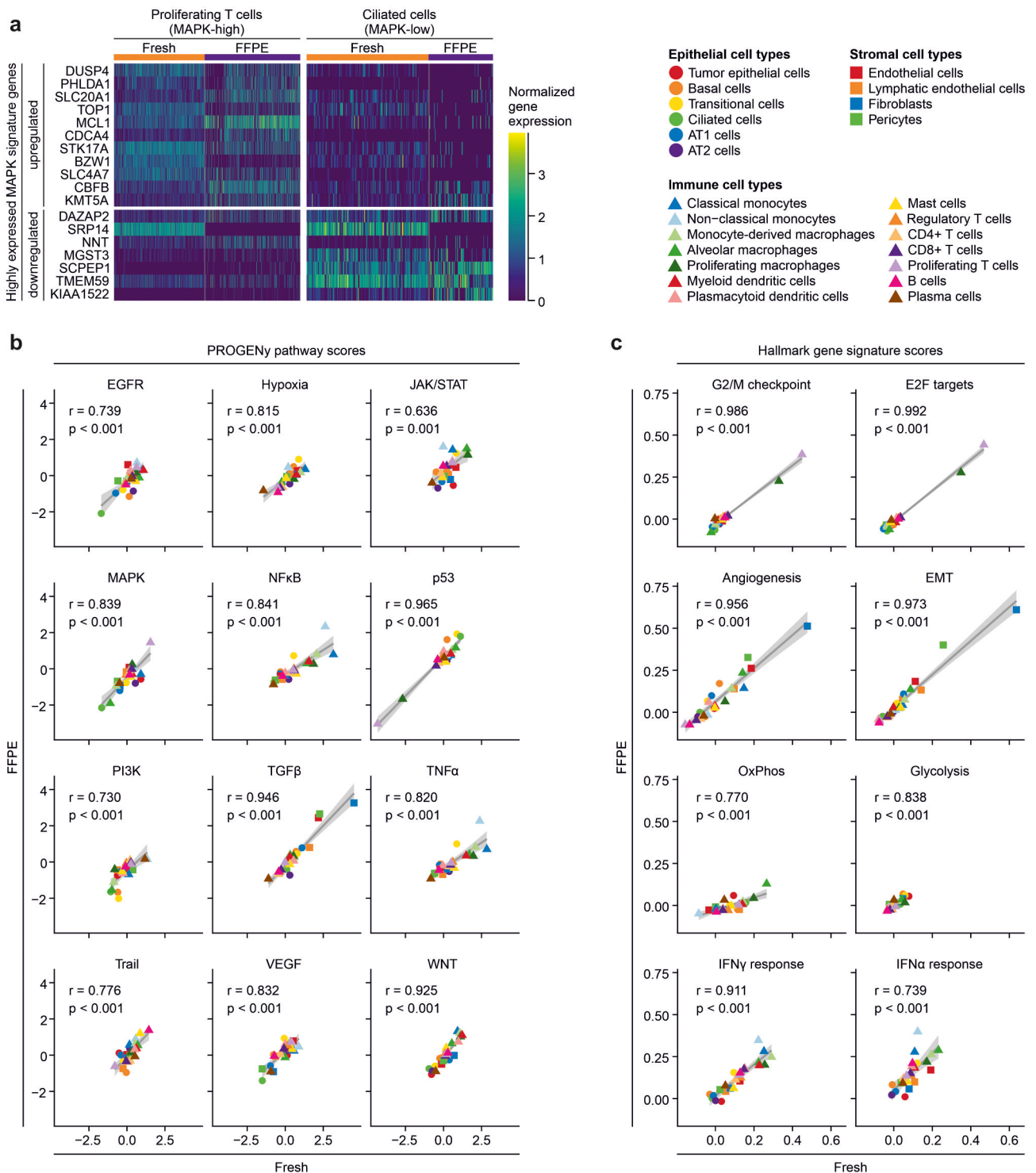
### 3.4 Assessment of clinical features from FFPE versus fresh tissue single-cell transcriptomes

We investigated whether patient-specific differences in tumor and microenvironmental characteristics can be



**Fig. 2** Cell type diversity in fresh versus FFPE tissue single-cell analysis. **a** UMAPs based on the top 10 principal components of fresh or FFPE single-cell transcriptomes, as indicated, color-coded by main cell type. **b** Relative proportions of epithelial, immune or stromal cells, compared between fresh and FFPE-derived single-cell libraries, and as quantified by IHC staining of FFPE tissue sections, see Fig. S2 for representative images, paired t-test per main cell type. **c, d** Analysis of epithelial transcriptomes. **c** Epithelial marker gene expression per

cell type in fresh or FFPE tissue-derived libraries. **d** UMAP and absolute cell numbers, color-coded by cell type. **e, f** Analysis of immune cell transcriptomes. **e** Immune marker gene expression per cell type in fresh or FFPE tissue-derived libraries. **f** UMAP and absolute cell numbers, color-coded by cell type. **g, h** Analysis of stromal cell transcriptomes. **g** Stromal marker gene expression per cell type in fresh or FFPE tissue-derived libraries. **h** UMAP and absolute cell numbers, color-coded by cell type



**Fig. 3** Cell trait quantification in fresh versus FFPE single-cell analysis. **a** Expression of selected PROGENy signature MAPK target genes in fresh tissue and FFPE libraries. Expression was normalized to scale across libraries. Data shown for cells assigned as Proliferating T cells (high MAPK activity, see Fig. 3b) or Ciliated cells (low MAPK activity, see Fig. 3b). **b** Correlations of PROGENy

pathway scores between FFPE and fresh tissue gene expression per cell type, Pearson correlation coefficient and *p*-value indicated per pathway. **c** Correlations of selected biologically-relevant Hallmark gene signatures between FFPE and fresh tissue gene expression per cell type, Pearson correlation coefficient and *p*-value indicated per gene signature

determined in fresh and FFPE tissue samples to the same degree. The histological grade of tumors (Fig. 4a) perfectly correlated with recently identified gene signatures of tumor differentiation [16] (Fig. 4b), regardless of whether the signatures were called from fresh or FFPE tissue-derived epithelial cell transcriptomes. The lowest histological tumor grade in patient P079 was further associated with higher activity of CAF-related pathways in fibroblasts [17] (Fig. 4c) and high expression of typical CAF marker genes on the single-cell mRNA level (Fig. 4d). With regard to oncogenic pathways, tumor epithelial cells of patient P079 scored highest for MAPK and WNT target expression, while P075 had the highest values of p53 pathway activity, in both fresh and FFPE tissue samples (Fig. S5a). In contrast, no correlation was observed for EGFR pathway activity. We inferred similar copy number profiles in tumor epithelial cells from both fresh and FFPE tissue-derived transcriptomes (Fig. S5b).

Histologically, tumors differed in expression of the immune evasion marker PD-L1 in tumor cells (Fig. 4e) which was similarly observed on the RNA level in fresh as well as FFPE tissue-derived tumor epithelial cell transcriptomes (Fig. 4f). In both fresh and FFPE transcriptome data, high PD-L1 expression in patient P079 correlated with high expression of the Interferon Gamma Response Hallmark signature in tumor cells and the tumor microenvironment, in particular in Mo-Macs [18] (Fig. 4g). In Mo-Macs, this feature was accompanied by high TNF $\alpha$ , NF $\kappa$ B and JAK/STAT pathway activity and high expression scores of the Inflammatory Response Hallmark signature in patient P079 (Fig. 4h) [19, 20]. Among T cells, high PD-L1 expression in P079 correlated with low cytotoxicity and high naiveness scores in CD4<sup>+</sup> T cells, and high scores for exhaustion in CD8<sup>+</sup> T cells (Fig. 4i) [21]. In summary, these correlations are in agreement with previous analysis of the lung cancer microenvironment [16, 22], and indicate that cell type characteristics relevant for clinical stratification can be retrieved faithfully from the FFPE snRNA-seq approach.

## 4 Discussion

Single-cell sequencing of surgical tissues has uncovered clinically relevant information on the patient and cohort levels. However, the use of fresh tissues comes with serious limitations in the clinic, such as the requirement for a rapid tissue handling pipeline. Recent technological advances allow single-nucleus sequencing from frozen and FFPE tissue, which promises broader availability of tissue, including routine pathology specimens. Here, we present a systematic comparison of fresh, frozen and FFPE single-cell analysis of clinical tissue. We find that

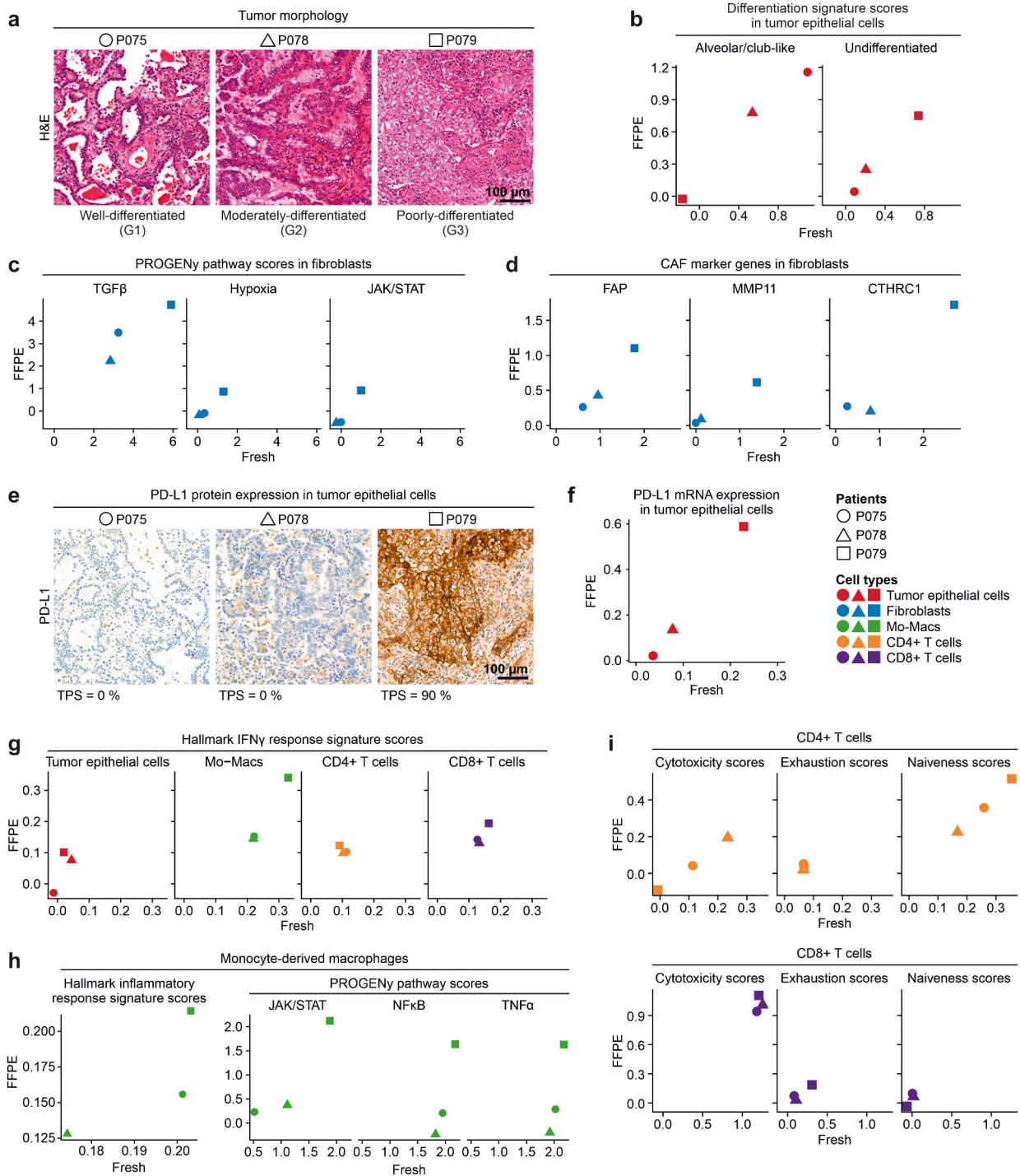
FFPE tissue robustly preserved clinically relevant information on cell types and patient characteristics comparable to fresh tissue.

Single-cell analysis of fresh solid tissue has been the gold standard to survey cell type composition of human tissues in health and disease [23–25]. Generally, the procedure relies on cell dissociation, most often by proteases, which can result in disproportionate release of cell types and preparation-related transcriptome artifacts, although procedures have been developed to minimize such effects. Moreover, fresh tissue requires timely processing, which can complicate standardization and result in batch effects. Potentially, analysis of transcriptionally inert nuclei from frozen or FFPE tissues could result in more even transcriptome representation across cell types in the absence of dissociation artifacts. We have here tested available procedures side-by-side on tissues of the same origin and found that single-cell sequencing of FFPE tissue results in transcriptomes representing a higher diversity of cell types compared to analysis of fresh tissue cell suspensions, an equally good representation of biological and clinical transcriptome features on a signature level, and lower induction of stress-related gene expression. In contrast, frozen tissue analysis was less reliable in our hands, probably due to handling of unfixed nuclei suspensions that are prone to RNA degradation or RNA diffusion.

In the clinic, fresh tissue single-cell analysis allowed for prospective collection of samples only. Importantly, single nucleus transcriptomics of FFPE tissues allow for retrospective analysis of cohorts that are annotated with long-term clinical follow-up data such as therapy response, relapse, metastasis, and patient survival. The FFPE blocks under analysis here were stored under standard archival conditions at room temperature for 4–5 months. It remains to be determined whether older FFPE blocks with longer follow-up periods perform equally well, as RNA quality could potentially deteriorate over time even in the fixed and embedded state. Moreover, it needs to be investigated to which degree prolonged time until fixation and tissue autolysis in large surgery specimens impair snRNA-seq analysis.

Pathology review is usually performed using FFPE sections, as these yield high-quality histology and immunohistochemistry information. Therefore, FFPE blocks can be annotated regionally with high confidence, unlike intraoperative fresh tissue samples. Thus, FFPE blocks allow analysis of cell composition of circumscribed areas of interest. We envision that careful selection of tissue regions in pathology, in conjunction with multiplexing techniques for cost-effectiveness, will bring single-cell analysis aiding therapy prediction or prognosis a step closer to clinical routine for cancer patients.





**Fig. 4** Quantification of clinically relevant gene expression patterns in fresh versus FFPE single-cell analysis. **a** Hematoxylin and eosin stained FFPE tumor sections of patients P075, P078, P079, showing a predominant lepidic, acinar or solid growth pattern, respectively, corresponding to well-differentiated (G1), moderately-differentiated (G2) or poorly-differentiated (G3) tumor morphology and grade, as indicated. **b** Correlation between FFPE and fresh tissue of Alveolar/club-like and Undifferentiated tumor cell signature scores in tumor epithelial cells. **c** Correlation between FFPE and fresh tissue of PROGENy pathway scores in fibroblasts. **d** Correlation between FFPE and fresh tissue of CAF marker expression in fibroblasts.

**e** Immunohistochemistry on FFPE tumor section for PD-L1 expression. TPS: Tumor proportion score. **f** Correlation between FFPE and fresh tissue of PD-L1 gene expression in tumor epithelial cells. **g** Correlation between FFPE and fresh tissue of Interferon Gamma Response Hallmark gene signature scores in various cell types, as indicated. Mo-Macs = Monocyte-derived macrophages. **h** Correlation between FFPE and fresh tissue of clinically and biologically relevant signature scores in Mo-Macs, as indicated. **i** Correlation between FFPE and fresh tissue of functional cell state scores in CD4+ and CD8+ T cells, as indicated

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s13402-024-00922-0>.

**Acknowledgements** We thank Caroline Braeuning at the BIH Core Unit Genomics for excellent technical support. Figure 1a was created with BioRender.com.

**Author contributions** Conceptualization: PB, SF, MM; Investigation: AT, PB, JK; Data Curation: MiM, DB, PB; Formal Analysis: PB; Validation: PB; Visualization: PB, MM; Writing – Original Draft: PB, MM, AT; Resources: MM, PB, CS, DH.

**Funding** Open Access funding enabled and organized by Projekt DEAL. PB is participant in the BIH Charité Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin, and the Berlin Institute of Health at Charité (BIH).

**Data availability** Single-cell gene expression data generated in this study and analysis scripts are available on Zenodo at <https://doi.org/10.5281/zenodo.7852154>.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. A. Maynard, C.E. McCoach, J.K. Rotow, L. Harris, F. Haderk, D.L. Kerr et al., Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell* **182**(5), 1232–1251.e22 (2020)
2. B. Chen, C.R. Scurrah, E.T. McKinley, A.J. Simmons, M.A. Ramirez-Solano, X. Zhu et al., Differential pre-malignant programs and microenvironment chart distinct paths to malignancy in human colorectal polyps. *Cell* **184**(26), 6262–6280.e26 (2021)
3. Y.C. Cohen, M. Zada, S.Y. Wang, C. Bornstein, E. David, A. Moshe et al., Identification of resistance pathways and therapeutic targets in relapsed multiple myeloma patients through single-cell sequencing. *Nat. Med.* **27**(3), 491–503 (2021)
4. L. Ma, M.O. Hernandez, Y. Zhao, M. Mehta, B. Tran, M. Kelly et al., Tumor cell biodiversity drives microenvironmental reprogramming in liver cancer. *Cancer Cell* **36**(4), 418–430.e6 (2019)
5. C.J. Hanley, S. Waise, M.J. Ellis, M.A. Lopez, W.Y. Pun, J. Taylor et al., Single-cell analysis reveals prognostic fibroblast subpopulations linked to molecular and immunological subtypes of lung cancer. *Nat. Commun.* **14**(1), 387 (2023)
6. C.A. Stewart, C.M. Gay, Y. Xi, S. Sivajothi, V. Sivakamasundari, J. Fujimoto et al., Single-cell analyses reveal increased intratumoral heterogeneity after the onset of therapy resistance in small-cell lung cancer. *Nat. Cancer* **1**, 423–436 (2020)
7. M. Slyper, C.B.M. Porter, O. Ashenberg, J. Waldman, E. Drokhllyansky, I. Wakiro et al., A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* **26**(5), 792–802 (2020)
8. A. Janesick, R. Shelansky, A.D. Gottscho, F. Wagner, M. Rouault, G. Beliakoff et al., High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue. *bioRxiv* 2022.10.06.510405 (2022). <https://doi.org/10.1101/2022.10.06.510405>
9. A.F. Vallejo, K. Harvey, T. Wang, K. Wise, L.M. Butler, J. Polo et al., snPATHO-seq: unlocking the FFPE archives for single nucleus RNA profiling. *bioRxiv* 2022.08.23.505054 (2022). <https://doi.org/10.1101/2022.08.23.505054>
10. H. Chung, A. Melnikov, C. McCabe, E. Drokhllyansky, V.W. N., E.M. Magee et al., SnFFPE-Seq: towards scalable single nucleus RNA-Seq of formalin-fixed paraffin-embedded (FFPE) tissue. *bioRxiv* 2022.08.25.505257 (2022). <https://doi.org/10.1101/2022.08.25.505257>
11. Y. Hao, S. Hao, E. Andersen-Nissen, W.M. Mauck, S. Zheng, A. Butler et al., Integrated analysis of multimodal single-cell data. *Cell* **184**(13), 3573–3587.e29 (2021)
12. M.D. Young, S. Behjati, SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**(12), gaa151 (2020)
13. M. Schubert, B. Klinger, M. Klünemann, A. Sieber, F. Uhlitz, S. Sauer et al., Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **9**(1), 20 (2018)
14. A. Neuschulz, O. Bakina, V. Badillo-Lisakowski, P. Olivares-Chauvet, T. Conrad, M. Gotthardt et al., A single-cell RNA labeling strategy for measuring stress response upon tissue dissociation. *Mol. Syst. Biol.* **19**(2), e11147 (2023)
15. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J.P. Mesirov, P. Tamayo, The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015)
16. P. Bischoff, A. Trinks, B. Obermayer, J.P. Pett, J. Wiederspahn, F. Uhlitz et al., Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. *Oncogene* **40**(50), 6748–6758 (2021)
17. G.J. Yoshida, Regulation of heterogeneous cancer-associated fibroblasts: the molecular pathology of activated signaling pathways. *J. Exp. Clin. Cancer Res.* **39**(1), 112 (2020)
18. F. Castro, A.P. Cardoso, R.M. Gonçalves, K. Serre, M.J. Oliveira, Interferon-gamma at the crossroads of tumor immune surveillance or evasion. *Front. Immunol.* **9**, 847 (2018)
19. M.G. Dorrington, I.D.C. Fraser, NF- $\kappa$ B signaling in macrophages: dynamics, crosstalk, and signal integration. *Front. Immunol.* **10**, 705 (2019)
20. M. Orecchioni, Y. Ghosheh, A.B. Pramod, K. Ley, Macrophage polarization: different gene signatures in M1(Lps+) vs. classically and M2(LPS-) vs. alternatively activated macrophages. *Front. Immunol.* **10**, 1084 (2019)
21. 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. *Nat. Med.* **24**(7), 978–985 (2018)
22. A.M. Gocher, C.J. Workman, D.A.A. Vignali, Interferon- $\gamma$ : teammate or opponent in the tumour microenvironment? *Nat. Rev. Immunol.* **22**(3), 158–172 (2021)
23. R.C. Jones, J. Karkanas, M.A. Krasnow, A.O. Pisco, S.R. Quake, J. Salzman et al., The Tabula Sapiens: a multiple-organ, single-cell

- transcriptomic atlas of humans. *Science* **376**(6594), eab14896 (2022)
24. J.M. Chan, Á. Quintanal-Villalonga, V.R. Gao, Y. Xie, V. Allaj, O. Chaudhary et al., Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. *Cancer Cell* **39**(11), 1479–1496.e18 (2021)
25. K.J. Travaglini, A.N. Nabhan, L. Penland, R. Sinha, A. Gillich, R.V. Sit et al., A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**(7835), 619–625 (2020)

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