

# Heterogeneity and targeted therapy-induced adaptations in lung cancer revealed by longitudinal single-cell RNA sequencing

Ashley Maynard<sup>1,A</sup>, Caroline E. McCoach<sup>2,3,A</sup>, Julia K. Rotow<sup>4,B</sup>, Lincoln Harris<sup>1,B</sup>, Franziska Haderk<sup>2,3,5</sup>, Lucas Kerr<sup>2</sup>, Elizabeth A. Yu<sup>2</sup>, Erin L. Schenk<sup>6</sup>, Weilun Tan<sup>1</sup>, Alexander Zee<sup>1,7</sup>, Michelle Tan<sup>1</sup>, Philippe Gui<sup>2,3</sup>, Tasha Lea<sup>3</sup>, Wei Wu<sup>2</sup>, Anatoly Urisman<sup>8</sup>, Kirk Jones<sup>8</sup>, Rene Sit<sup>1</sup>, Pallav K. Kolli<sup>9</sup>, Eric Seeley<sup>2</sup>, Yaron Gesthalter<sup>2</sup>, Daniel D. Le<sup>1</sup>, Kevin A. Yamauchi<sup>1</sup>, David Naeger<sup>10,11</sup>, Nicholas J. Thomas<sup>2</sup>, Anshal Gupta<sup>2</sup>, Mayra Gonzalez<sup>2</sup>, Hien Do<sup>2</sup>, Lisa Tan<sup>2</sup>, Rafael Gomez-Sjoberg<sup>1</sup>, Matthew Gubens<sup>2</sup>, Thierry Jahan<sup>2</sup>, Johannes R. Kratz<sup>12</sup>, David Jablons<sup>12</sup>, Norma Neff<sup>1</sup>, Robert C. Doebele<sup>6</sup>, Jonathan Weissman<sup>5,13</sup>, Collin M. Blakely<sup>2,3,✉</sup>, Spyros Darmanis<sup>1,✉</sup>, and Trevor G. Bivona<sup>2,3,5,✉</sup>

<sup>1</sup>Chan-Zuckerberg Biohub San Francisco, CA

<sup>2</sup>Department of Medicine, University of California, San Francisco, CA

<sup>3</sup>University of California, San Francisco, Helen Diller Family Comprehensive Cancer Center, San Francisco, CA

<sup>4</sup>Lowe Center for Thoracic Oncology, Dana-Farber Cancer Institute; Boston, MA

<sup>5</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA

<sup>6</sup>Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO

<sup>7</sup>Department of Biomolecular Engineering, University of California, Santa Cruz, CA

<sup>8</sup>Department of Pathology University of California, San Francisco, CA

<sup>9</sup>Department of Radiology and Biomedical Imaging; University of California, San Francisco, CA

<sup>10</sup>Denver Health Medical Center, Denver, CO

<sup>11</sup>Department of Radiology, University of Colorado, Aurora, CO

<sup>12</sup>Department of Surgery, University of California, San Francisco, CA

<sup>13</sup>Howard Hughes Medical Institute, University of California, San Francisco, CA

<sup>A</sup>These authors contributed equally

<sup>B</sup>These authors contributed equally

Lung cancer, the leading cause of cancer mortality, exhibits heterogeneity that enables adaptability, limits therapeutic success, and remains incompletely understood. Single-cell RNA sequencing (scRNAseq) of metastatic lung cancer was performed using 44 tumor biopsies obtained longitudinally from 27 patients before and during targeted therapy. Over 20,000 cancer and tumor microenvironment (TME) single-cell profiles exposed a rich and dynamic tumor ecosystem. scRNAseq of cancer cells illuminated targetable oncogenes beyond those detected clinically. Cancer cells surviving therapy as residual disease (RD) expressed an alveolar-regenerative cell signature suggesting a therapy-induced primitive cell state transition, whereas those present at on-therapy progressive disease (PD) upregulated kynurenine, plasminogen, and gap junction pathways. Active T-lymphocytes and decreased macrophages were present at RD and immunosuppressive cell states characterized PD. Biological features revealed by scRNAseq were biomarkers of clinical outcomes in independent cohorts. This study highlights how therapy-induced adaptation of the multi-cellular ecosystem of metastatic cancer shapes clinical outcomes.

single cell RNA sequencing | NSCLC | EGFR | ALK | immune microenvironment | mutation | heterogeneity | landscape

Correspondence:

✉collin.blakely@ucsf.edu

✉spyros.darmanis@czbiohub.org

✉trever.bivona@ucsf.edu

## Introduction

Heterogeneity is a property of many biological systems and diseases such as cancer. Biological plasticity in cancer cells is one form of heterogeneity that allows for early adaptation to treatment and limits the success of precision approaches

for cancer treatment (Yuan et al., 2019; Xue et al., 2017). In addition to cancer-cell intrinsic heterogeneity, cells within the tumor microenvironment (TME) further contribute to tumor heterogeneity in a cancer-cell extrinsic manner. While these tumor compartments and tumor heterogeneity have been characterized in many cancer subtypes (Gerlinger et al., 2012; Alexandrov et al., 2013; Lawrence et al., 2013; Brannon et al., 2014; Lee et al., 2014; Vignot et al., 2015; Hata et al., 2016), our understanding of how these properties evolve and interact longitudinally in response to systemic treatment remains incomplete, particularly in metastatic tumors.

Many oncogene-driven cancers such as those with oncogenic alterations in *EGFR*, *ALK*, *ROS1*, and *BRAF* are treated with targeted therapies against the cognate oncoprotein or signaling pathway. This has led to improvements in clinical outcomes of metastatic solid cancers such as lung cancer and melanoma as well as hematologic malignancies (Flaherty et al., 2012; Mok et al., 2009; Shaw et al., 2013). A paradigm for molecular therapeutics and the study of tumor evolution is lung cancer, the leading cause of cancer mortality worldwide (Siegel et al., 2018), with non-small cell lung cancer (NSCLC) constituting the major subtype (Swanton and Govindan, 2016). Despite the success of targeted therapy in cancers such as NSCLC, tumors typically respond incompletely in the form of residual disease that is a prelude to tumor progression due to therapy-induced tumor evolution. Bulk tumor sampling after tumor progression on targeted therapy has identified resistance mechanisms and demonstrated that tumors become increasingly molecularly heterogeneous after therapy progression (Blakely et al.,

2017a; Camidge et al., 2014; McCoach et al., 2018; Rotow and Bivona, 2017).

Single-cell RNA sequencing (scRNAseq) is one established approach to dissect the heterogeneity of complex biological systems (Chung et al., 2017; Darmanis et al., 2017; Tirosh et al., 2016). Comparison of single-cell transcriptional profiles of individual cells at different treatment timepoints could lead to deeper insight into the evolution of cell states of both cancer cells and TME cells occurring during treatment. There is currently a paucity of single-cell transcriptome studies that sample metastatic solid malignancies and prior scRNAseq studies of metastatic disease largely focused on individual treatment timepoints (Chung et al., 2017a; Darmanis et al., 2017; Lambrechts et al., 2018a; Patel et al., 2014a; Tirosh et al., 2016a; Wang et al., 2019; Zhang et al., 2019a). This is due, in part, to challenges associated with obtaining high-quality biopsies from metastatic tumors, instead of the larger tumor specimens that can be obtained from surgically resected earlier-stage disease. Obtaining multiple tumor biopsies longitudinally at different timepoints during systemic therapy of patients, individually and as a cohort, with metastatic solid tumors is an additional challenge to overcome to allow for the study of tumor adaptation to systemic therapy.

Beyond the analysis of tumor biopsies obtained both before treatment and at tumor progression, analysis of advanced-stage tumor samples that reflect the residual disease (RD) state during therapy could reveal clinically relevant biological events that permit cancer cell persistence. Developing a deeper understanding of evolving cell-state changes during treatment, and particularly at the RD state, is essential to improve the durability and magnitude of response to precision therapies for cancer. RD sample analysis could help identify therapeutic targets in cancer and TME cells for exploitation in a window-of-opportunity during therapy to preempt the subsequent evolution to absolute drug resistance.

We performed scRNAseq analyses on human NSCLC tumor biopsies collected at different treatment states. Our goal was to help elucidate cancer cell phenotypic plasticity and the contribution and dynamic nature of the TME during therapy that individually, and collectively, promote RD and subsequent progressive disease during targeted therapy. To accomplish this, we instituted a unique clinical tumor rebiopsy protocol wherein advanced-stage NSCLC samples were obtained from patients longitudinally before systemic targeted therapy (*treatment naïve, TN*), at the residual disease state during treatment response (*residual disease, RD*) and upon the subsequent establishment of acquired drug resistance (*progression, PD*). We developed a custom tissue processing and analytic framework adapted to the relatively small and challenging advanced-stage lung tumor biopsy samples to: 1) evaluate the expressed mutational landscape of tumor samples, 2) characterize transcriptional gene signatures unique to different treatment timepoints, 3) investigate the properties of the TME as it evolves during treatment, and 4) describe the interplay between cancer cells and the TME during targeted

therapy.

## Results

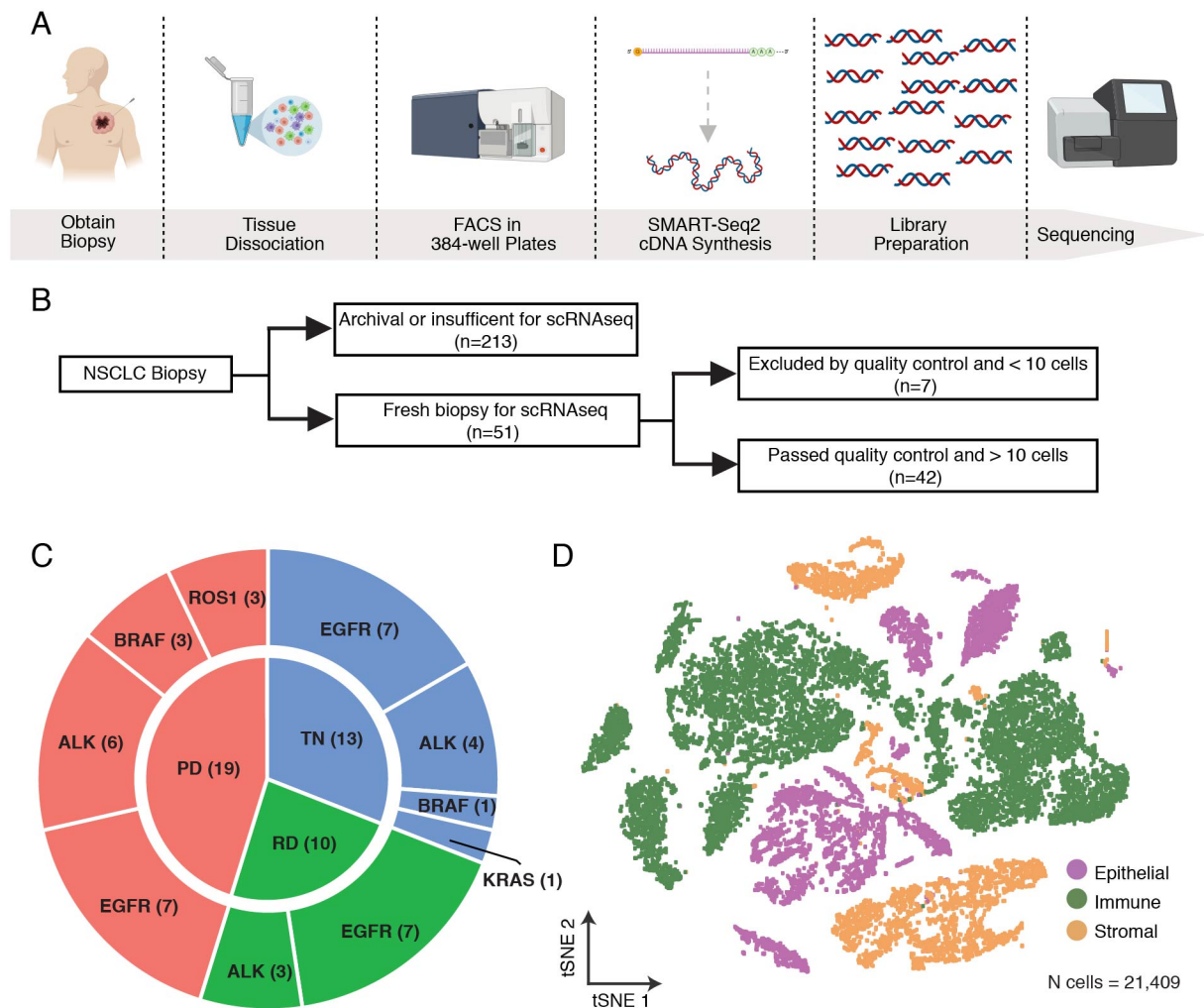
**Longitudinal scRNAseq analysis of advanced-stage NSCLC during targeted therapy.** We used scRNAseq to profile 44 samples (41 lung adenocarcinomas, 1 squamous cell carcinoma and 2 normal adjacent tissues) (Figure 1A), corresponding to 27 individual patients. We used a rapid processing workflow customized to isolate viable single cells primarily from fine needle aspirates (FNA), core needle biopsies, and thoracentesis samples in addition to standard surgical resections (Figure 1B). Tumor biopsy samples were collected and immediately processed before and during systemic targeted therapy (Figure 1C, Supplemental Table 1).

Gene expression profiles of 21,409 cells were retained after quality control filtering (Supplemental Figure 1A). Following gene expression normalization, we performed principal component analysis (PCA) using a set of variably expressed genes across all cells. We clustered cells using graph-based clustering on the informative principal component space (npcs=20). The resulting cell clusters were annotated as immune cells, stromal cells (fibroblasts, endothelial cells, and melanocytes), and epithelial cells (Figure 1D) through gene expression analysis of established marker genes (Lambrechts et al., 2018; Schiller et al., 2019; Tabula Muris Consortium et al., 2018; Treutlein et al., 2014) (Supplemental Table 2). Epithelial cells (n=5,109), were subset and re-clustered into 31 discrete epithelial clusters (Supplemental Figure 1B).

**Clustering-based copy number variation resolves cancer from non-cancer epithelial cells.** Given the known association between cancer and large-scale chromosomal alterations, we utilized an established analytical framework to infer copy number variation (CNV) from RNA expression to classify epithelial cells as either cancer or non-cancer (Patel et al., 2014; Puram et al., 2017; Tirosh et al., 2016; Venteicher et al., 2017; <https://github.com/broadinstitute/inferCNV>). Compared to fibroblasts and endothelial cells, which were used as controls, cancer cells displayed larger excursions from relative expression intensities in multiple regions of the genome (Supplemental Figure 1C). Following identification of cancer and non-cancer epithelial cells, non-cancer epithelial cell clusters (n=13) were further annotated as either pulmonary alveolar type 2 (AT2), pulmonary alveolar type 1 (AT1), hepatocytes, club cells, basal cells, ionocytes or stromal cells (Supplemental Figure 1D).

As noted by others (Zhang et al., 1997), we found that these cancer cells expressed an elevated number of unique genes compared to non-cancer cells (Supplemental Figure 1E). This difference in the number of uniquely expressed genes was not explained by sequencing depth (Pearson correlation = 0.17).

All cancer cells (n=3,620) were subset and re-clustered, resulting in 24 unique clusters across 38 total samples (Supplemental Figure 2A, 2B). For each cluster, we calculated patient occupancy, defined as the number of cells of the highest



**Fig. 1. Patient characteristics and experimental overview** Tissue processing pipeline for scRNAseq. Patient samples were first mechanically, and enzymatically disaggregated and then single cells were sorted into microtiter plates using FACS. cDNA synthesis was performed using the Smart-seq2 protocol and libraries were sequenced on Illumina platforms. (B) Consort diagram of sample collection. A total of 51 biopsies were processed. Only samples with cells that passed quality control were used ( $n = 42$ ). (C) Circle plot illustrating the clinically identified oncogenic driver (outer circle) and treatment timepoint (inner circle) for each sample. (D) t-SNE plot of all 21,409 cells colored by their cellular identity as inferred from gene expression profiling (Epithelial cells ( $n = 5,109$ ), Immune Cells ( $n = 12,077$ ), Stromal Cells ( $n = 4,223$ )).

contributing individual patient over the total number of cells for that cluster for both non-cancer and cancer epithelial cells (Supplemental Figure 2C). The majority of the cancer cell clusters are patient specific, having high patient occupancy scores, similar to findings reported in other studies (Chung et al., 2017; Darmanis et al., 2017; Puram et al., 2017; Tirosh et al., 2016). Conversely, non-cancer cell types exhibited lower patient occupancy, with multiple patients contributing cells to each cluster (Supplemental Figure 2C). Thus, patient-specific malignant cell clustering reflects the unique molecular signatures of an individual patient's tumor rather than potential technical issues.

**scRNA-seq analysis reveals a rich complexity of expressed gene alterations in cancer cells.** Additional genetic alterations detected by bulk tumor DNA-based analysis can co-exist with a primary targetable oncogenic “driver” alteration (e.g. oncogenic *EGFR*, *ALK*, *BRAF*, *ROS1*, *KRAS* and *MET*) and may help promote tumor progression and

limit therapy response (Blakely et al., 2017; Kim et al., 2019; Scheffler et al., 2019; Yang et al., 2019). We queried scRNAseq transcripts from each cancer cell to identify somatic alterations, namely single nucleotide polymorphisms (SNPs), insertions/deletions (indels), and gene fusions (Figure 2A, 2B, 2C). We identified 17 of 38 tumor biopsy samples that contained cancer cells harboring the clinically-identified oncogenic driver (Figure 2B, Supplemental Table 3), consistent with the known potential drop-out occurrence in scRNAseq analyses (Kharchenko et al., 2014). Within these 17 samples in which we found at least one cell harboring the clinically actionable mutation, there were 7 samples in which at least one other cell harbored an additional oncogenic alteration not detected by clinical-grade testing of a tumor sample from the same patient (i.e. occult genetic alterations) (Figure 2B). One such example from our data is illustrated by sample LTS47. This tumor was determined to harbor an *EML4-ALK* oncogenic gene rearrangement by clinical-grade bulk DNA analysis. However, scRNAseq additionally revealed that this

sample contained cancer cells harboring *KRAS* G13D and *KRAS* G12C occult mutations; neither population of *KRAS* mutant cells showed evidence of the *ALK* gene rearrangement (Supplemental Figure 2D). This sample was obtained from the patient after multiple lines of therapy which likely allowed for evolution of multiple mechanisms of resistance, including the 2 different oncogenic forms of *KRAS*, which are known mechanisms of resistance to *ALK* inhibitor treatment (Doebele et al., 2012; Hrustanovic and Bivona, 2015; Shaw and Engelman, 2013).

We also queried scRNAseq transcripts for mutations from the COSMIC (Catalogue of Somatic Mutations In Cancer) lung adenocarcinoma tier 1 mutations (Supplemental Table 2), (Forbes et al., 2017; Shihab et al., 2015), focusing on mutations that are classified as 'pathogenic'. Many of the mutations we identified had not been previously reported by the clinical-grade assay conducted on the patient's tumor despite having been included in the clinical panel (Figure 2C, Supplemental Tables 2 and 3). Though this may reflect the difference in biopsy technique or tumor clonality at the time of clinical testing, these results also demonstrate that clinical-grade bulk DNA-based testing may underestimate tumor heterogeneity. To assess the clinical outcomes of patients harboring multiple oncogenic alterations and to determine the broader translational impact of our findings, we utilized the MSK-Impact NSCLC dataset, which contains the clinical and sequencing data for over 800 advanced-stage NSCLC patients (Zehir et al., 2017). Those patients with greater than or equal to 2 mutations from the tier one COSMIC mutations found in the scRNAseq dataset (mutation high), had significantly lower overall survival ( $p < 0.01$ ) compared to those with less than 2 mutations (mutation low) Patients with more than two cosmic tier 1 mutations (Figure 2D). Thus, scRNAseq analysis can complement current clinical-grade bulk DNA-based assays to provide increased granularity into cancer cell genomic heterogeneity and provides insight into the mutational landscape that is expressed at the RNA level in cancer cells of advanced-stage tumors, with clinical implications.

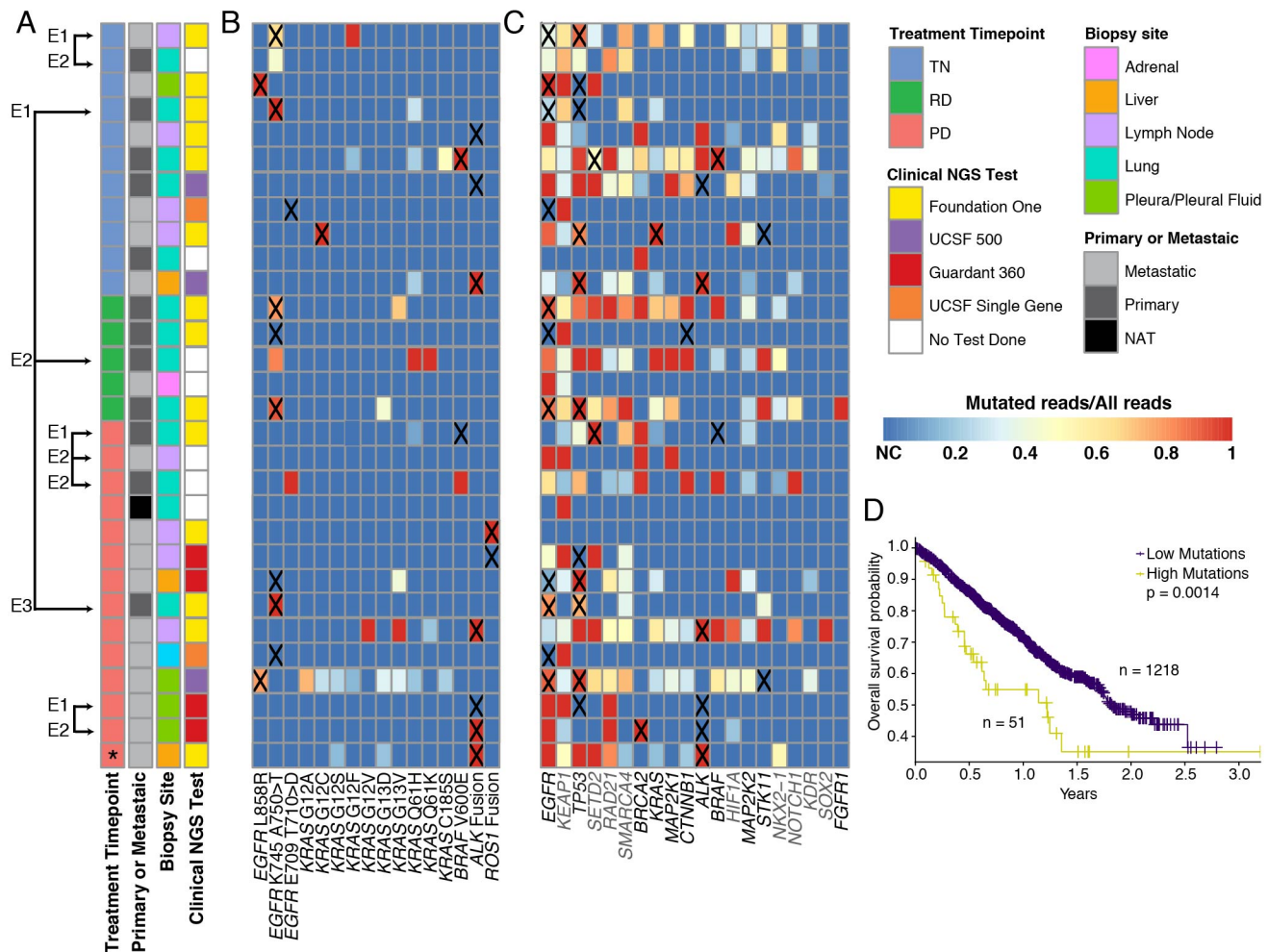
**Transcriptional differences between TN and RD cancer cells detected by scRNAseq analysis reveal cell state-specific biological programs.** We hypothesized that defining the biological programs activated in cancer cells during therapy response may identify signaling pathways that promote adaptation and survival of cancer cells that comprise residual disease during initial treatment. We compared the transcriptional profiles of individual cancer cells obtained from tumor samples from TN to RD (Supplemental Table 4) and focused our attention on the 691 significantly ( $p < 0.05$ ) upregulated genes in RD cancer cells as a proxy for evidence of pathway activation. We found individual genes associated with specific cancer-associated pathways such as survival, cell growth, invasion and metastasis (Supplemental Table 5). As a control and consistent with the expectation that during targeted treatment cancer cells surviving and persisting drug therapy are generally less proliferative (Sharma et al., 2010;

Zhu et al., 2001), we found that RD cancer cells expressed decreased proliferation marker genes compared to both TN and PD (Supplemental Figure 3A) (Hsiao et al., 2019).

Interestingly, we identified an alveolar cell gene expression signature composed of 17 established gene markers of alveolar cells (Vieira Braga et al., 2019; Wade et al., 2006) that showed significantly ( $p < 0.0001$ ) increased expression in RD versus TN timepoints (Figure 3A, Supplemental Figure 3B, Supplemental Table 2). Alveolar cells are partitioned into alveolar type 1 (AT1) and type 2 (AT2) subtypes and form the lining of the lung alveoli. AT2 cells produce surfactants and can act as stem-like progenitor cells (Nabhan et al., 2018) which become active and proliferate in the setting of diverse types of lung injury. AT2 cells are also suspected to be the cell of origin in oncogene-driven lung cancers (Desai et al., 2014; Hanna and Onaitis, 2013; Nabhan et al., 2018). AT1 cells are the dominant population in alveoli and mediate gas exchange and, when injured or dying, can release proliferation and regenerative signals (Desai et al., 2014). The alveolar signature we detected in the cancer cells at RD includes increased expression of both AT1- and AT2-associated genes (Supplemental Table 2), including *AQP4*, *SFTPB/C/D*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC* for AT2 cells (Desai et al., 2014; Liu et al., 2003; Nabhan et al., 2018; Wade et al., 2006; Xu et al., 2016; Zhou et al., 2018) and *AGER* and *HOPX* for AT1 cells (Nabhan et al., 2018a; Serveaux-Dancer et al., 2019) (Supplemental Figure 3B). Additional analysis demonstrated that the alveolar cell state we identified in cancer cells was not derived from misannotated non-cancer alveolar cells within our cancer cell populations (Supplemental Figure 3C).

We validated the activation of the alveolar cell signature at RD using orthogonal approaches. First, we used an established preclinical model consisting of patient-derived EGFR-mutant NSCLC cells (PC9) (Lee et al., 1985) to develop analogs of the TN, RD, and PD clinical states and measured the expression of a subset of the alveolar cell state genes (by RT-PCR) detected at RD in the clinical tumor samples. Cells were treated with vehicle (DMSO) or a standard EGFR inhibitor (osimertinib) and sampled under control conditions (day 2 in DMSO), during RD (days 7 and 19), and at PD (day 70, acquired resistance). We found significantly increased expression of *NKX2-1*, a hallmark alveolar cell signature gene upregulated in the RD clinical samples, at days 7 and 19 compared to control and day 70 (Figure 3B). This demonstrates that our findings from the scRNAseq analysis can be reproduced under controlled conditions *in vitro*. Furthermore, immunohistochemistry (IHC) analysis showed induction of AQP4 protein expression, another marker of the alveolar cell signature, at the plasma membrane of RD clinical samples compared to both TN and PD clinical samples (Figure 3C, Supplemental Figure 3D-E). The collective data support key scRNAseq findings arising from the clinical tumor biopsies.

We next determined whether the alveolar cell signature was more broadly clinically-relevant by examining whether it is a



**Fig. 2. scRNAseq infers patient mutational status and reveals a complex mutational landscape in cancer cells** (A) Clinical characteristics of the 30 NSCLC samples from which oncogenic drivers and/or COSMIC tier 1 genes were identified. Columns indicate treatment timepoint (TN, RD, PD), primary or metastatic sample origin, biopsy site, and clinical NGS test performed. Longitudinal samples from the same patient are indicated by chronological encounter numbers (e.g. E1, E2) and grouped with arrows. (B) Cancer cell mutational landscape for each patient sample as determined by scRNAseq represented as a heatmap. Color indicates the number of mutant reads for each genomic region and sample divided by the total number of reads for that region in that sample. The expected oncogenic driver, as determined by a clinical NGS assay, is marked with an "X". NC-No Coverage for the specific sample. (C) Mutational landscape of COSMIC tier 1 genes. Color indicates the number of mutant reads for each genomic region and sample divided by the total number of reads for that region in that sample. Genes names in black indicate inclusion in all clinical NGS assays (aside from the UCSF single gene test), whereas gene names in grey are included in only select assays (Supplemental Table 2). (D) Kaplan-Meier plot showing overall survival of 1269 NSCLC patients within the MSK-Impact dataset. Patients were stratified by high ( $\geq 2$ ) and low ( $< 2$ ) mutations from the 141 mutations that are found in both the MSK-Impact dataset and panel (C).

biomarker of patient survival in the TCGA lung adenocarcinoma bulk RNAseq dataset generated by the TCGA Research Network (<https://www.cancer.gov/tcga>, Cancer Genome Atlas Research Network et al., 2013). We found a significant ( $p < 0.0001$ ) association between high expression of our alveolar cell type signature and improved overall survival (OS) (Figure 3D). These collective findings support the assertion that there is a distinct alveolar cell type gene expression signature characterizing RD cancer cells that is associated with improved patient survival. A plausible model is that our identified alveolar signature reflects a cell injury and repair signal in which the machinery of injury-response in malignant AT1/2 cells activates a regenerative signaling pathway and transition to a more primitive cell state. This could serve to aid repair from injury from cell death during treatment to support cancer cell persistence, while at the same time generat-

ing a less aggressive malignant state. This is consistent with notion that RD represents a "persister" cell state observed in preclinical models of slow-cycling cancer cells that survive without rapid proliferation (as in Supplemental Figure 3A), as a prelude to the onset of aggressive tumor progression upon absolute drug resistance (Hata et al., 2016).

The molecular details of the alveolar and cell injury repair signature are notable. In our RD cohort, WNT/ $\beta$ -catenin-associated pathway genes *AFF3*, *SUSD2*, and *CAVI* exhibited increased expression (Supplemental Table 5). *AFF3* and *SUSD2* are activated downstream targets of the WNT pathway (Lefèvre et al., 2015; Umeda et al., 2018; Xu et al., 2018) while *CAVI* can promote nuclear localization of  $\beta$ -catenin (*CTNNB1*) and transcriptional activation of the WNT/ $\beta$ -catenin pathway (Yu et al., 2014). In NSCLC, the WNT/ $\beta$ -catenin signaling pathway contributes to tumorige-

nesis (Juan et al., 2014; Nakayama et al., 2014; Pacheco-Pinedo et al., 2011), repair, and regeneration after cell injury (Huch et al., 2013; Tammela et al., 2017). The self-renewal and injury response in AT2 cells specifically can utilize the WNT/ $\beta$ -catenin signaling pathway (Nabhan et al., 2018b; Stewart, 2014). Additionally, in *EGFR*-mutant NSCLC the WNT/ $\beta$ -catenin pathway may limit EGFR inhibitor response and can lead to survival of a persisting cell population during EGFR inhibitor therapy *in vitro* (Arasada et al., 2018; Blakely et al., 2017; Casás-Selves et al., 2012; Nakayama et al., 2014). Overall, the RD state is characterized by signals of cellular injury and survival which act, in part, through the WNT/ $\beta$ -catenin pathway, which may be therapeutically targetable (Krishnamurthy and Kurzrock, 2018).

### **Transcriptional differences between TN and PD cancer cells reveal immune modulation and cellular invasion as key features of cancer progression.**

When comparing cancer cells from TN and PD samples, we found 958 differentially expressed genes with higher expression in PD cancer cells (Supplemental Table 4). Within those genes, we identified genes involved in the kynurenine pathway and multiple genes and pathways associated with tumor invasion, metastasis, cell viability, and inflammation (Supplemental Table 5). We observed a significant ( $p < 0.0001$ ) increase in the expression of *IDO1*, *KYNU*, and *QPRT*, which are each involved in the kynurenine pathway, in the PD versus TN cancer cells (Figure 3E, Supplemental Figure 3F, Supplemental Table 2). The kynurenine pathway metabolizes tryptophan generating catabolite intermediates that are further processed to replenish cellular NAD<sup>+</sup> and support immunomodulatory roles (Triplett et al., 2018; Zhai et al., 2018). Expression of *IDO1*, *KYNU*, and *QPRT* can result in immunosuppressive behavior (Triplett et al., 2018). Our data indicate that cancer cells within PD tumors may inhibit the activity of the immune system during targeted therapy. The identification of this pathway as a mediator of immune suppression within PD tumors has important potential therapeutic implications, as *IDO1* is known to be upregulated in many cancers (Cheong and Sun, 2018; Hornyák et al., 2018; Liu et al., 2018). Multiple clinical trials have attempted to block this pathway using *IDO1* inhibitors as a monotherapy as well as in combination with immune checkpoint inhibitors or hormone therapy (Ricciuti et al., 2019), albeit with limited success. *QPRT* also exhibited increased expression specifically at day 70 of EGFR inhibitor treatment (i.e. acquired resistance and the analog of clinical PD) of our *in vitro* model using PC9 cells (Figure 3F), reinforcing the assertion that this pathway is indicative of cancer progression under the selective pressure of treatment.

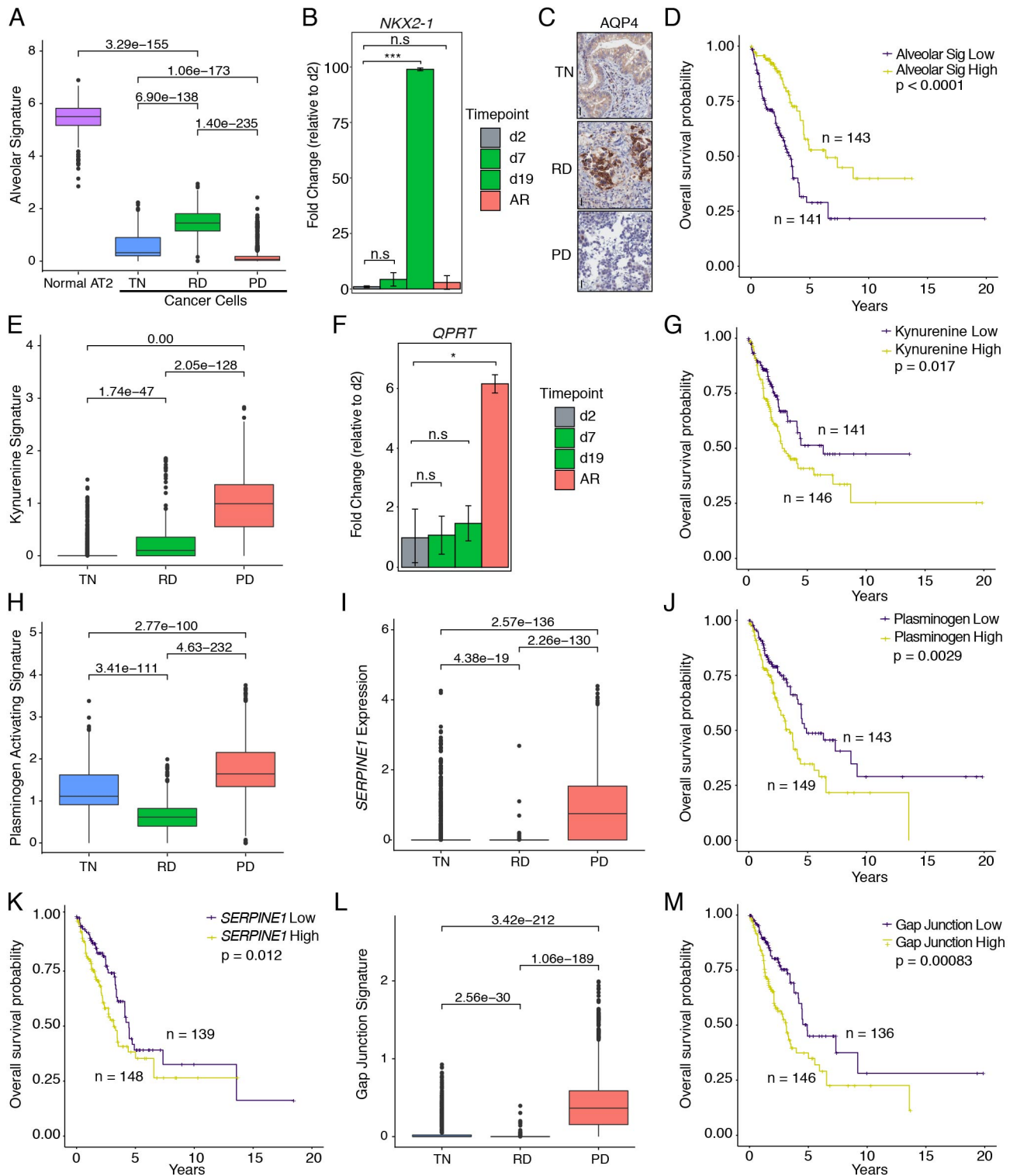
To further demonstrate the clinical relevance of the kynurenine pathway, we again used the TCGA lung adenocarcinoma RNAseq dataset. Increased expression of the kynurenine pathway signature was a biomarker of worse OS ( $p = 0.017$ ) (Figure 3G). This is consistent with the notion that activation of this pathway leads to immunosuppression and an inability of the immune system to effectively surveil and

eradicate cancer cells.

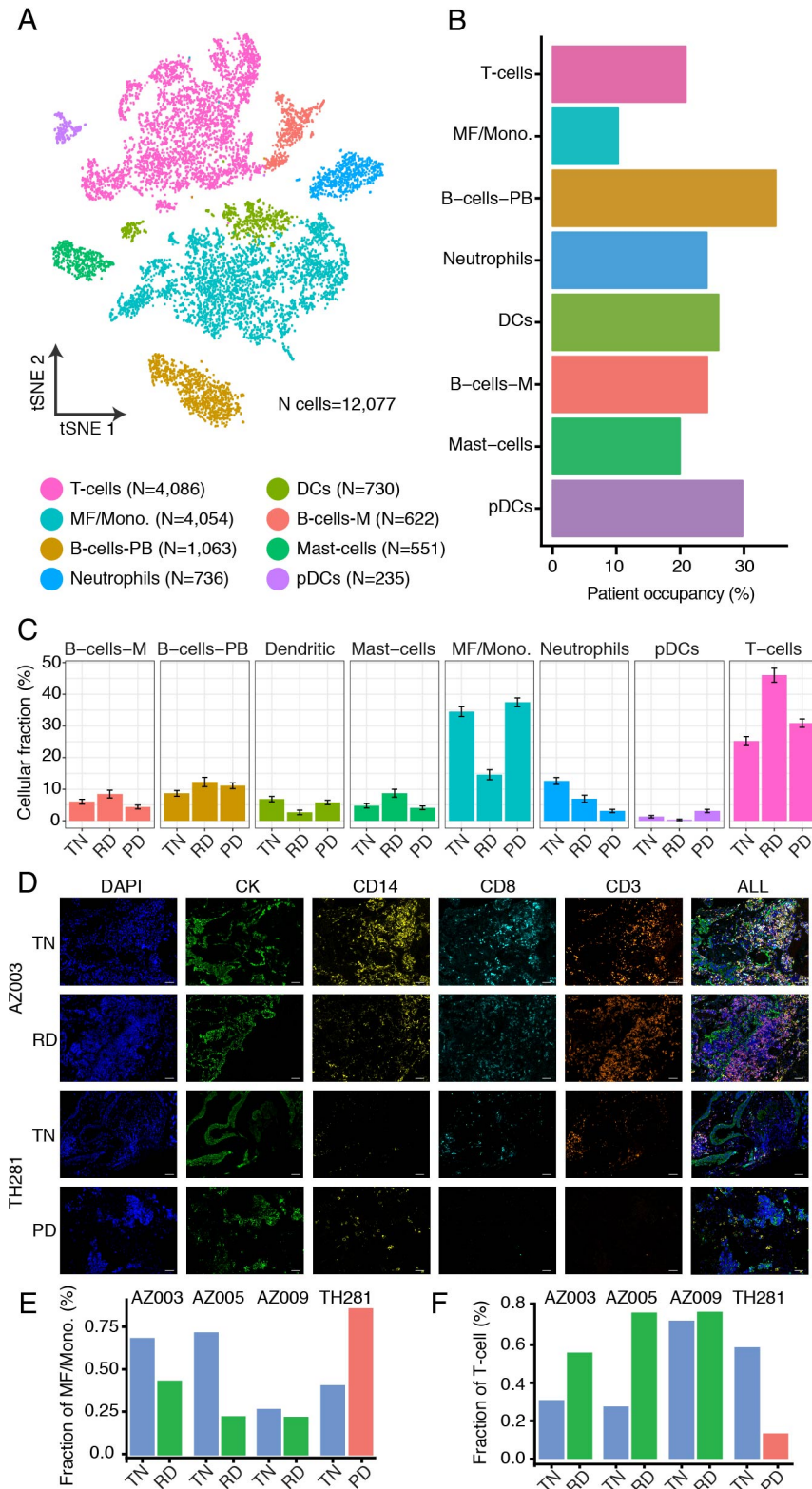
### **Longitudinal scRNAseq profiles of cancer cells change from RD to PD.**

We compared cancer cells from RD and PD patient samples to elucidate the differences that occur during the outgrowth of PD from RD and found a total of 2011 genes which had significantly ( $p < 0.001$ ) increased expression in either RD or PD ( $N_{RD}=874$ ,  $N_{PD}=1,137$ ) (Supplemental Table 4). Among the differentially overexpressed genes at RD were genes associated with the alveolar cell signature, cell growth, differentiation, and cell motility (Supplemental Table 5). RD cancer cells overexpress surfactant genes (*SFTPB/C/D* and *SFTA3*) which are part of the alveolar cell signature (Figure 3A, Supplemental Figure 3B) (Desai et al., 2014; Treutlein et al., 2014; Wang et al., 2018). Furthermore, RD cancer cells also overexpress the putative tumor suppressor *DLCI*, which is a member of the RhoGAP family of negative regulators of Rho GTPases and can be downregulated in NSCLC (Healy et al., 2008; Yuan et al., 2004). High expression of *DLCI* was associated with an improved survival in NSCLC (Sun et al., 2019). *NKX2-1* and *NFIX* were overexpressed in RD cancer cell and are associated with decreased cell motility (Ge et al., 2018; Rahman et al., 2017a; Winslow et al., 2011). Low expression of *NKX2-1* leads to loss of differentiation and enhanced tumor seeding ability (Winslow et al., 2011). The collective findings arising from this and the previous RD cancer cell analyses suggest that an injury-repair and regenerative cell state may promote cancer cell indolence, increased tumor control, and improved clinical outcomes.

By contrast, PD cancer cells differentially overexpressed genes associated with invasion, cell-to-cell communication, differentiation and immune modulation (Supplemental Table 5). Several genes in the plasminogen activation pathway were significantly overexpressed (*ANXA2*, *PLAT*, *PLAUR*, *PLAU*) (Figure 3H) along with the plasminogen inhibitor *SERPINE1* (*PAI1*) ( $p < 0.0001$ , Figure 3I, Supplemental Figure 3G). *ANXA2* and *PLAUR* are the receptor proteins in the plasminogen activation cascade and involved in inflammation, angiogenesis, invasion and metastasis, via degradation of the extracellular matrix (Kubala et al., 2018; Zhu et al., 2017). Signaling is initiated when *ANXA2* or *PLAU* binds to *PLAT* (uPa) or *PLAU* (uPa), respectively. Plasminogen is then degraded to plasmin through the activity of *PLAT* and/or *PLAU* leading to activation of metalloproteinases and degradation of fibrin. *SERPINE1* shows increased expression in a number of cancer subtypes and plays important roles in cell adhesion, invasion, tumor vascularization, radio-resistance, and immunosuppression (Kubala et al., 2018; Zhu et al., 2017). High expression of the plasminogen activation signature correlated with worse OS ( $p = 0.0029$ ) within the TCGA lung adenocarcinoma RNAseq dataset (Figure 3J). Similarly, in this independent dataset high expression of *SERPINE1* was associated with worse OS ( $p = 0.012$ ) (Figure 3K). Interestingly, EGFR inhibitor therapy can induce expression of *SERPINE1* and EGFR mutation positive patients with greater than two-fold induction of *SERPINE1* (*PAI1*) plasma levels



**Fig. 3. Differential gene-expression analysis between treatment timepoints reveals treatment stage specific transcriptional signatures** (A) Boxplots showing the expression level of the alveolar signature across different treatment timepoints as well as non-cancerous AT2 cells from our cohort. (B) Fold change expression of *NKX2-1* as quantified by RT-PCR in PC9 cells after treatment with vehicle (2 days), osimertinib at 7 days (RD), 19 days (RD) and at the acquired resistance state for this cell line, day 70 (see methods) (AR), \*\*\* indicates  $p < 0.001$ . (C) Representative IHC images of TN, RD and PD tumor tissue sections stained for AQP4 demonstrating increased expression at the RD timepoint. Scale bars correspond to 50 $\mu$ m. See Supplemental Figure 3D, E for validation cohort. (D) Kaplan-Meier plot of the relationship between the alveolar signature and patient OS within the TGCA dataset. Patients were stratified by high ( $n=143$ ) and low ( $n=141$ ) signature expression. (E) Box plots showing the expression levels of the kynurenine signature expression across different treatment timepoints. (F) Fold change expression of *QPRT* as quantified by RT-PCR in PC9 cells after treatment with osimertinib as in (B) (see methods) (AR), \* indicates  $p < 0.05$ . (G) Kaplan-Meier plot of the relationship between the kynurenine signature and patient OS within the TGCA dataset. Patients were stratified by high ( $n=146$ ) and low ( $n=141$ ) signature expression. (H, I) Box plots showing the expression levels of the plasminogen activation, *SERPINE1* pathway signatures across different treatment timepoints. (J, K) Kaplan-Meier plots of relationships between the plasminogen activating pathway signature and *SERPINE1* and patient OS within the TGCA dataset, respectively. Patients were stratified by high and low signature expression. (L) Box plot showing the expression levels of the gap junction signature across treatment timepoints. (M) Kaplan-Meier plots of relationships between the gap junction signature and patient OS within the TGCA dataset. Patients were stratified by high and low signature expression.



**Fig. 4. Changes in the composition of the tumor microenvironment within each tumor** (A) t-SNE plot of all immune cells colored by immune cell type. Number of cells for each cell type is provided on the figure legend. (B) Patient occupancy for each immune cell type. (C) Fractional changes for each immune cell type across the three treatment states. Error bars indicate the 95% confidence interval for the calculated relative frequencies. (D) Representative in situ immunofluorescence images of changes from TN to RD and TN to PD in tumor tissue sections from two separate samples at two separate timepoints; AZ003(TN and RD), TH281 (TN and PD). Scale bars correspond to 50  $\mu$ m. (E and F) Quantification of the fractional changes of macrophages and T-cells, respectively, across treatment timepoints from the images in D and Supplemental Figure 5F.



during EGFR inhibitor treatment demonstrate shorter progression free survival (Arasada et al., 2018). Collectively, our longitudinal scRNAseq-based findings shed light on the clinical relevance and potential role of the plasminogen activation cascade in inferior clinical outcomes and targeted therapy resistance.

Additionally, we found several gap junction proteins differentially overexpressed in PD cancer cells compared to RD cancer cells ( $p < 0.0001$ , Figure 3L, Supplemental Table 2). Gap junction proteins (e.g. connexins) are integral membrane proteins that allow for cytosolic exchange of ions, metabolites and secondary messengers between cells (Aasen et al., 2016; Sinyuk et al., 2018). While some have been identified as tumor suppressors, we found that high expression of *GJB2/3/4/5* (Supplemental Figure 3H) was linked to worse survival in the TCGA lung adenocarcinoma RNAseq dataset ( $p = 0.00083$ ) (Figure 3M). These collective findings suggest a pro-tumor effect not only in our cohort but also in NSCLC more generally.

In summary, within cancer cells we identified a rich complexity of clinically-relevant, expressed mutations that may impact therapy response. Furthermore, evaluation of transcriptional profiles of individual tumor cells longitudinally across different treatment timepoints identified several clinically-relevant cell state changes (Supplemental Figure 3I). The therapy-induced cancer cell adaptations highlighted in our study offer new potential biomarkers and therapeutic strategies for patients with advanced-stage NSCLC.

**Longitudinal scRNAseq analysis of an individual patient's tumor during treatment.** Obtaining consecutive clinical tumor biopsies from individual advanced-stage lung cancer patients before and during treatment is challenging. Nevertheless, we obtained samples from the same primary tumor site from 3 treatment timepoints from a patient (TH226) whose tumor contained a standard EGFR exon 19 deletion oncogenic mutation and was treated with the EGFR inhibitor osimertinib (Supplemental Figure 4A-C). In all 3 biopsies, we identified by scRNAseq RNA expression of the EGFR exon 19 driver mutation in the cancer cells and several other mutations of interest (Supplemental Figure 4D).

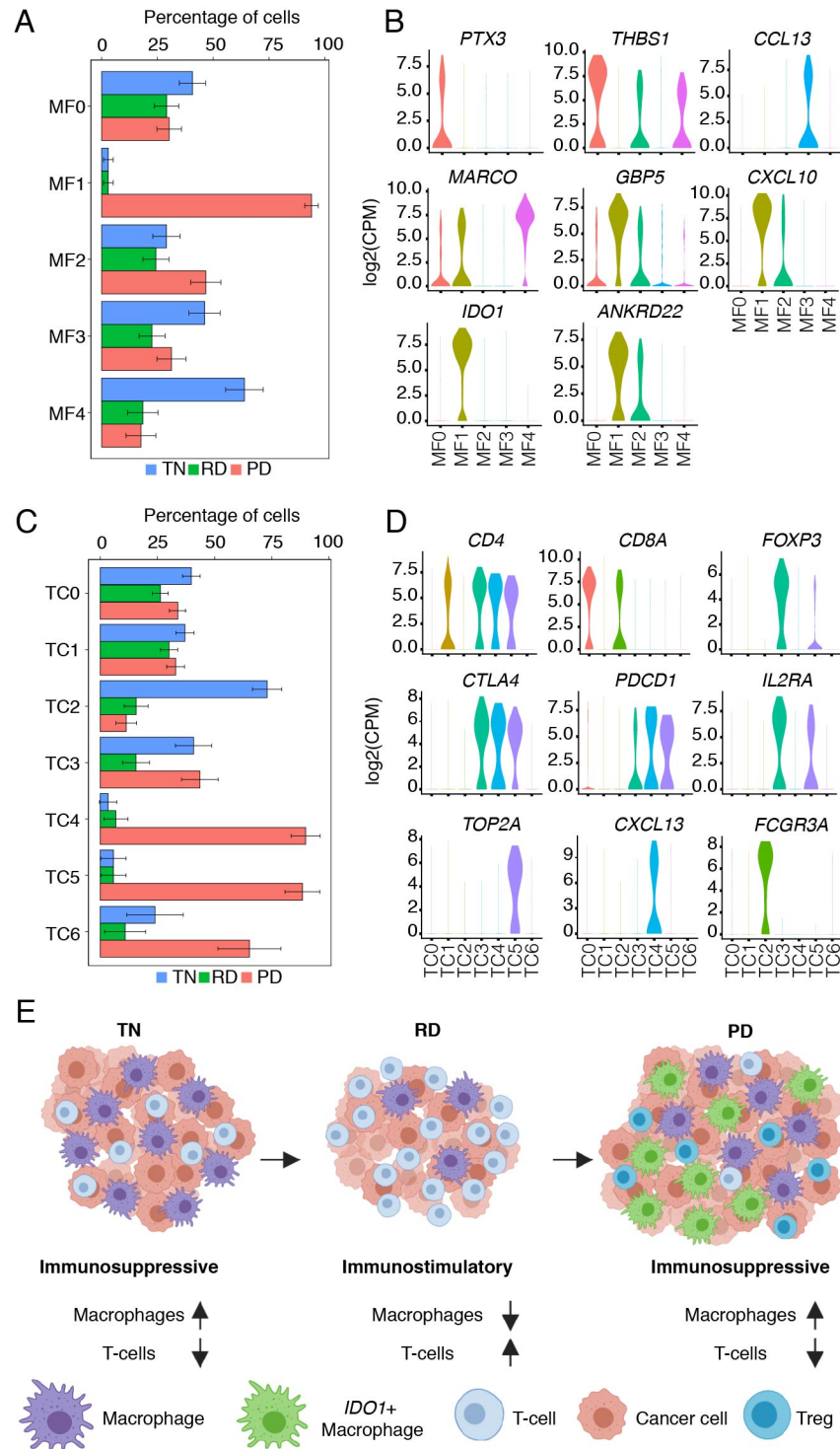
When comparing TH226 to the rest of scRNAseq dataset, we found overlapping differentially expressed genes and signatures (Supplemental Table 4, Supplemental Figure 4E-H). Intriguingly, we also found numerous genes associated with squamous cell differentiation (*KRT16*, *KRT14*, *KRT6A*, *KRT5*, *CLCA2*, *PKPI*, *ANXA8*, *DSG3*) overexpressed at PD compared to TN and RD timepoints ( $p < 0.0001$ , Supplemental Figure 4I, Supplemental Tables 4-5) (Ben-Hamo et al., 2013; Chao et al., 2006; Goodwin et al., 2017). This is particularly interesting given that the patient's lung tumor biopsy at PD demonstrated a histologic shift to squamous cell carcinoma from that of prior biopsies that showed pure adenocarcinoma histology (Supplemental Figure 4C). Histologic transformation to squamous cell carcinoma is a mechanism of EGFR inhibitor resistance in EGFR-mutant NSCLC

(Izumi et al., 2018; Jukna et al., 2016). Thus, scRNAseq has the power to provide a high resolution, gene and pathway level view of biological and histological plasticity that arises during cancer drug treatment.

**Longitudinal inversion of myeloid and lymphoid infiltration within the TME at progressive disease compared to residual disease.** We next addressed the evolution of the TME during targeted treatment. The immune cells ( $n = 12,077$ ) were separately clustered and annotated as primary immune cell types (Figure 4A, Supplemental Table 4). In contrast to clusters of cancer cells, which clustered primarily from a single patient sample (Supplemental Figure 2C), immune cell type clusters were each composed of cells derived from multiple different patients and biopsies (Figure 4B). This is consistent with the expectation of finding common immune cell phenotypes across patients and samples. We compared the immune cell composition across all 3 timepoints, expressed as the correlation between fractional immune cell abundance vectors. The immune composition within RD was the most dissimilar from the other two treatment states ( $r = 0.63$  versus TN samples,  $r = 0.7$  versus PD samples, Pearson's correlation coefficient) (Supplemental Figure 5A). Across all treatment timepoints, T-cells and macrophages were the dominant cell populations and demonstrated an inversion in relative abundance during tumor response and resistance to treatment, a finding we examined further as described below (Figure 4C). T-cells comprised a larger fraction of all immune cells within the TME at RD compared to TN or PD samples (25% T-cells TN, 46% RD, 32% PD). Macrophage infiltration followed the inverse pattern, with a decrease in macrophages at RD compared to TN and PD (34% Macrophages TN, 15% RD, 37% PD).

In 2 patients, we examined the immune cells from matched tumor biopsies obtained at different treatment timepoints (TH226 and TH266, Supplemental Figure 5B and C, respectively). In the 2 tumor biopsies available for patient TH266, both macrophages and T-cells showed identical patterns to those observed across the entire cohort: a reduction in the fraction of macrophages and an increase in the fraction of T-cells from TN to RD (Supplemental Figure 5D). TH226 exhibited a similar pattern with the fraction of macrophages decreasing at RD after initiation of treatment and increasing again at PD (Supplemental Figure 5E). In an orthogonal dataset, we performed immunofluorescence analysis using macrophage and T-cell specific markers (see methods, Figure 4D, Supplemental Figure 5F). These analyses validated our scRNAseq findings and demonstrated consistent macrophage and T-cell population changes during treatment (Figure 4E-F). Additionally, we deconvoluted TCGA bulk transcriptome data for NSCLC into fractions of immune cell types (see Methods) and found that TCGA samples with high fractions of macrophages had significantly worse OS ( $p < 0.01$ ) (Supplemental Figure 5G). This suggests additional clinical relevance of our observations.

These findings are particularly intriguing given their similarity to melanoma tumors treated with PD-1 inhibitor (Riaz et



**Fig. 5. Immune cell subpopulations demonstrate unique transcriptional profiles within each treatment timepoint** (A) Fraction of cells belonging to each treatment stage for each lung-derived macrophage cluster in Supplementary Figure 6. Error bars indicate the 95% confidence interval for the calculated relative frequencies. (B) Violin plots showing the expression level distribution of notable individual genes (C) Fraction of cells belonging to each treatment stage for each lung-derived T-cell cluster in Supplementary Figure 6. Error bars indicate the 95% confidence interval for the calculated relative frequencies. (D) Violin plots showing the expression level distribution of notable individual genes. (E) Graphical summary of immune microenvironment changes across treatment timepoints.

al., 2017), albeit here in the distinct context of oncoprotein-targeted therapy in lung cancer. Specifically, an increase in the number of CD8+ T-cells and NK cells and a decrease in M1 macrophages were observed in melanoma during PD-1 inhibition. There may be common responses in NK/T cells and macrophages during treatment across different tumor histologies and treatments. Hence, conserved approaches to targeting RD across different cancer subtypes and therapeutic modalities may exist, an area for future investigation.

**An *IDO1*-expressing macrophage population is enriched at PD.** Macrophages from lung tumor biopsies (n=1,034) were subset and clustered into 5 distinct groups (Supplemental Figure 6A). We then evaluated the differential gene expression in each resulting cluster (Supplemental Figure 6B, Supplemental Table 4). We calculated the fraction of cells originating from each of the three treatment groups in each of the 5 macrophage clusters (Figure 5A).

Cluster MF0, found equally among all three treatment time-points, was enriched for *THBS1* and *PTX3* (Figure 5B, Supplemental Figure 6B), which are associated with resolution of inflammation, wound healing, and with inhibition of IL-1 $\beta$  (Bouhlef et al., 2007; Faz-López et al., 2016; Martinez and Gordon, 2014; Puig-Kröger et al., 2009; Shiraki et al., 2016; Stein et al., 2016). Macrophage clusters MF3 and MF4 were relatively enriched for TN cells. Both clusters were characterized by expression of different genes associated with an immunosuppressive phenotype (*FOLR2*, *MARCO*, *RETN*, *PPARG*, *CCL13*) (Figure 6B, Supplemental Figure 6B, Supplemental Table 4). Macrophages at PD are overrepresented in group MF1 (Figure 5A), and expressed pro-inflammatory cytokines *CXCL9*, *CXCL10*, and *CXCL11* (Figure 5B, Supplemental Figure 6B), which favor lymphocyte recruitment into the TME (Nagarsheth et al., 2017). Top differentially expressed genes in this population also included the guanylate-binding family proteins *GBP1* and *GBP5*, which are induced in IFN- $\gamma$ -activated macrophages and promote inflammatory signaling within the innate immune system via inflammatory assembly (Shenoy et al., 2012). (Figure 5B, Supplemental Figure 6B).

Despite the expression of pro-inflammatory genes within the MF1 macrophages, the top differentially expressed gene within this group of PD-specific macrophages was *IDO1* (Figure 5B). *IDO1* is induced by inflammation within the TME and promotes a tolerogenic environment through immunosuppressive myeloid cell populations, regulatory T cell (Treg) differentiation, and an immunosuppressive cytokine milieu (Munn and Mellor, 2016).

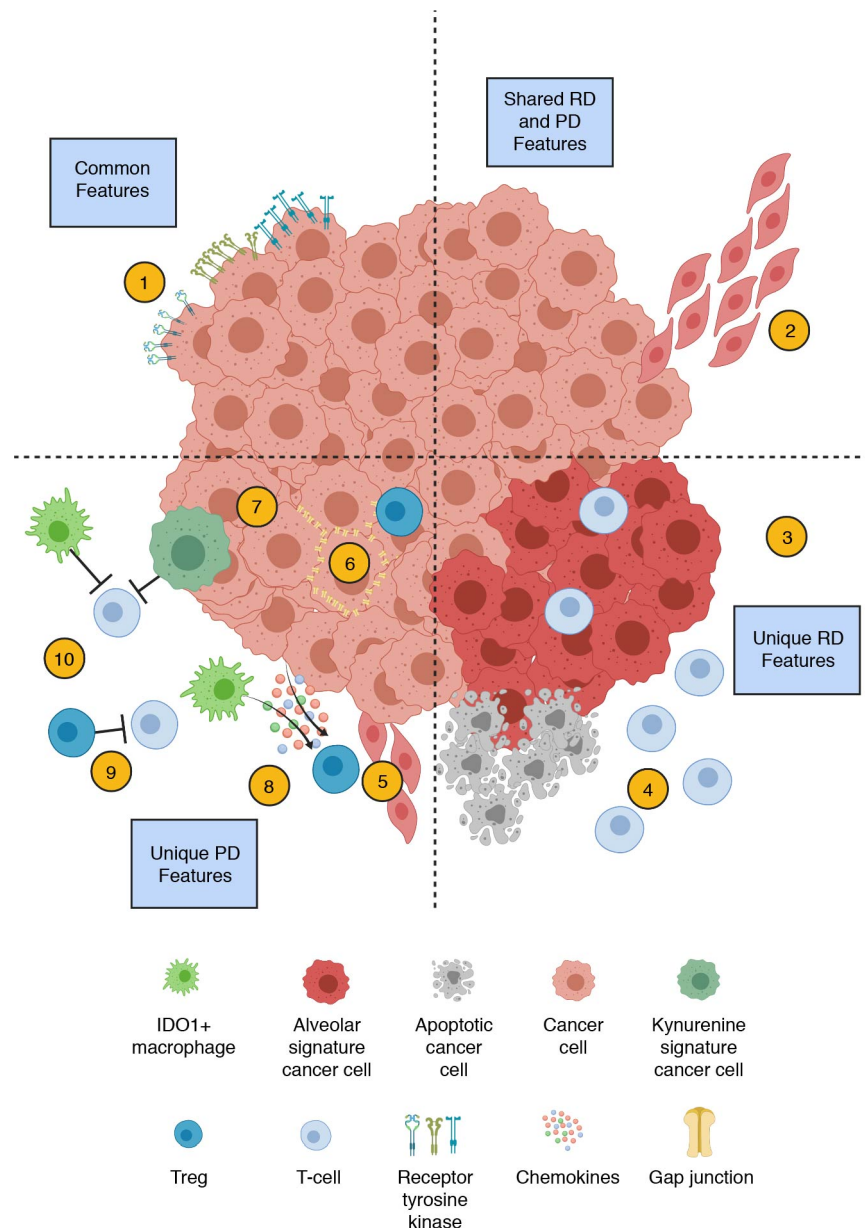
**An immunosuppressive T-cell phenotype is predominant within the TME at PD.** T-cells and NK cells (n=1,725) were analyzed in the same manner as macrophages and resulted in 7 distinct T/NK cell populations (Figure 5C). These included one population (TC2) enriched in TN samples, and 3 populations (TC4, TC5, and TC6) enriched at PD (Figure 5C-D, Supplemental Figure 6D). There was a high fraction of T-cells in RD tumors (Figure 4C) and there was no single

T-cell cluster which demonstrated an excess of T-cells in RD (Figure 5C).

Both TN and PD T-cells demonstrated a relative decrease in T-cell infiltration (Figure 4C). TC2 cluster cells, enriched in TN, expressed markers consistent with an NKT cell phenotype, including both T-cell (*CD3*, *CD8*) and NK cell markers (*KIR2DL3*, *FCGR3A*), which were less represented in RD and PD samples (Figure 5D, Supplemental Figure 6D). While overall T-cell infiltrate remained limited at PD (Figure 4C), there was relative enrichment for T-cell phenotypes with immunosuppressive features, including T-cell clusters TC4, TC5, and TC6 (Figure 5C). TC4 was identified as a CD4+ T-cell cluster with an exhausted phenotype, characterized by expression of the inhibitory receptors *PDCD1* and *CTLA4* (Wherry and Kurachi, 2015) (Figure 5D). The T-cells within this cluster also expressed *CXCL13*, which suggests that these cells may have features of PD-1-expressing T-follicular helper cells (Tfh). Tfh cells function in crosstalk with the humoral immune system to promote B-cell activation and immunologic memory (Crotty, 2014). TC5 was composed of proliferating Treg cells (expressing *FOXP3*, *IL2RA*, *Ki67*, *TOP2A*). Finally, TC6 is composed of NK cells (*KIR2DLA*, *GNLY*), the majority of which did not highly express CD3 or *FCGR3A* (CD16) (Figure 5D). This suggests limited cytotoxic function and is consistent with reports of infiltration by poorly cytotoxic NK cells within NSCLC more generally (Carrega et al., 2008), albeit without a prior link to tumors progressing on targeted therapy.

Tumor biopsies obtained at RD revealed the presence of a more pro-inflammatory, “hot”, TME which was absent in TN or PD biopsy samples as manifested by increased overall proportion of T-cells and enrichment of effector T-cell phenotypes (Figure 5E). The majority of the RD T-cells were contained in TC0 and TC1. Cluster TC0 was composed of cytotoxic CD8+ T-cells (*CD8*, *IFNG*, *GZMK*), while cluster TC1 was composed of CD4+ T-cells that are negative for co-inhibitory receptor expression (*CD4*, *CCR7*, *IL7R*) (Figure 5D, Supplemental Figure 6D), in contrast to the *PDCD1/CTLA4* expressing CD4+ T-cells seen in the PD samples.

In summary, both the TN and PD TME were characterized by the relative predominance of macrophage infiltration over T-cell infiltration; however, the phenotypic characteristics of these infiltrating immune cells differ between the two groups. At PD, there was infiltration by an *IDO1*+ macrophage population, of proliferating regulatory T-cells, and of exhausted CD4+ T-cells which were minimally present at earlier phases of treatment. In contrast, the TN state was characterized by a predominance of more classically immunosuppressive M2-like macrophages (Sica et al., 2008). By distinction, in RD there was increased infiltration of effector T-cell populations with signatures of activity and decreased immunosuppressive macrophage infiltration (Figure 5E).



**Fig. 6. scRNAseq profiles reveal clinical-state specific features of the tumor cellular ecosystem** Features of cancer progression and survival. Common features, found at all treatment timepoints, are shown in the upper left quadrant and include the presence of multiple oncogenic drivers (1). Features shared in RD and PD are shown in the upper right quadrant and include various invasive signaling pathways (2). Features unique at RD, shown in the lower right quadrant, include the Alveolar signature (3) and increased T-cell fraction (4). Features unique to PD, shown in the lower left quadrant, include upregulation of the plasminogen activation pathway (5), expression of gap junction proteins (6), loss of tumor suppressor genes (7), expression of pro-inflammatory chemokines (8), increased Treg population (9), and increased kynurenine signature expression (10).

## Discussion

Though emerging, there remains an incomplete catalogue of single-cell transcriptional data to understand the cell states and therapy-induced evolution of biological heterogeneity of major diseases such as cancer. Precision medicine treatments such as small molecule targeted therapies and immunotherapies have improved cancer patient survival. However, tumors continuously evolve and resistance to these therapies is nearly inevitable. Tumor heterogeneity impacts treatment response and resistance and pertains to cancer cells and the

immune and stromal components of the TME (Andor et al., 2016; Morris et al., 2016; Rybinski and Yun, 2016; Swanton, 2012). Cellular plasticity, as a contributor to heterogeneity, has been under-explored in longitudinal clinical tumor biopsies of advanced-stage solid malignancies and is likely to be a key contributor to tumor resilience to systemic therapy and the development of treatment resistance. Single-cell transcriptional data are invaluable in our attempt to understand the cell states and therapy-induced evolution of biological heterogeneity in cancer, particularly in metastatic disease.

Our scRNAseq analyses of advanced-stage NSCLC biopsies obtained longitudinally from individual patients and the clinical cohort as a whole elucidate the rich mutational and transcriptional diversity within individual tumor samples and the dynamic changes in the transcriptional profiles of cancer cells and the TME composition during treatment. These features are clinically-relevant in NSCLC, as we link them to distinct treatment timepoints and clinical states and to differential survival outcomes in independent clinical cohorts such as the MSK-IMPACT and TCGA datasets. Our findings provide a roadmap that highlights the underlying cellular ecosystem and mechanisms to further explore to improve response to treatment.

An important first step was to develop and optimize a clinical pipeline to capture tumor biopsy samples from a treatment phase that is infrequently captured in solid malignancies: namely the phase of RD during initial molecular treatment. Our study offers a rare view of the clinically-relevant biological processes that characterize this poorly-understood phase of the evolution of advanced-stage solid malignancies (here, lung cancer) and creates a resource for both discovery purposes and validation of pre-clinical studies (Blakely et al., 2015; Ramirez et al., 2016; Sharma et al., 2010; Spitzer et al., 2017). We demonstrate successful use of these tumor samples in combination with plate-based scRNAseq in metastatic NSCLC samples. This is critical, as patients with metastatic disease do not routinely receive surgical resection as part of their treatment. Thus, techniques for single-cell profiling that require larger amounts of tissue are not suitable for the interrogation of tissue samples from metastatic disease (Lambrechts et al., 2018; Schelker et al., 2017).

Our scRNAseq data was leveraged to query for the presence of cancer-relevant mutations within expressed regions of the genome. This revealed widespread intra-tumoral heterogeneity in oncogenic alterations that are expressed in cancer cells.

The scRNAseq data complement and extend current clinical analyses which typically: (1) profile a limited subset of genes, (2) provide a view of DNA changes without validation of gene expression, and (3) are performed at bulk cellular scale (Figure 2). Our data suggest that individual cancer cell populations can show expression of the putative oncogenic driver and multiple additional mutations in genes of known oncogenic importance (Figure 6, #1). This is relevant as it provides a potential explanation for why complete responses to treatment are rare. Cancer cells, and their occult genetic subpopulations as revealed by high-resolution scRNAseq, in human tumors already harbor the appropriate genetic framework and evolutionary playbook to evolve resistance. These ‘hard-wired’ properties that can remain undetected by current bulk sampling analysis are further bolstered by the therapy-induced transcriptional plasticity that we demonstrated by longitudinal scRNAseq profiling.

We uncovered transcriptional signatures specific to different treatment time points and clinical states (Figure 3, Figure 6 #3, #5, #6, #7). The majority of these signatures (kynurenine pathway, plasminogen activation pathway, *SERPINE1*, and gap junction genes) were biomarkers of significantly worse overall survival in TCGA lung adenocarcinoma samples and were most pronounced at PD. Conversely, we found the alveolar cell signature was enriched at RD and was associated with improved survival. Our data highlight a connection from the alveolar cell signature to the WNT/  $\beta$ -catenin pathway as a mechanism of injury-response regeneration. Though the WNT/  $\beta$ -catenin pathway is potentially therapeutically-targetable, (Krishnamurthy and Kurzrock, 2018) it will be critical to determine how to best modulate this pathway to impact residual cancer cell survival. A general principle our data highlight is that by employing targeted treatments that take advantage of specific cell states we may be able to engineer cancer cell fate(s) to improve therapeutic responses in

ID	Feature	Therapeutic approach	References
1	Multiple targetable mutations	Combination targeted therapy	(McCoach and Bivona, 2019)
2	Invasion pathways	Targeted inhibition	(Rahman et al., 2017b; Zhang et al., 2016, 2019b)
3	Increased immunostimulatory T-cells	Immune system modulation	(Ha et al., 2019; Jenkins et al., 2018; Souza-Fonseca-Guimaraes et al., 2019; Valkenburg et al., 2018)
4	Alveolar signature	Targeted inhibition	(Nabhan et al., 2018c; Zhang et al., 2019b)
5	Plasminogen activation signature	Targeted inhibition	(Mahmood et al., 2018; Zhang et al., 2019b)
5	SERPINE1 signature	Targeted inhibition and immune modulation	(Placencio and DeClerck, 2015)
6	Gap Junction signature	Targeted inhibition	(Wu and Wang, 2019; Mulkearns-Hubert et al., 2019)
7	Loss of Tumor suppressors	Targeting acquired vulnerabilities	(Ding et al., 2019)
8	Increased pro-inflammatory chemokines	Immune system modulation	(Tokunaga et al., 2018)
9	Increased Tregs	Immune system modulation	(Tanaka and Sakaguchi, 2019)
10	Kyurenine signature	Targeted inhibition and immune modulation	(Labadie et al., 2019)

**Table 1.** Table of common and unique features in different treatment timepoints and possible therapeutic approaches

metastatic solid malignancies, expanding upon this concept in practice in acute promyelocytic leukemia (Chomienne et al., 1990). If deployed at the appropriate time, treatments that target liabilities of a specific cell state or prevent further adaptation may help improve patient survival by constraining continued tumor evolution towards complete drug resistance. It is possible that inhibition of the factors driving the primitive cell state transition at RD could remove this protective biological state and eliminate RD cancer cells. It remains unclear whether the cell state(s) present at RD is unique to cancer cells receiving targeted therapy or also applies to chemotherapy and immunotherapy, an area for future investigation. We have highlighted ongoing efforts targeting signatures we discovered in Table 1 that could be explored to improve targeted therapy response.

Beyond cancer cell-intrinsic signatures, we also explored changes in the TME during treatment. We found a relatively low T-cell infiltration in the TME of TN and PD patients (Figure 5C-D), consistent with prior reports of low cytotoxic T-cell infiltration in treatment-naïve EGFR-mutant NSCLC (Gainor et al., 2016). Our results uncovered an induction of a more inflammatory phenotype during RD on targeted therapy, hallmarked by infiltration of cytotoxic T-cells (Figure 6, #4) and decreased infiltration of immunosuppressive macrophages (Figure 5E). This inflammatory state may represent a complement to the alveolar cell, injury-repair and regenerative state present in the cancer cell compartment (described above), with the potential for crosstalk between the cancer cells and TME. These TME changes were transient, as at PD there was enrichment for IDO1-expressing macrophages, regulatory T-cells, and other immunosuppressive T-cell populations. These are all features of an environment hostile to the establishment of an effective immunologic response (Figure 6, #9, #10). The induction of a more immunostimulatory phenotype during targeted therapy (i.e. in RD) may offer a window-of-opportunity to introduce novel TME target-based combination therapies around the time of RD in the context of a more favorable TME to increase initial response and consolidate treatment in a multi-modal approach.

Cancer cell signaling and the TME are linked and there may be treatment strategies which target both compartments concurrently. We identified two such examples: the kynurenine pathway and *SERPINE1*. In the kynurenine pathway, we identified increased pathway activation in cancer cells and myeloid cells at PD (Figure 6, #5). *IDO1*, as a rate limiting enzyme in the kynurenine pathway, can influence diverse components of the TME including T-cell and myeloid cell populations as well as angiogenesis in favor of immunosuppression (Munn and Mellor, 2016). The use of *IDO1* inhibitors as part of a combination immunotherapy strategy with PD1/PDL1 checkpoint inhibitors showed promise in early-phase studies (Siefker-Radtke et al., 2018), yet ultimately failed to demonstrate improved outcomes in advanced-stage melanoma (Long et al., 2018). We demonstrated distinct evolving TME states, suggesting that there may be a window-of-opportunity at which point kynurenine

pathway inhibitors may be more effective (Figure 6, Table 1). Similarly, *SERPINE1* is notable for its activity in both cancer cells, as we identified, and the immune system (Placencio and DeClerck, 2015). Utilizing therapies that leverage this multi-cellular crosstalk may improve patient outcomes.

In summary, the scRNAseq dataset presented here demonstrates the feasibility of performing scRNAseq on tumor biopsies obtained longitudinally at clinically-relevant time-points during the active targeted treatment of advanced-stage solid malignancy patients. Understanding each of the features we discovered and highlighted (Figure 6) in individual cells, as revealed by scRNAseq, is critical to deconvolute tumor biological heterogeneity and evolution during systemic therapy. The data provide a more granular biological foundation to develop strategies for the elimination or neutralization of RD to induce more durable responses for patients with advanced-stage NSCLC and potentially other solid malignancies across different therapeutic modalities.

#### ACKNOWLEDGEMENTS

This project is supported by the NIH / NCI U54CA224081, R01CA204302, R01CA211052 and R01CA169338 (to T.G.B.), the Van Auken Foundation (to T.G.B. and C.M.B.), and Novartis Pharmaceuticals (to T.G.B.), AstraZeneca, as well as grants from the University of California Cancer League (to C.E.M.), The Damon Runyon Cancer Research Foundation P0528804 (to C.M.B.), Doris Duke Charitable Foundation grant # 2018110 (to C.M.B.), and V Foundation V2018-020 (to C.M.B.), K12 CA086913 (to E.S.). F.H. was supported by the Mildred Scheel postdoctoral fellowship from the German Cancer Aid. E.A.Y. is supported by T32 HL007185 from the NHLBI. E.L.S. is supported by K12 CA086913. Special thanks to Bing Wu and Lillian Cohn for their insights and support.

#### References

- Aasen, T., Mesnil, M., Naus, C.C., Lampe, P.D., and Laird, D.W. (2016). Gap junctions and cancer: communicating for 50 years. *Nat. Rev. Cancer* 16, 775–788.
- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
- Andor, N., Graham, T.A., Jansen, M., Xia, L.C., Aktipis, C.A., Petritsch, C., Ji, H.P., and Maley, C.C. (2016). Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* 22, 105–113.
- Arasada, R.R., Shilo, K., Yamada, T., Zhang, J., Yano, S., Ghanem, R., Wang, W., Takeuchi, S., Fukuda, K., Katakami, N., et al. (2018a). Notch3-dependent  $\beta$ -catenin signaling mediates EGFR TKI drug persistence in EGFR mutant NSCLC. *Nat. Commun.* 9, 3198.
- Blakely, C.M., Pazarentzos, E., Olivas, V., Asthana, S., Yan, J.J., Tan, I., Hrutanovic, G., Chan, E., Lin, L., Neel, D.S., et al. (2015). NF- $\kappa$ B-activating complex engaged in response to EGFR oncogene inhibition drives tumor cell survival and residual disease in lung cancer. *Cell Rep.* 11, 98–110.
- Blakely, C.M., Watkins, T.B.K., Wu, W., Gini, B., Chabon, J.J., McCoach, C.E., McGranahan, N., Wilson, G.A., Birkbak, N.J., Olivas, V.R., et al. (2017b). Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. *Nat. Genet.* 49, 1693–1704.
- Brannon, A.R., Vakiani, E., Sylvester, B.E., Scott, S.N., McDermott, G., Shah, R.H., Kania, K., Viale, A., Oswald,

- D.M., Vacic, V., et al. (2014). Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome Biol.* 15, 454.
- Camidge, D.R., Pao, W., and Sequist, L.V. (2014). Acquired resistance to TKIs in solid tumours: learning from lung cancer. *Nat. Rev. Clin. Oncol.* 11, 473–481.
- Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120.
- Carrega, P., Morandi, B., Costa, R., Frumento, G., Forte, G., Altavilla, G., Ratto, G.B., Mingari, M.C., Moretta, L., and Ferlazzo, G. (2008). Natural killer cells infiltrating human nonsmall-cell lung cancer are enriched in CD56 bright CD16(-) cells and display an impaired capability to kill tumor cells. *Cancer* 112, 863–875.
- Casás-Selves, M., Kim, J., Zhang, Z., Helfrich, B.A., Gao, D., Porter, C.C., Scarborough, H.A., Bunn, P.A., Chan, D.C., Tan, A.C., et al. (2012). Tankyrase and the canonical Wnt pathway protect lung cancer cells from EGFR inhibition. *Cancer Res.* 72, 4154–4164.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsen, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404.
- Cheong, J.E., and Sun, L. (2018). Targeting the IDO1/TDO2–KYN–AhR Pathway for Cancer Immunotherapy – Challenges and Opportunities. *Trends Pharmacol. Sci.* 39, 307–325.
- Chomienne, C., Ballerini, P., Balitrand, N., Daniel, M.T., Fenaux, P., Castaigne, S., and Degos, L. (1990). All-trans retinoic acid in acute promyelocytic leukemias. II. In vitro studies: structure-function relationship. *Blood* 76, 1710–1717.
- Chung, W., Eum, H.H., Lee, H., Lee, K., Lee, H., Kim, K., Ryu, H.S., Kim, S., Lee, J.E., Park, Y.H., et al. (2017b). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.* 8, 15081.
- Crotty, S. (2014). T follicular helper cell differentiation, function, and roles in disease. *Immunity* 41, 529–542.
- Darmanis, S., Sloan, S.A., Croote, D., Mignardi, M., Chernikova, S., Samghabadi, P., Zhang, Y., Neff, N., Kowarsky, M., Caneda, C., et al. (2017). Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep.* 21, 1399–1410.
- Desai, T.J., Brownfield, D.G., and Krasnow, M.A. (2014). Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* 507, 190–194.
- Ding, H., Zhao, J., Zhang, Y., Yu, J., Liu, M., Li, X., Xu, L., Lin, M., Liu, C., He, Z., et al. (2019). Systematic Analysis of Drug Vulnerabilities Conferred by Tumor Suppressor Loss. *Cell Rep.* 27, 3331–3344.e6.
- Doebele, R.C., Pilling, A.B., Aisner, D.L., Kutateladze, T.G., Le, A.T., Weickhardt, A.J., Kondo, K.L., Linderman, D.J., Heasley, L.E., Franklin, W.A., et al. (2012). Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 18, 1472–1482.
- Enge, M., Arda, H.E., Mignardi, M., Beusang, J., Bottino, R., Kim, S.K., and Quake, S.R. (2017). Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell* 171, 321–330.e14.
- Fedchenko, N., and Reifenrath, J. (2014). Different approaches for interpretation and reporting of immunohistochemistry analysis results in the bone tissue - a review. *Diagn. Pathol.* 9, 221.
- Flaherty, K.T., Infante, J.R., Daud, A., Gonzalez, R., Kefford, R.F., Sosman, J., Hamid, O., Schuchter, L., Cebon, J., Ibrahim, N., et al. (2012). Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations. *N. Engl. J. Med.* 367, 1694–1703.
- Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 45, D777–D783.
- Gainor, J.F., Shaw, A.T., Sequist, L.V., Fu, X., Azzoli, C.G., Piotrowska, Z., Huynh, T.G., Zhao, L., Fulton, L., Schultz, K.R., et al. (2016). EGFR Mutations and ALK Rearrangements Are Associated with Low Response Rates to PD-1 Pathway Blockade in Non-Small Cell Lung Cancer: A Retrospective Analysis. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 22, 4585–4593.
- Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6, p11.
- Ge, J., Dong, H., Yang, Y., Liu, B., Zheng, M., Cheng, Q., Peng, L., and Li, J. (2018). NFIX downregulation independently predicts poor prognosis in lung adenocarcinoma, but not in squamous cell carcinoma. *Future Oncol.* 14, 3135–3144.
- Gerlinger, M., Rowan, A.J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* 366, 883–892.
- Ha, D., Tanaka, A., Kibayashi, T., Tanemura, A., Sugiyama, D., Wing, J.B., Lim, E.L., Teng, K.W.W., Adeegbe, D., Newell, E.W., et al. (2019). Differential control of human Treg and effector T cells in tumor immunity by Fc-engineered anti-CTLA-4 antibody. *Proc. Natl. Acad. Sci. U. S. A.* 116, 609–618.
- Haderk, F., Olivas, V., and Bivona, T.G. (2019). Immunohistochemistry to Study YAP in Human Tissue Samples. *Methods Mol. Biol. Clifton NJ* 1893, 89–95.
- Hanna, J.M., and Onaitis, M.W. (2013). Cell of origin of lung cancer. *J. Carcinog.* 12.
- Hata, A.N., Niederst, M.J., Archibald, H.L., Gomez-Caraballo, M., Siddiqui, F.M., Mulvey, H.E., Maruvka, Y.E.,

- Ji, F., Bhang, H.C., Krishnamurthy Radhakrishna, V., et al. (2016). Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat. Med.* 22, 262–269.
- Hornýák, L., Dobos, N., Koncz, G., Karányi, Z., Páll, D., Szabó, Z., Halmos, G., and Székvölgyi, L. (2018). The Role of Indoleamine-2,3-Dioxygenase in Cancer Development, Diagnostics, and Therapy. *Front. Immunol.* 9.
- Hrustanovic, G., and Bivona, T.G. (2015). RAS-MAPK in ALK targeted therapy resistance. *Cell Cycle Georget. Tex* 14, 3661–3662.
- Hsiao, C.J., Tung, P., Blischak, J.D., Burnett, J.E., Barr, K.A., Dey, K.K., Stephens, M., and Gilad, Y. (2019). Characterizing and inferring quantitative cell cycle phase in single-cell RNA-seq data analysis (Genomics).
- Huch, M., Bonfanti, P., Boj, S.F., Sato, T., Loomans, C.J.M., van de Wetering, M., Sojoodi, M., Li, V.S.W., Schuijers, J., Gracanin, A., et al. (2013). Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. *EMBO J.* 32, 2708–2721.
- Izumi, H., Yamasaki, A., Ueda, Y., Sumikawa, T., Maeta, H., Nakamoto, S., and Shimizu, E. (2018). Squamous Cell Carcinoma Transformation from EGFR-mutated Lung Adenocarcinoma: A Case Report and Literature Review. *Clin. Lung Cancer* 19, e63–e66.
- Jenkins, R.W., Barbie, D.A., and Flaherty, K.T. (2018). Mechanisms of resistance to immune checkpoint inhibitors. *Br. J. Cancer* 118, 9–16.
- Juan, J., Muraguchi, T., Iezza, G., Sears, R.C., and McMahon, M. (2014). Diminished WNT → β-catenin → c-MYC signaling is a barrier for malignant progression of BRAFV600E-induced lung tumors. *Genes Dev.* 28, 561–575.
- Jukna, A., Montanari, G., Mengoli, M.C., Cavazza, A., Covi, M., Barbieri, F., Bertolini, F., and Rossi, G. (2016). Squamous Cell Carcinoma “Transformation” Concurrent with Secondary T790M Mutation in Resistant EGFR-Mutated Adenocarcinomas. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* 11, e49-51.
- Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* 11, 740–742.
- Kim, Y., Lee, B., Shim, J.H., Lee, S.-H., Park, W.-Y., Choi, Y.-L., Sun, J.-M., Ahn, J.S., Ahn, M.-J., and Park, K. (2019). Concurrent Genetic Alterations Predict the Progression to Target Therapy in EGFR-Mutated Advanced NSCLC. *J. Thorac. Oncol.* 14, 193–202.
- Krishnamurthy, N., and Kurzrock, R. (2018). Targeting the Wnt/β-catenin pathway in cancer: Update on effectors and inhibitors. *Cancer Treat. Rev.* 62, 50–60.
- Kubala, M.H., Punj, V., Placencio-Hickok, V.R., Fang, H., Fernandez, G.E., Sposto, R., and DeClerck, Y.A. (2018). Plasminogen Activator Inhibitor-1 Promotes the Recruitment and Polarization of Macrophages in Cancer. *Cell Rep.* 25, 2177-2191.e7.
- Labadie, B.W., Bao, R., and Luke, J.J. (2019). Reimagining IDO Pathway Inhibition in Cancer Immunotherapy via Downstream Focus on the Tryptophan-Kynurenine-Aryl Hydrocarbon Axis. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 25, 1462–1471.
- Lambrechts, D., Wauters, E., Boeckx, B., Aibar, S., Nittner, D., Burton, O., Bassez, A., Decaluwé, H., Pircher, A., Van den Eynde, K., et al. (2018b). Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* 24, 1277–1289.
- Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218.
- Lee, S.Y., Haq, F., Kim, D., Jun, C., Jo, H.-J., Ahn, S.-M., and Lee, W.-S. (2014). Comparative genomic analysis of primary and synchronous metastatic colorectal cancers. *PLoS One* 9, e90459.
- Lee, Y.C., Saijo, N., Sasaki, Y., Takahashi, H., Sakurai, M., Ishihara, J., Hoshi, A., Chen, K.M., and Hamburger, A.W. (1985). Clonogenic patterns of human pulmonary adenocarcinoma cell lines (PC-9, PC-13 and PC-14) and how they influence the results of test for chemosensitivity to cisplatin in the human tumor clonogenic assay. *Jpn. J. Clin. Oncol.* 15, 637–644.
- Lefèvre, L., Omeiri, H., Drougat, L., Hantel, C., Giraud, M., Val, P., Rodriguez, S., Perlemoine, K., Blugeon, C., Beuschlein, F., et al. (2015). Combined transcriptome studies identify AFF3 as a mediator of the oncogenic effects of β-catenin in adrenocortical carcinoma. *Oncogenesis* 4, e161.
- Liu, H., Hooper, S.B., Armugam, A., Dawson, N., Ferraro, T., Jeyaseelan, K., Thiel, A., Koukoulas, I., and Wintour, E.M. (2003). Aquaporin gene expression and regulation in the ovine fetal lung. *J. Physiol.* 551, 503–514.
- Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018a). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 173, 400-416.e11.
- Liu, M., Wang, X., Wang, L., Ma, X., Gong, Z., Zhang, S., and Li, Y. (2018b). Targeting the IDO1 pathway in cancer: from bench to bedside. *J. Hematol. Oncol. J Hematol Oncol* 11, 100.
- Mahmood, N., Mihalciou, C., and Rabbani, S.A. (2018). Multifaceted Role of the Urokinase-Type Plasminogen Activator (uPA) and Its Receptor (uPAR): Diagnostic, Prognostic, and Therapeutic Applications. *Front. Oncol.* 8, 24.
- McCoach, C.E., and Bivona, T.G. (2019). Engineering Multi-dimensional Evolutionary Forces to Combat Cancer. *Cancer Discov.* 9, 587–604.
- McCoach, C.E., Blakely, C.M., Banks, K.C., Levy, B., Chue, B.M., Raymond, V.M., Le, A.T., Lee, C.E., Diaz, J., Waqar, S.N., et al. (2018). Clinical Utility of Cell-Free DNA for the Detection of ALK Fusions and Genomic Mechanisms of ALK Inhibitor Resistance in Non-Small Cell Lung Cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 24, 2758–2770.
- Mok, T.S., Wu, Y.-L., Thongprasert, S., Yang, C.-H., Chu, D.-T., Saijo, N., Sunpaweravong, P., Han, B., Margono, B., Ichinose, Y., et al. (2009). Gefitinib or carboplatin-paclitaxel



- in pulmonary adenocarcinoma. *N. Engl. J. Med.* 361, 947–957.
- Morris, L.G.T., Riaz, N., Desrichard, A., Şenbabaoğlu, Y., Hakimi, A.A., Makarov, V., Reis-Filho, J.S., and Chan, T.A. (2016). Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget* 7, 10051–10063.
- Munn, D.H., and Mellor, A.L. (2016). IDO in the Tumor Microenvironment: Inflammation, Counter-Regulation, and Tolerance. *Trends Immunol.* 37, 193–207.
- Nabhan, A.N., Brownfield, D.G., Harbury, P.B., Krasnow, M.A., and Desai, T.J. (2018a). Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* 359, 1118–1123.
- Nagarsheth, N., Wicha, M.S., and Zou, W. (2017). Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nat. Rev. Immunol.* 17, 559–572.
- Nakayama, S., Sng, N., Carretero, J., Welner, R., Hayashi, Y., Yamamoto, M., Tan, A.J., Yamaguchi, N., Yasuda, H., Li, D., et al. (2014).  $\beta$ -catenin contributes to lung tumor development induced by EGFR mutations. *Cancer Res.* 74, 5891–5902.
- Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457.
- Pacheco-Pinedo, E.C., Durham, A.C., Stewart, K.M., Goss, A.M., Lu, M.M., Demayo, F.J., and Morrissey, E.E. (2011). Wnt/ $\beta$ -catenin signaling accelerates mouse lung tumorigenesis by imposing an embryonic distal progenitor phenotype on lung epithelium. *J. Clin. Invest.* 121, 1935–1945.
- Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014a). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401.
- Placencio, V.R., and DeClerck, Y.A. (2015). Plasminogen Activator Inhibitor-1 in Cancer: Rationale and Insight for Future Therapeutic Testing. *Cancer Res.* 75, 2969–2974.
- Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A., Emerick, K.S., et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* 171, 1611–1624.e24.
- Rahman, N.I.A., Abdul Murad, N.A., Mollah, M.M., Jamal, R., and Harun, R. (2017a). NFIX as a Master Regulator for Lung Cancer Progression. *Front. Pharmacol.* 8, 540.
- Rahman, N.I.A., Abdul Murad, N.A., Mollah, M.M., Jamal, R., and Harun, R. (2017b). NFIX as a Master Regulator for Lung Cancer Progression. *Front. Pharmacol.* 8, 540.
- Ramirez, M., Rajaram, S., Steininger, R.J., Osipchuk, D., Roth, M.A., Morinishi, L.S., Evans, L., Ji, W., Hsu, C.-H., Thurley, K., et al. (2016). Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat. Commun.* 7, 10690.
- Riaz, N., Havel, J.J., Makarov, V., Desrichard, A., Urba, W.J., Sims, J.S., Hodi, F.S., Martín-Algarra, S., Mandal, R., Sharfman, W.H., et al. (2017). Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* 171, 934–949.e16.
- Ricciuti, B., Leonardi, G.C., Puccetti, P., Fallarino, F., Bianconi, V., Sahebkar, A., Baglivo, S., Chiari, R., and Pirro, M. (2019). Targeting indoleamine-2,3-dioxygenase in cancer: Scientific rationale and clinical evidence. *Pharmacol. Ther.* 196, 105–116.
- Rotow, J., and Bivona, T.G. (2017). Understanding and targeting resistance mechanisms in NSCLC. *Nat. Rev. Cancer* 17, 637–658.
- Rybicki, B., and Yun, K. (2016). Addressing intra-tumoral heterogeneity and therapy resistance. *Oncotarget* 7, 72322–72342.
- Scheffler, M., Ihle, M.A., Hein, R., Merkelbach-Bruse, S., Scheel, A.H., Siemanowski, J., Brägelmann, J., Kron, A., Abedpour, N., Ueckerth, F., et al. (2019). K-ras Mutation Subtypes in NSCLC and Associated Co-occurring Mutations in Other Oncogenic Pathways. *J. Thorac. Oncol.* 14, 606–616.
- Schelker, M., Feau, S., Du, J., Ranu, N., Klipp, E., MacBeath, G., Schoeberl, B., and Raue, A. (2017). Estimation of immune cell content in tumour tissue using single-cell RNA-seq data. *Nat. Commun.* 8, 2032.
- Schiller, H.B., Montoro, D.T., Simon, L.M., Rawlins, E.L., Meyer, K.B., Strunz, M., Vieira Braga, F.A., Timens, W., Koppelman, G.H., Budinger, G.R.S., et al. (2019). The Human Lung Cell Atlas: A High-Resolution Reference Map of the Human Lung in Health and Disease. *Am. J. Respir. Cell Mol. Biol.* 61, 31–41.
- Serveaux-Dancer, M., Jabaudon, M., Creveaux, I., Belville, C., Blondonnet, R., Gross, C., Constantin, J.-M., Blanchon, L., and Sapin, V. (2019). Pathological Implications of Receptor for Advanced Glycation End-Product (AGER) Gene Polymorphism. *Dis. Markers* 2019, 2067353.
- Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* 141, 69–80.
- Shaw, A.T., and Engelman, J.A. (2013). ALK in lung cancer: past, present, and future. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 31, 1105–1111.
- Shaw, A.T., Kim, D.-W., Nakagawa, K., Seto, T., Crinó, L., Ahn, M.-J., De Pas, T., Besse, B., Solomon, B.J., Blackhall, F., et al. (2013). Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N. Engl. J. Med.* 368, 2385–2394.
- Shenoy, A.R., Wellington, D.A., Kumar, P., Kassa, H., Booth, C.J., Cresswell, P., and MacMicking, J.D. (2012). GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. *Science* 336, 481–485.
- Shihab, H.A., Rogers, M.F., Gough, J., Mort, M., Cooper, D.N., Day, I.N.M., Gaunt, T.R., and Campbell, C. (2015). An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* 31, 1536–1543.
- Sica, A., Larghi, P., Mancino, A., Rubino, L., Porta, C., To-

- taro, M.G., Rimoldi, M., Biswas, S.K., Allavena, P., and Mantovani, A. (2008). Macrophage polarization in tumour progression. *Semin. Cancer Biol.* 18, 349–355.
- Siefker-Radtke, A.O., Necchi, A., Park, S.H., Garca-Donas, J., Huddart, R.A., Burgess, E.F., Fleming, M.T., Rezazadeh, A., Mellado, B., Varlamov, S., et al. (2018). First results from the primary analysis population of the phase 2 study of erdafitinib (ERDA; JNJ-42756493) in patients (pts) with metastatic or unresectable urothelial carcinoma (mUC) and FGFR alterations (FGFRalt). *J. Clin. Oncol.* 36, 4503–4503.
- Siegel, R.L., Miller, K.D., and Jemal, A. (2018). Cancer statistics, 2018. *CA. Cancer J. Clin.* 68, 7–30.
- Sinyuk, M., Mulkearns-Hubert, E.E., Reizes, O., and Lathia, J. (2018). Cancer Connectors: Connexins, Gap Junctions, and Communication. *Front. Oncol.* 8, 646.
- Souza-Fonseca-Guimaraes, F., Cursons, J., and Huntington, N.D. (2019). The Emergence of Natural Killer Cells as a Major Target in Cancer Immunotherapy. *Trends Immunol.* 40, 142–158.
- Spitzer, M.H., Carmi, Y., Reticker-Flynn, N.E., Kwek, S.S., Madhireddy, D., Martins, M.M., Gherardini, P.F., Prestwood, T.R., Chabon, J., Bendall, S.C., et al. (2017). Systemic Immunity Is Required for Effective Cancer Immunotherapy. *Cell* 168, 487–502.e15.
- Stewart, D.J. (2014). Wnt signaling pathway in non-small cell lung cancer. *J. Natl. Cancer Inst.* 106, djt356.
- Swanton, C. (2012). Intratumor heterogeneity: evolution through space and time. *Cancer Res.* 72, 4875–4882.
- Swanton, C., and Govindan, R. (2016). Clinical Implications of Genomic Discoveries in Lung Cancer. *N. Engl. J. Med.* 374, 1864–1873.
- Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, and Principal investigators (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 562, 367–372.
- Tammela, T., Sanchez-Rivera, F.J., Cetinbas, N.M., Wu, K., Joshi, N.S., Helenius, K., Park, Y., Azimi, R., Kerper, N.R., Wesselhoeft, R.A., et al. (2017). A Wnt-producing niche drives proliferative potential and progression in lung adenocarcinoma. *Nature* 545, 355–359.
- Tanaka, A., and Sakaguchi, S. (2019). Targeting Treg cells in cancer immunotherapy. *Eur. J. Immunol.* 49, 1140–1146.
- Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016a). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196.
- Tokunaga, R., Zhang, W., Naseem, M., Puccini, A., Berger, M.D., Soni, S., McSkane, M., Baba, H., and Lenz, H.-J. (2018). CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - A target for novel cancer therapy. *Cancer Treat. Rev.* 63, 40–47.
- Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014a). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509, 371–375.
- Triplett, T.A., Garrison, K.C., Marshall, N., Donkor, M., Blazek, J., Lamb, C., Qerqez, A., Dekker, J.D., Tanno, Y., Lu, W.-C., et al. (2018). Reversal of indoleamine 2,3-dioxygenase-mediated cancer immune suppression by systemic kynurenine depletion with a therapeutic enzyme. *Nat. Biotechnol.* 36, 758–764.
- Umeda, S., Kanda, M., Miwa, T., Tanaka, H., Tanaka, C., Kobayashi, D., Suenaga, M., Hattori, N., Hayashi, M., Yamada, S., et al. (2018). Expression of sushi domain containing two reflects the malignant potential of gastric cancer. *Cancer Med.* 7, 5194–5204.
- Valkenburg, K.C., de Groot, A.E., and Pienta, K.J. (2018). Targeting the tumour stroma to improve cancer therapy. *Nat. Rev. Clin. Oncol.* 15, 366–381.
- Venteicher, A.S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M.G., Hovestadt, V., Escalante, L.E., Shaw, M.L., Rodman, C., et al. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* 355.
- Vieira Braga, F.A., Kar, G., Berg, M., Carpaij, O.A., Polanski, K., Simon, L.M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J., et al. (2019). A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* 25, 1153–1163.
- Vignot, S., Lefebvre, C., Frampton, G.M., Meurice, G., Yelensky, R., Palmer, G., Capron, F., Lazar, V., Hannoun, L., Miller, V.A., et al. (2015). Comparative analysis of primary tumour and matched metastases in colorectal cancer patients: evaluation of concordance between genomic and transcriptional profiles. *Eur. J. Cancer Oxf. Engl.* 1990 51, 791–799.
- Wade, K.C., Guttentag, S.H., Gonzales, L.W., Maschhoff, K.L., Gonzales, J., Kolla, V., Singhal, S., and Ballard, P.L. (2006). Gene Induction during Differentiation of Human Pulmonary Type II Cells In Vitro. *Am. J. Respir. Cell Mol. Biol.* 34, 727.
- Wang, J., Xu, R., Yuan, H., and Zhang, Y. (2019). Single-cell RNA sequencing reveals novel gene expression signatures of trastuzumab treatment in HER2+ breast cancer: A pilot study. *Medicine (Baltimore)* 98, e15872.
- Wang, Y., Tang, Z., Huang, H., Li, J., Wang, Z., Yu, Y., Zhang, C., Li, J., Dai, H., Wang, F., et al. (2018). Pulmonary alveolar type I cell population consists of two distinct subtypes that differ in cell fate. *Proc. Natl. Acad. Sci.* 115, 2407–2412.
- Wherry, E.J., and Kurachi, M. (2015). Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* 15, 486–499.
- Winslow, M.M., Dayton, T.L., Verhaak, R.G.W., Kim-Kiselak, C., Snyder, E.L., Feldser, D.M., Hubbard, D.D., DuPage, M.J., Whittaker, C.A., Hoersch, S., et al. (2011). Suppression of Lung Adenocarcinoma Progression by Nkx2-1. *Nature* 473, 101–104.
- Xu, Y., Mizuno, T., Sridharan, A., Du, Y., Guo, M., Tang,

- J., Wikenheiser-Brokamp, K.A., Perl, A.-K.T., Funari, V.A., Gokey, J.J., et al. (2016). Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 1, e90558.
- Xu, Y., Miao, C., Jin, C., Qiu, C., Li, Y., Sun, X., Gao, M., Lu, N., and Kong, B. (2018). SUSD2 promotes cancer metastasis and confers cisplatin resistance in high grade serous ovarian cancer. *Exp. Cell Res.* 363, 160–170.
- Xue, Y., Hou, S., Ji, H., and Han, X. (2017). Evolution from genetics to phenotype: reinterpretation of NSCLC plasticity, heterogeneity, and drug resistance. *Protein Cell* 8, 178–190.
- Yang, J.-J., Zhang, C., Zhao, J., Dai, P., Lin, G., Han, L., Chen, C., Guan, Y.-F., Yang, L., Yi, X., et al. (2019). ALK resistance mutations and co-occurring genetic alterations to the ALK tyrosine kinase inhibitors in lung cancer. *J. Clin. Oncol.* 37, e20675–e20675.
- Yu, H., Shen, H., Zhang, Y., Zhong, F., Liu, Y., Qin, L., and Yang, P. (2014). CAV1 promotes HCC cell progression and metastasis through Wnt/ $\beta$ -catenin pathway. *PloS One* 9, e106451.
- Yuan, S., Norgard, R.J., and Stanger, B.Z. (2019). Cellular Plasticity in Cancer. *Cancer Discov.* 9, 837–851.
- Zehir, A., Benayed, R., Shah, R.H., Syed, A., Middha, S., Kim, H.R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S.M., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* 23, 703–713.
- Zhai, L., Ladomersky, E., Lenzen, A., Nguyen, B., Patel, R., Lauing, K.L., Wu, M., and Wainwright, D.A. (2018). IDO1 in cancer: a Gemini of immune checkpoints. *Cell. Mol. Immunol.* 15, 447–457.
- Zhang, D., Dai, Y., Cai, Y., Suo, T., Liu, H., Wang, Y., Cheng, Z., and Liu, H. (2016). PEBP4 promoted the growth and migration of cancer cells in pancreatic ductal adenocarcinoma. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* 37, 1699–1705.
- Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B., and Kinzler, K.W. (1997). Gene expression profiles in normal and cancer cells. *Science* 276, 1268–1272.
- Zhang, Q., He, Y., Luo, N., Patel, S.J., Han, Y., Gao, R., Modak, M., Carotta, S., Haslinger, C., Kind, D., et al. (2019a). Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* 179, 829–845.e20.
- Zhang, Y., Xie, Z.-Y., Guo, X.-T., Xiao, X.-H., and Xiong, L.-X. (2019b). Notch and breast cancer metastasis: Current knowledge, new sights and targeted therapy. *Oncol. Lett.* 18, 2743–2755.
- Zhou, B., Flodby, P., Luo, J., Castillo, D.R., Liu, Y., Yu, F.-X., McConnell, A., Varghese, B., Li, G., Chimge, N.-O., et al. (2018). Claudin-18-mediated YAP activity regulates lung stem and progenitor cell homeostasis and tumorigenesis. *J. Clin. Invest.* 128, 970–984.
- Zhu, C., Shen, H., Zhu, L., Zhao, F., and Shu, Y. (2017). Plasminogen Activator Inhibitor 1 Promotes Immunosuppression in Human Non-Small Cell Lung Cancers by Enhancing TGF- $\beta$ 1 Expression in Macrophage. *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.* 44, 2201–2211.
- Zhu, X.F., Liu, Z.C., Xie, B.F., Li, Z.M., Feng, G.K., Yang, D., and Zeng, Y.X. (2001). EGFR tyrosine kinase inhibitor AG1478 inhibits cell proliferation and arrests cell cycle in nasopharyngeal carcinoma cells. *Cancer Lett.* 169, 27–32.

## Author Contributions

Conceptualization, T.G.B, C.M.B., R.C.D., J.W. S.D. C.E.M., A.M.; Methodology, T.G.B, C.M.B., R.C.D., S.D. C.E.M., A.M., J.K.R.; Software Programming, A.M., L.H. S.D., Wei.W.; Validation, F.H., D.L.K., C.E.M, A.M., P.G., E.L.S., E.A.Y., J.K.R.; Formal Analysis, A.M., L.H. S.D., A.Z., W.T., M.T., R.S., K.A.Y., C.E.M., W.W, J.K.R, E.A.Y, D.L.K, F.H.; Resources, C.E.M. C.M.B, J.K.R., A.M., S.D., N.N., T.L., A.U., K.J, P.K.K., E.S., Y.G., D.M.N., N.J.T., A.G., M.Go, H.D., L.T, M.Gu., T.J. J.R.K., D.J., E.L.S., E.A.Y. ; Data Curation Management, A.M, C.E.M.; S.D.; Writing- Original Draft C.E.M., A.M., J.K.R., S.D.; Writing – Revisions and Editing, All Authors; Visualization, A.M., C.E.M., J.K.R. S.D., C.M.B., T.G.B.; Supervision, S.D., C.M.B., T.G.B.; J.W.; R.C.D., R.G., N.N.; Project Administration, S.D., C.M.B., T.G.B; Funding Acquisition, S.D., C.M.B., T.G.B, C.E.M.

## Code

All code can be found on github [czbiohub/scell\\_lung\\_adenocarcinoma](https://github.com/czbiohub/scell_lung_adenocarcinoma) and [czbiohub/cerebra](https://github.com/czbiohub/cerebra).

## Materials and Methods

**Patient population.** All patients gave informed consent for collection of clinical correlates, tissue collection, research testing under Institutional Review Board (IRB)-approved protocols (CC13-6512 and CC17-658, NCT03433469). Formalin-fixed paraffin embedded (FFPE), frozen, and fresh tissue samples were obtained according to the safety standards of the interventional radiologist, pulmonologist, or surgeon. Demographic and clinical history for each patient was obtained from chart review. Days until progression were determined based on imaging studies which demonstrated definitive growth of a known tumor site or new extra-CNS metastatic deposits. Residual disease state was determined by serial imaging demonstrating continued reduction or stability tumor with no evidence of progression. Patient studies were conducted according to the Declaration of Helsinki, the Belmont Report, and the U.S. Common Rule.

**Sample preparation of core and resection samples.** Tissue was first cut sample into small pieces and placed into a 1.5 mL tube (or multiple tubes if necessary). 1.5 mL of collagenase buffer (10mL DMEM (GE Life Sciences, SH30081.01), 0.20 g Collagenase Type 2 (Worthington Biochemical, LS004176)) was added to the tube and the sample was digested for 30 minutes at 37°C, shaking in a thermomixer @ 800-1000 rpm. The sample was manually agitated by pipetting up and down 15 times then returned to the thermomixer for 25 minutes. After incubation, the sample was removed from the thermomixer, agitated again by pipetting the sample up and down 15 times before passing the sample through a 100-micron filter (Fisherbrand, 22363548) into a new 15 mL falcon tube. The filter was washed with 1-2 mL of collagenase buffer before the sample was spun in the centrifuge at 500xg for 10 minutes. If the resulting cell pellet was red, 0.5 mL RBC lysis buffer (Thermo Fisher Scientific, A1049201) was added to sample tubes and allowed to sit at room temperature for 3 minutes before quenching with 1.0 mL DMEM (GE Life Sciences, SH30081.01) + 6% FBS (Omega Scientific, Inc, FB-11) and spun in the centrifuge at 500xg for 5 minutes. Remaining cells were stained with 10 µl CD45-FITC (Miltenyi Biotec, 130-080-202) and 1 µl of Hoechst stain (Thermo Fisher Scientific, H3570). Samples incubated on ice in the dark for 20 minutes. One mL of FACS Buffer was then added to the stained cells and spun at 500xg for 10 minutes before aspirating off supernatant. Cells were resuspended with 0.5 mL of FACS Buffer. PI (Life Technologies, P3566) was added immediately prior to sorting (protocols.io: dx.doi.org/10.17504/protocols.io.65rhg56).

**Sample preparation of thoracentesis.** samples Cells were filtered through a 100 µm strainer (Fisherbrand 22363548), pelleted (500xg, 5 min, 4°C), and resuspended in FACS buffer. Cells were then stained with CD45-FITC (Miltenyi Biotec, 130-080-202) for 20 min at 4°C in the dark. Cells were then pelleted (500xg, 5 min, 4°C) and resuspended in FACS buffer before being transferred to a FACS tube (Falcon 14-956-3C). Sytox Blue dead cell stain (Thermo Fisher Scientific, S34867) was added immediately prior to sorting.

**Lysis plate preparation.** Lysis plates were created by dispensing 0.4 µl lysis buffer (0.5U Recombinant RNase Inhibitor (Takara Bio, 2313B), 0.0625% Triton™ X-100 (Sigma, 93443-100ML), 3.125 mM dNTP mix (Thermo Fisher, R0193), 3.125 µM Oligo-dT30VN (IDT, 5'AAGCAGTGGTATCAACGCAGAGTACT30VN-3') and 1:600,000 ERCC RNA spike-in mix (Thermo Fisher, 4456740)) into 384-well hard-shell PCR plates (Biorad HSP3901) using a Tempest liquid handler (Formulatrix). All plates were then spun down for 1 minute at 3220xg and snap frozen on dry ice. Plates were stored at -80°C until used for sorting.

**FACS sorting.** Cells were sorted into 384-well plates using SH800S (Sony) sorter. Cells were sorted using the “Ultra purity” setting on the sorter. For a typical sort, a tube containing 0.3-1ml the pre-stained cell suspension was vortexed gently and loaded onto the FACS machine. A small number of cells were flowed at low pressure to check cell concentration and amount of

debris. Then the pressure was adjusted, flow was paused, the first destination plate was unsealed and loaded. Single-cell sorting was done where half the plate was sorted for CD45+/PI-/Hoechst+ while the second half was sorted for CD45-/PI-/Hoechst+. Immediately after sorting, plates were sealed with a pre-labeled aluminum seal, centrifuged and flash frozen on dry ice.

**cDNA synthesis and library preparation.** cDNA synthesis was performed using the Smart-seq2 protocol (Picelli et al., 2013, 2014). Briefly, 384-well plates containing single-cell lysates were thawed on ice followed by first strand synthesis. 0.6  $\mu$ l of reaction mix (16.7 U/ $\mu$ l SMARTScribe Reverse Transcriptase (Takara Bio, 639538), 1.67 U/ $\mu$ l Recombinant RNase Inhibitor (Takara Bio, 2313B), 1.67X First-Strand Buffer (Takara Bio, 639538), 1.67  $\mu$ M TSO (Exiqon, 5'-AAGCAGTGGTATCAACGCAGACTACATrGrG+G-3'), 8.33 mM DTT (Bioworld, 40420001-1), 1.67 M Betaine (Sigma, B0300-5VL), and 10 mM MgCl<sub>2</sub> (Sigma, M1028-10X1ML)) was added to each well using a Tempest liquid handler or Mosquito (TTP Labtech). Reverse transcription was carried out by incubating wells on a ProFlex 2x384 thermal-cycler (Thermo Fisher) at 42°C for 90 min and stopped by heating at 70°C for 5 min.

Subsequently, 1.5  $\mu$ l of PCR mix (1.67X KAPA HiFi HotStart ReadyMix (Kapa Biosystems, KK2602), 0.17  $\mu$ M IS PCR primer (IDT, 5'-AAGCAGTGGTATCAACGCAGAGT-3'), and 0.038 U/ $\mu$ l Lambda Exonuclease (NEB, M0262L)) was added to each well with a Mantis liquid handler (Formulatrix) or Mosquito, and second strand synthesis was performed on a ProFlex 2x384 thermal-cycler by using the following program: 1. 37°C for 30 minutes, 2. 95°C for 3 minutes, 3. 23 cycles of 98°C for 20 seconds, 67°C for 15 seconds, and 72°C for 4 minutes, and 4. 72°C for 5 minutes.

The amplified product was diluted with a ratio of 1 part cDNA to 10 parts 10mM Tris-HCl (Thermo Fisher, 15568025). 0.6  $\mu$ l of diluted product was transferred to a new 384-well plate using the Viaflow 384 channel pipette (Integra). Illumina sequencing libraries were prepared as described in (Darmanis et al., 2015). Briefly, tagmentation was carried out on double-stranded cDNA using the Nextera XT Library Sample Preparation kit (Illumina, FC-131-1096). Each well was mixed with 0.8  $\mu$ l Nextera tagmentation DNA buffer (Illumina) and 0.4  $\mu$ l Tn5 enzyme (Illumina), then incubated at 55°C for 10 min. The reaction was stopped by adding 0.4  $\mu$ l "Neutralize Tagment Buffer" (Illumina) and spinning at room temperature in a centrifuge at 3220xg for 5 min. Indexing PCR reactions were performed by adding 0.4  $\mu$ l of 5  $\mu$ M i5 indexing primer, 0.4  $\mu$ l of 5  $\mu$ M i7 indexing primer, and 1.2  $\mu$ l of Nextera NPM mix (Illumina). All reagents were dispensed with the Mantis or Mosquito liquid handlers. PCR amplification was carried out on a ProFlex 2x384 thermal cycler using the following program: 1. 72°C for 3 minutes, 2. 95°C for 30 seconds, 3. 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds, and 72°C for 1 minute, and 4. 72°C for 5 minutes.

**Library pooling, quality control, and sequencing.** Following library preparation, wells of each library plate were pooled using a Mosquito liquid handler. Pooling was followed by two purifications using 0.7x AMPure beads (Fisher, A63881). Library quality was assessed using capillary electrophoresis on a Fragment Analyzer (Agilent) or TapeStation (Agilent), and libraries were quantified by qPCR (Kapa Biosystems, KK4923) on a CFX96 Touch Real-Time PCR Detection System (Biorad). Plate pools were normalized to 2 nM and equal volumes from library plates were mixed together to make the sequencing sample pool.

**Sequencing libraries from 384-well plates.** Libraries were sequenced on the NextSeq or NovaSeq 6000 Sequencing System (Illumina) using 2 x 100bp paired-end reads and 2 x 8bp or 2 x 12bp index reads. NextSeq runs used high output kits, whereas NovaSeq runs used either a 200 or 300-cycle kit (Illumina, 20012860). PhiX control library was spiked in at 1%.

**Alignment and gene counts.** Sequences from the Illumina sequencing were demultiplexed using bcl2fastq version 2.19.0.316. Reads were aligned using the hg38 genome using STAR version 2.5.2b with parameters TK. Gene counts were produced using HTSEQ version 0.6.1p1 with default parameters except stranded was set to false and mode was set to intersection-nonempty.

**General clustering.** Standard procedures for filtering, variable gene selection, dimensionality reduction, and clustering were performed using the Seurat package (Satija et al., 2019) (version 2.3.4) in R (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>), where cells with fewer than 500 genes and 50,000 reads were excluded. Samples with less than 10 total cells were filtered from the analysis. Counts were log-normalized, then scaled by linear regression against the number of reads. Variable genes (Ngenes=6,112) were selected using a threshold for dispersion, with z-scores normalized by expression level. The variable genes were projected onto a low-dimensional subspace using principal component analysis. The number of principal components (Npcs) were selected based on inspection of the plot of variance explained (Npcs = 20). A shared-nearest-neighbors graph was constructed based with metric the Euclidean distance in the low-dimensional subspace. Cells were visualized using a 2-dimensional tSNE on the same distance metric (Res = 0.5, Kparam = 30, script 03). Cell types were assigned to each cluster of cells using the abundance of known marker genes (Supplemental Table 2, script S01-03 and script NI01).

**Epithelial subset clustering, identification of tumor cells, and annotation of non-tumor epithelial cells.** Cells previously annotated as epithelial (n = 5,109) were subset and re-clustered using methods described above and the following parameters: Ngenes = (6,919), Npcs = 20, Res = 0.9, Kparam = 10 (script NI02). Malignant epithelial cells were identified using inferCNV (Tickle T, Tirosh I, Georgescu C, Brown M, Haas B (2019). inferCNV of the Trinity CTAT Project. Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. <https://github.com/broadinstitute/inferCNV>.) inferCNV which works by finding cells with large copy number variations as determined by sorting expressed genes by their chromosomal location and applying a moving average, a sliding window of 100 genes within each chromosome, to the relative expression values (Patel et al., 2014; Puram et al., 2017; Tirosh et al., 2016). All epithelial cells as well as 300 fibroblasts and 300 endothelial cells were used as input (script NI03). An additional 500 fibroblasts and 500 endothelial cells were used as reference controls. We scored each cell for the extent of CNV signal and plotted cells on a dendrogram which was then cut at the highest point in which all the spiked in endothelial and fibroblasts cells belonged to one cluster (k = 3). All cells that clustered together with spiked in controls were labeled “nontumor”, whereas the remaining two clusters were labeled as “tumor”.

Nonmalignant epithelial cells (n = 1,489), as determined as those cells lacking large chromosomal aberrations from InferCNV analysis, were subset and re-clustered using the following parameters: Ngenes = (7,213), Npcs = 20, Res = 0.3, Kparam = 30 (script NI05). Cell types were assigned to each cluster of cells using the abundance of known marker genes (Supplemental Table 2) and differentially expressed genes as found by using the Seurat function *FindAllMarkers* using the default Wilcoxon rank sum test (arguments used: only.pos = TRUE, min.pct = 0.25, thresh.use = 0.25).

**Tumor cell subset clustering and differential gene expression.** Malignant epithelial cells (n = 3,620), as determined as those cells harboring large chromosomal aberrations from InferCNV analysis, were subset and re-clustered using the following parameters: Ngenes = 6,797, Npcs = 15, Res = 0.9, Kparam = 30 (script NI04). We found the differences in gene expression between the three treatment response groups (TN, RD, and PD) using the Seurat function *FindMarkers* using the default Wilcoxon rank sum test. Three separate tests were used to ascertain the differences between: 1) TN and RD, 2) TN and PD and 3) RD and PD (Supplemental Tables 5). Resulting differential gene lists were then filtered to limit patient specific effects. This is achieved by setting a threshold for non-zero expressing cells per patient (RD = 3, 37% of RD patients and PD = 6, 33% of PD patients) and removing differentially expressed genes explained by less than the thresholds set. The top 100 genes from each comparison were manually curated to evaluate for pathway activation. Decreased expression could indicate lack of detection due to the stochasticity of scRNAseq and thus for analysis of activated pathways we focused on upregulated genes. Gene signatures (Supplemental Table 2) were compiled using differentially expressed as well as known cell marker genes. Specifically, the alveolar signature (Supplemental Figure 3B) is made of differentially expressed AT1/AT2 genes among the cancer cell timepoint comparisons as well as additional known AT1/AT2 genes (Vieira Braga et al., 2019; Wade et al., 2006). The remaining signatures were identified directly from top differentially expressed genes.

To ensure that we were not misclassifying healthy AT2 cells as cancer cells, we compared the expression levels of our combined alveolar gene signature between the three timepoints (TN, RD, PD) and non-cancer AT2 cells from our dataset as well as additional non-cancer AT2 cells from an external dataset (Vieira Braga et al., 2019). Non-cancer AT2 cells from our dataset were more similar to the external AT2 cells than any of our cancer cells across all timepoints ( $\rho=0.73, -0.14, 0.23, -0.19$ , for non-cancer AT2 cells, and TN, RD, PD cancer cells respectively) (Supplemental Figure 3C).

Longitudinal analysis of a single patient was done by subsetting all cell originating from patient TH226. As above, the differences in gene expression between the three treatment response groups (TN, RD, and PD) was found by applying the Seurat function *FindMarkers* using the default Wilcoxon rank sum test. Three separate tests were used to ascertain the differences between: 1) TN and RD, 2) TN and PD and 3) RD and PD (script NI07-08, Supplemental Table 5)

**Survival analysis using cancer cell gene signatures within the TCGA.** TCGA LUAD data were downloaded from [xenabrowser.net/datapages/](http://xenabrowser.net/datapages/). Metadata was downloaded from An Integrated TCGA Pan-Cancer Clinical Data Resource (Liu et al., 2018a). Mean expression of each cancer cell expression signature (alveolar, kynurenine, plasminogen activating, *SERPINE1*, and gap junction) was calculated per TCGA sample. TCGA samples were then split by quartile groups. Only quartile 4 (high expression) and quartile 1 (low expression) were plotted using library packages survival (Therneau T (2015). A Package for Survival Analysis in S. version 2.38, <https://CRAN.R-project.org/package=survival>.) and survminer (Aldoukadel Kassambara, Marcin Kosinski and Przemyslaw Biecek (2019). survminer: Drawing Survival Curves using ‘ggplot2’. R package version 0.4.5. <http://CRAN.R-project.org/package=survminer>) in R (script NI10).

**Immunohistochemistry.** All specimens were acquired from individuals with NSCLC as noted above. 4-micron thick formalin-fixed paraffin embedded (FFPE) human tissue sections were processed using previously published method (Haderk

et al., 2019) and manufacturer's recommendations for antigen retrieval. Staining was performed overnight AQP4 rabbit monoclonal antibody (#59678, Cell Signaling Technology, 1:100 dilution). Stained slides were digitized using an Aperio ScanScope XT Slide Scanner (Leica Biosystems) with an either a 20X or 40X objective. Tumor populations were annotated, then scored in a blinded, randomized analysis by a pathologist for percent tumor positivity and subcellular staining intensity at the membrane, cytosolic, and nuclear compartments. Staining intensity was graded as negative, weak, intermediate, or strong and received scores of 0, 1, 2, or 3 respectively. Percent tumor positivity coefficient was graded as 0, negative; 1, less than 10% immunopositive; 2, between 10-50% immunopositive; 3, between 51-80% immunopositive; 4, greater than 80% immunopositive. Calculation of immunoreactivity scores was performed by multiplying the staining intensity score (0-3) with the percent tumor positive coefficient (0-4) to yield a value between 0 and 12 (Fedchenko and Reifenrath, 2014).

**Mutation detection from scRNAseq.** Alignment bam files for all non-immune cells (stroma and epithelial) were passed to GATK HaplotypeCaller which was run from the latest available Docker container (broadinstitute/gatk:4.0.11.0) using the following options:

```
-disable-read-filter MappingQualityReadFilter  
-disable-read-filter GoodCigarReadFilter  
-disable-read-filter NotSecondaryAlignmentReadFilter  
-disable-read-filter MappedReadFilter  
-disable-read-filter MappingQualityAvailableReadFilter  
-disable-read-filter NonZeroReferenceLengthAlignmentReadFilter  
-disable-read-filter NotDuplicateReadFilter  
-disable-read-filter PassesVendorQualityCheckReadFilter  
-disable-read-filter WellformedReadFilter
```

Disabling these specific read filters proved necessary for scRNAseq, as inherent low-coverage causes the vast majority of reads to be flagged for removal otherwise. The full human variant set (dbSNP) was downloaded from NCBI ([https://www.ncbi.nlm.nih.gov/variation/docs/human\\_variation\\_vcf/](https://www.ncbi.nlm.nih.gov/variation/docs/human_variation_vcf/)), and every variant call was assessed for its presence/absence in the human variant database. dbSNP is a public, living catalog of 674 million human somatic SNPs and indels that have been reported by peer-reviewed publications.

Cloud-based parallelization of HaplotypeCaller jobs was achieved with Reflow, a workflow engine for distributed, incremental data processing in the cloud (<https://github.com/grailbio/reflow>). HaplotypeCaller outputs a separate variant calling format file (VCF) for each cell, which were processed with the python package cerebra (<https://github.com/czbiohub/cerebra>). Variants found in dbSNP were removed, not to be included in further analysis. We reasoned that by removing 'common', population-level variants, we could better hone-in on disease specific variation.

In addition to scRNAseq reads, we obtained bulk DNA reads from peripheral blood for the majority of our patients (with the exception of three). These PBMC reads were run through HaplotypeCaller to establish 'germline' mutation profiles for each of our patients. Germline mutations were then subtracted out from each of that patient's single cell VCFs. This filtering step was omitted for the three patients for which we did not obtain peripheral blood, however, these single cell VCFs were still passed through our dbSNP filter.

We also applied a *fathmm* filter to all cells (Shihab et al., 2015). *fathmm* takes a machine learning approach to predict the likelihood of a given SNP to be pathogenic, integrating ENCODE annotations for things like transcription factor binding sites, histone modifications, cross-species sequence alignment and conservation scores, etc. Only variants computationally predicted to be pathogenic were included in our analysis, i.e. those variants with a *fathmm* score > 0.7.

The remaining variants were then filtered through the COSMIC (Catalogue of Somatic Mutations in Cancer) complete mutation – genome screens database (<https://cancer.sanger.ac.uk/cosmic/download>). Only SNPs/indels associated with 'Lung' as per their COSMIC annotation were kept. Variant calls were mapped to their corresponding genes, and per-patient / per-sample mutational profiles were established. We used the ERCCs spiked into each cell sample as a negative control for false positive mutations, which can arise due to technical artifacts such as PCR errors. We found the median false positive mutation rate to be 0.000256% per base (Enge et al., 2017).

**Fusion detection from scRNAseq.** Fusion transcripts were detected with STAR-fusion (<https://github.com/STAR-Fusion/STAR-Fusion/wiki>) version 1.6.0, run from a Docker container (trinityctat/ctatfusion:1.5.0). The following options were used:

```
-FusionInspector validate  
-examine_coding_effect  
-denovo_reconstruct
```

Distributed processing of STAR-fusion jobs was accomplished with Reflow. Output files were processed with cerebra, then combined with variant calls to create per-cell and per-sample summary tables.

**Mutational analysis of tumor cells.** Information extracted from cerebra was summarized by sample. Coverage information was provided by a secondary output from cerebra summarized by sample and gene. Where all cells are summarized by sample and all *fathmm* filtered ROIs are summarized by corresponding gene (script NI06). Plots were generated using the R pheatmap package (Raivo Kolde (2019). pheatmap. Pretty Heatmaps. R package version 1.0.12. <https://CRAN.R-project.org/package=pheatmap>).

Survival analysis using number of mutations within the MSK-Impact data MSK-Impact data was downloaded from cBioPortal (Cerami et al., 2012; Gao et al., 2013) and subset to only NSCLC samples MSK-Impact data was subset to only those mutations that were also found in the scRNAseq dataset of mutations (n=141 unique mutations). We stratified MSK-Impact samples by those with greater than or equal to 2 mutations from the tier one COSMIC mutations found in the scRNAseq dataset (mutation high), and those less than 2 mutations (mutation low) (Figure 2D). Kaplan-Meier plots were visualized with the lifelines package in python (Cameron Davidson-Pilon, Jonas Kalderstam, Paul Zivich, Ben Kuhn, Andrew Fiore-Gartland, Luis Moneda, ... André F. Rendeiro. (2019, September 4). CamDavidsonPilon/lifelines: v0.22.4 (Version v0.22.4). Zenodo. <http://doi.org/10.5281/zenodo.3386382>; Python Software Foundation. Python Language Reference, version 3.4. Available at (script NI12).

**Immune subset clustering and differential gene expression.** All cells annotated as immune (n=12,077) were subset and clustered as described above (script IM01) using the following parameters (Ngenes=1,690, Npc=20, Res=0.3, k.param=10). The resulting 17 clusters were assigned to different major immune cells types using a list of curated gene markers (Supplemental Table 2) and by manual curation of differentially expressed genes for each cluster (Supplemental Table 5). The different cell types and number of cells belonging to each type are described in the main text.

To assess changes in fractional abundance of different immune cell populations we used all cells though excluded thoracentesis and brain samples due to difference in the immune makeup of these tumor environments which would skew the data. The function freqCI from the R package REdaS (Maier MJ (2015). Companion Package to the Book R: Einführung durch angewandte Statistik. R package version 0.9.3. <http://CRAN.R-project.org/package=RedaS>) (script IM02) was used to calculate confidence intervals for relative frequencies.

Macrophages (n=1,034) and T-cells (n=1,725) from lung biopsies were subset and clustered as described above (script IM03 and IM04 respectively) using the following parameters for MFs (Ngenes=1,637, Npc=10, res=0.3, k.param=10) and T-cells (Ngenes=1,475, Npc=10, res=0.3, k.param=10). The resulting clusters are discussed in the main text and the lists of differentially expressed genes are provided (Supplemental Table 5). We repeated this analysis where we subset the data to only patients with multiple biopsies and sufficient cells (TH226 and TH266) (script IM05).

**Multiplex Immunofluorescence.** Multiplex immunofluorescence staining was performed on sequential 4 micron FFPE slides, utilizing the Opal IHC Multiplex Assay (Perkin Elmer). Slides were stained with one of two 6 antibody panels. Panel 1 contained primary antibodies against PD-L1 (clone E1L3N, dilution 1:50, Cell Signaling Technologies), CD68 (clone KP1, dilution 1:500, Dako), IDO (clone D5J4E, dilution 1:100, Cell Signaling Technologies), HLA-DR (clone CR3/43, dilution 1:250, Abcam), CD14 (clone SP192, dilution 1:100, Abcam), and cytokeratin (polyclonal Z0622, dilution 1:250, Dako). Panel 2 consisted of primary antibodies against CD3 (clone LN10, Leica), PD-1 (clone NAT105, dilution 1:100, Abcam), CD14, CD8 (clone C8/144B, dilution 1:100, Dako), FoxP3 (clone 236A/E7, dilution 1:200, Abcam), and cytokeratin. Whole slide scans were acquired at 10X via the Vectra imaging system (Perkin Elmer, version 3.0). Three to six regions from each slide containing tumor and stroma were selected utilizing Phenochart (v1.0.8, Perkin Elmer) for high resolution multispectral acquisition on the Vectra system at 20X magnification. Spectral libraries were generated from single color stains on human tonsil tissue. Spectral unmixing, cell segmentation, and cell phenotype assignment as tumor, macrophage, T cell, or other cell population were performed utilizing InForm image analysis software (Perkin Elmer, version 2.4). Fractions of macrophage and T-cell populations were calculated as: (population of interest) / (macrophage + T-cell populations) and plotted using 'ggplot2' in R (H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016) (script NI15).

**Survival analysis using fractional immune population within the TCGA.** As with the survival analysis using cancer cell gene signatures, we used the downloaded TCGA LUAD dataset and metadata to access patient survival outcomes as they pertain to the fractional changes of immune populations within a given tumor. We used CIBERSORT (Newman et al., 2015) to deconvolute the bulk TCGA samples into relative fractions of immune cell populations as determined by using the LM22 reference. The total macrophage population was found by combining fractions for Monocytes, Macrophages.M0, Macrophages.M1, and Macrophages.M2. The total T-cell population was found by combining fractions of T.cells.CD8, T.cells.CD4.naive, T.cells.CD4.memory.resting, T.cells.CD4.memory.activated, T.cells.follicular.helper, T.cells.regulatory..Tregs, T.cells.gamma.delta, NK.cells.resting, and NK.cells.activated. TCGA samples were then split by quartile groups. Only quartile 4 (high population fraction) and quartile 1 (low population fraction) were plotted using library packages survival (Therneau T (2015). A Package for Survival Analysis in S. version 2.38, <https://CRAN.R-project.org/package=survival>) and survminer (Aldoukadel Kassambara, Marcin Kosinski and Prze-

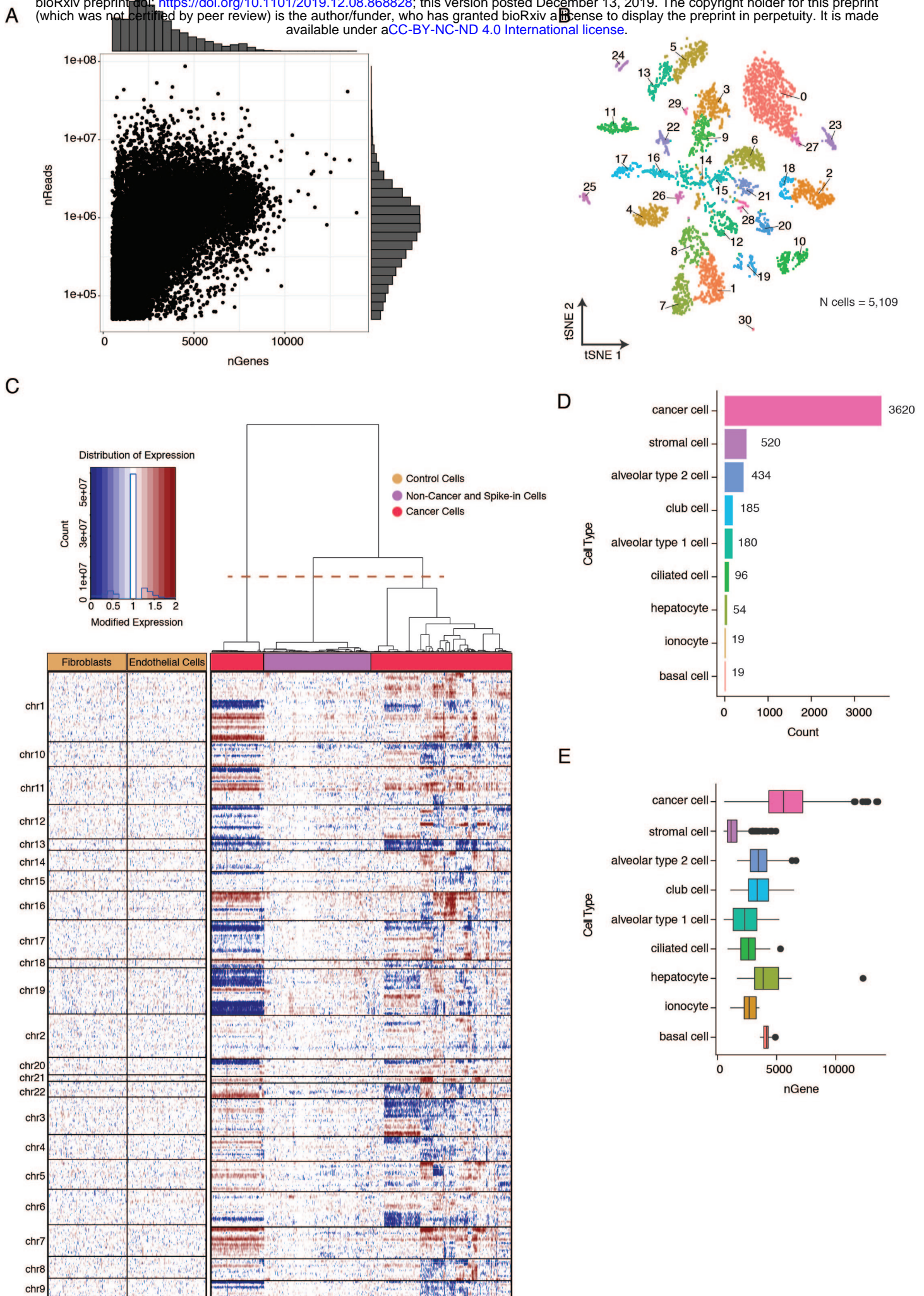


mylsław Biecek (2019). survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.5. <http://CRAN.R-project.org/package=survminer> in R (script NI10).

**Generation of in vitro drug tolerant persister and acquired resistance cells and RT PCR.** For validation of candidate gene expression via a RT2 Profiler PCR array (Qiagen, CLAH34795), human lung cancer PC9 (EGFR del19) cells were treated for 48 hours (day 2) with DMSO (TN) or for 7 and 19 days with 2 $\mu$ M Osimertinib (Selleck Chemicals LLC, S7297) with replenishment of drug every 3-4 days (RD), respectively. PD samples were derived from an acquired resistant PC9 cell line (Osimertnib IC50 = 89 $\mu$ M), that was generated by continuous treatment with 2 $\mu$ M Osimertinib with replenishment of drug every 3-4 days and presented active proliferation under drug when resistance was called. RNA was extracted via RNeasy Mini Kit (Qiagen, 74104). RNA quality was confirmed as RIN <sup>3</sup> 7.5 via Bioanalyzer RNA 6000 Pico kit (Agilent, 5067-1514) and RNA was quantified via Qubit RNA HS Assay kit (Thermo Fisher Scientific, Q32852). A total of 400ng of RNA was reverse transcribed using the First Strand Synthesis Kit (Qiagen, 330401) and then loaded into a custom 384 well RT2 profiler array (Qiagen, CLAH34795). Fold Change was calculated by determining the ratio of mRNA levels to control (day 2) values using the delta threshold cycle (Ct) method (DCt). A t-test was used to find the significance of change between baseline (day 2) and treated timepoints (days 7, 19 and 70) based on normalized Cts to baseline (script NI14).

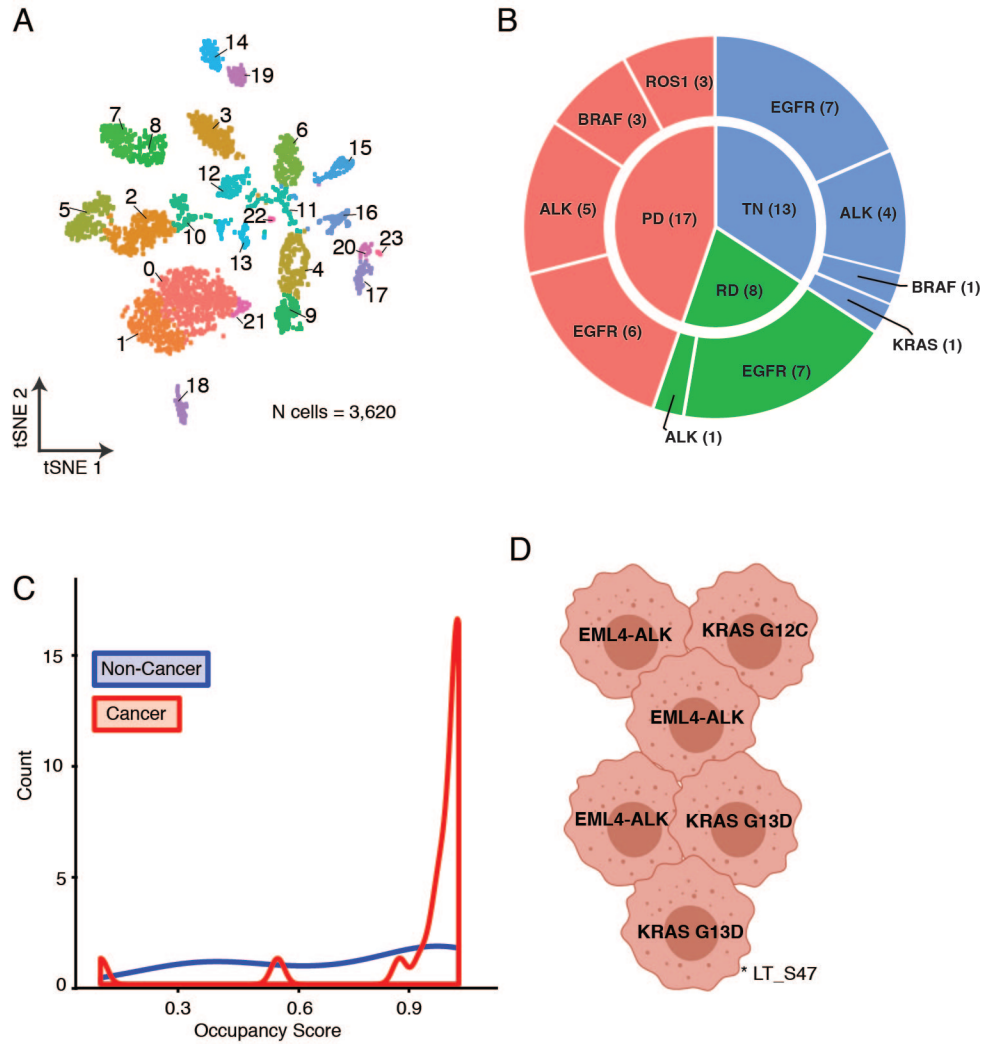
## Supplementary Figures

DRAFT



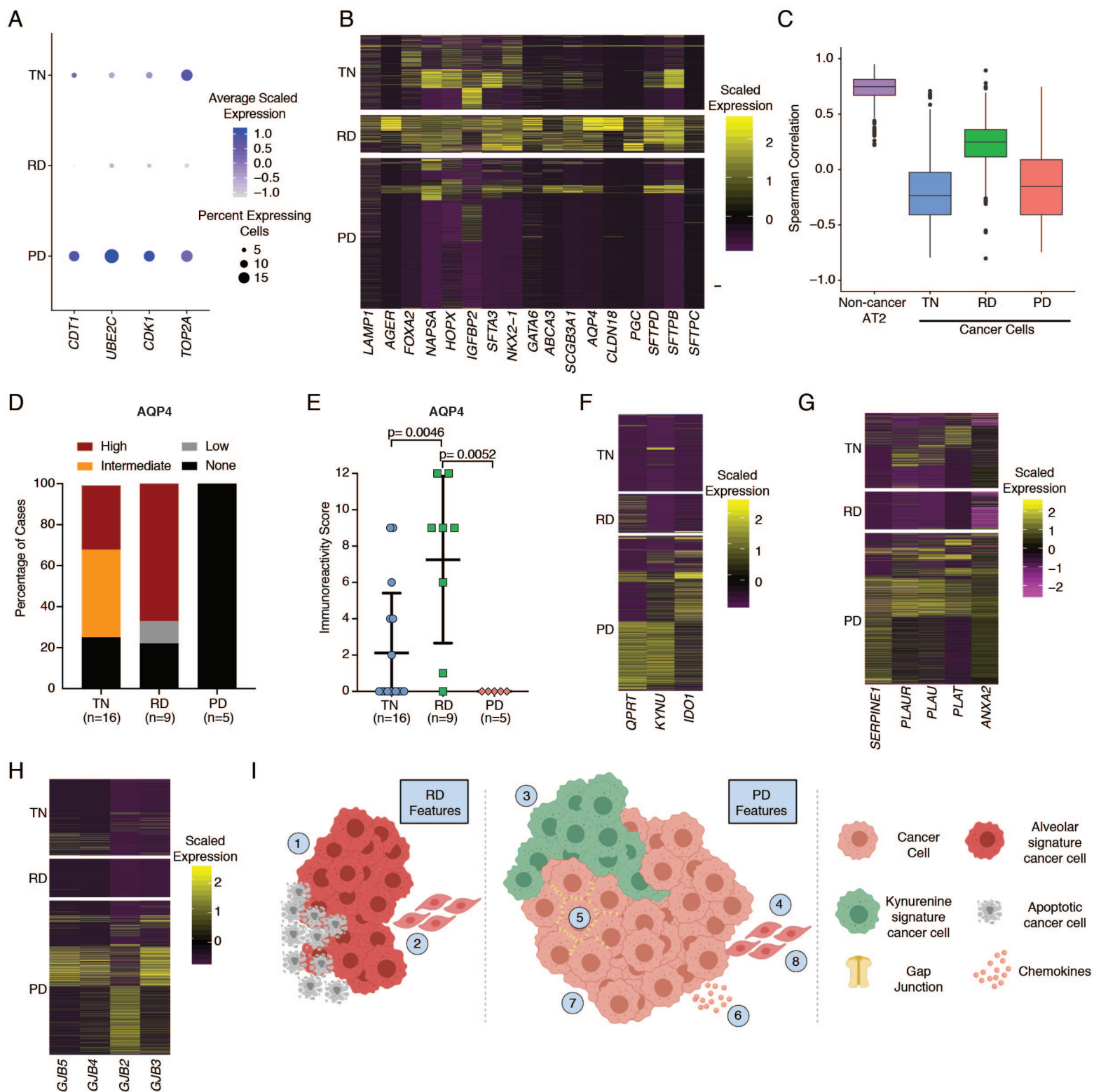
**Fig S1 related to Figure 1** (A) Scatterplot with marginal histograms showing the number of unique genes plotted against the number of reads for all cells passing quality control. (B) t-SNE of all epithelial cells (n=5109), numbers correspond to individual clusters. (C) Inferred large-scale copy number variations (CNVs) help identify cancer (pink) and non-cancer cells (purple). Epithelial and spike in control cells are included in the x-axis and chromosomal regions on the y-axis. Amplifications (red) or deletions (blue) were inferred by averaging expression over 100-gene stretches on the respective chromosomes. (D) Bar plot of cell counts for annotated epithelial cells. (E) Bar plot of the number of unique genes across all annotated epithelial cell types.

DRAFT



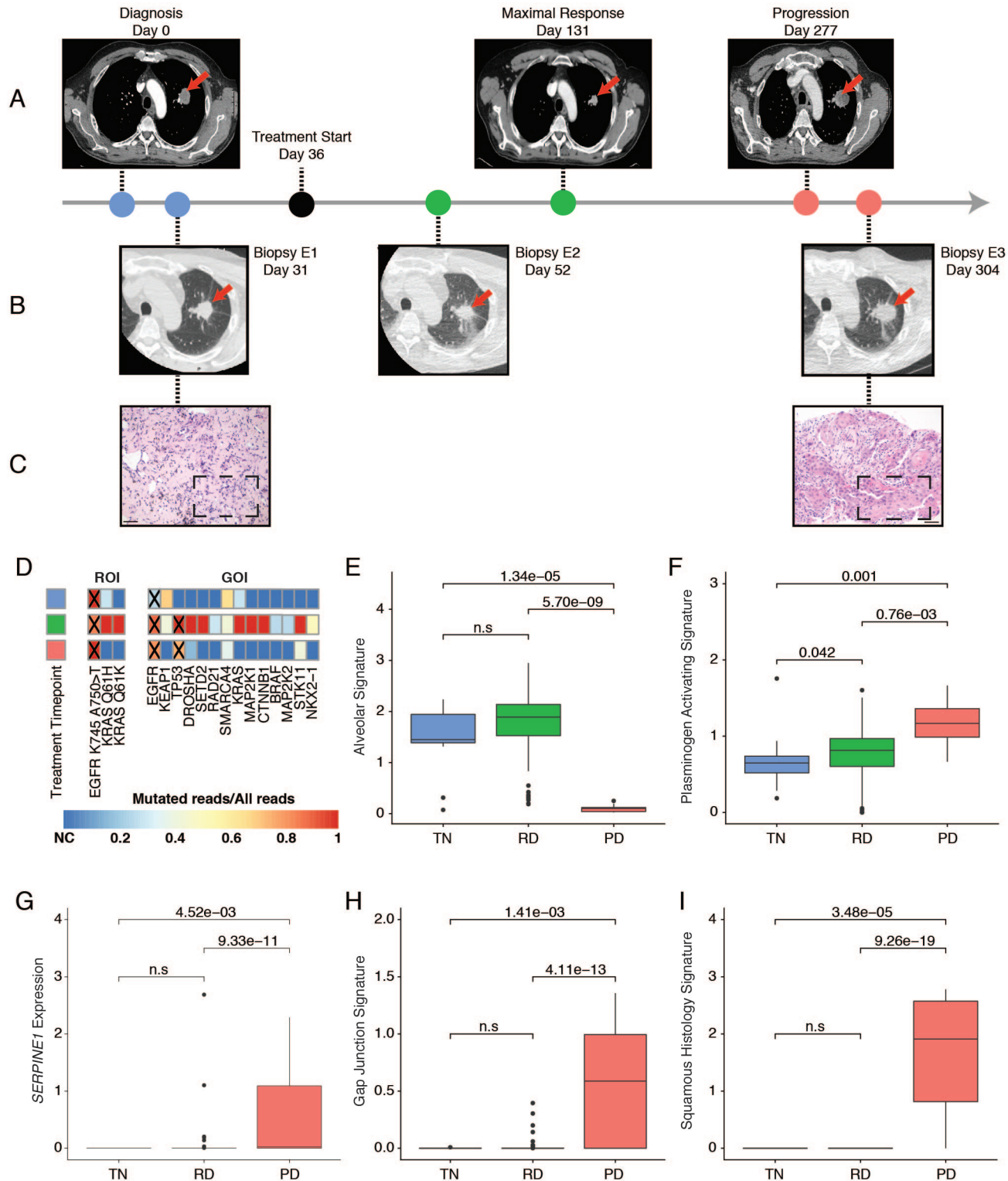
**Fig S2 related to Figure 2** (A) t-SNE plot of 3620 cancer cells from 38 samples, numbers indicate individual clusters. (B) Circle plot illustrating the clinically identified oncogenic driver (outer circle) and timepoints (inner circle) of each biopsy, only for cancer cells. (C) Density distribution of cluster occupancy of cancer (red) and non-cancer (blue) epithelial cell clusters, calculated as the percentage of the highest contributing individual patient over the total number of cells for that cluster. (D) Illustration of heterogeneity of primary driver mutated cancer cells found in exemplary sample LTS47 (\* in Figure 2A).

DRAFT



**Fig S3 related to Figure 3** (A) Dot plot of the relative expression of established cellular proliferation genes (x-axis) across treatment timepoints (y-axis). The color intensity scale reflects the average gene expression and the size scale indicates the number of cells expressing the gene within that treatment timepoint. Applying grouped, pairwise comparisons of treatment timepoints of the average scaled expression of all genes demonstrated significantly different expression ( $p < 0.000001$ ) in all comparisons. (B) Heatmap showing the expression of genes in the alveolar signature. Cells are grouped by treatment timepoint. (C) Box plot of Spearman correlations of cancer cells from all treatment timepoints and healthy AT2 cells to an external reference of healthy AT2 cells. Non-cancer AT2 cells from our dataset were more similar to the external, healthy AT2 cells than any of our cancer cells across all timepoints (mean spearman correlation=0.73, -0.14, 0.23, -0.19, for healthy AT2 cells, and TN, RD, PD cancer cells, respectively). (D) Scoring of IHC stains for AQP4 in TN (n=19), RD (n=9), and PD (n=5) tumor tissue sections. Chi-square testing demonstrated a significant difference ( $p = 0.0038$ ) in proportions of staining intensity groups between treatment timepoints. (E) Immunoreactivity score for AQP4 across all timepoints from S3D. (F, G, H) Heatmaps showing the expression of genes within each signature (kynurenine, SERPINE1/plasminogen activation, and gap junction, respectively) grouped by treatment timepoint. (I) Graphical summary of cancer cell expression changes across treatment timepoints. RD features include (1) Alveolar signature, and (2) various RD specific invasive signaling pathways. PD features include: (3) kynurenine signature, (4) plasminogen activation and SERPINE1 signatures, (5) gap junction proteins, (6) expression of pro-inflammatory chemokines, (7) loss of tumor suppressor genes, and (8) various PD specific invasive signaling pathways.

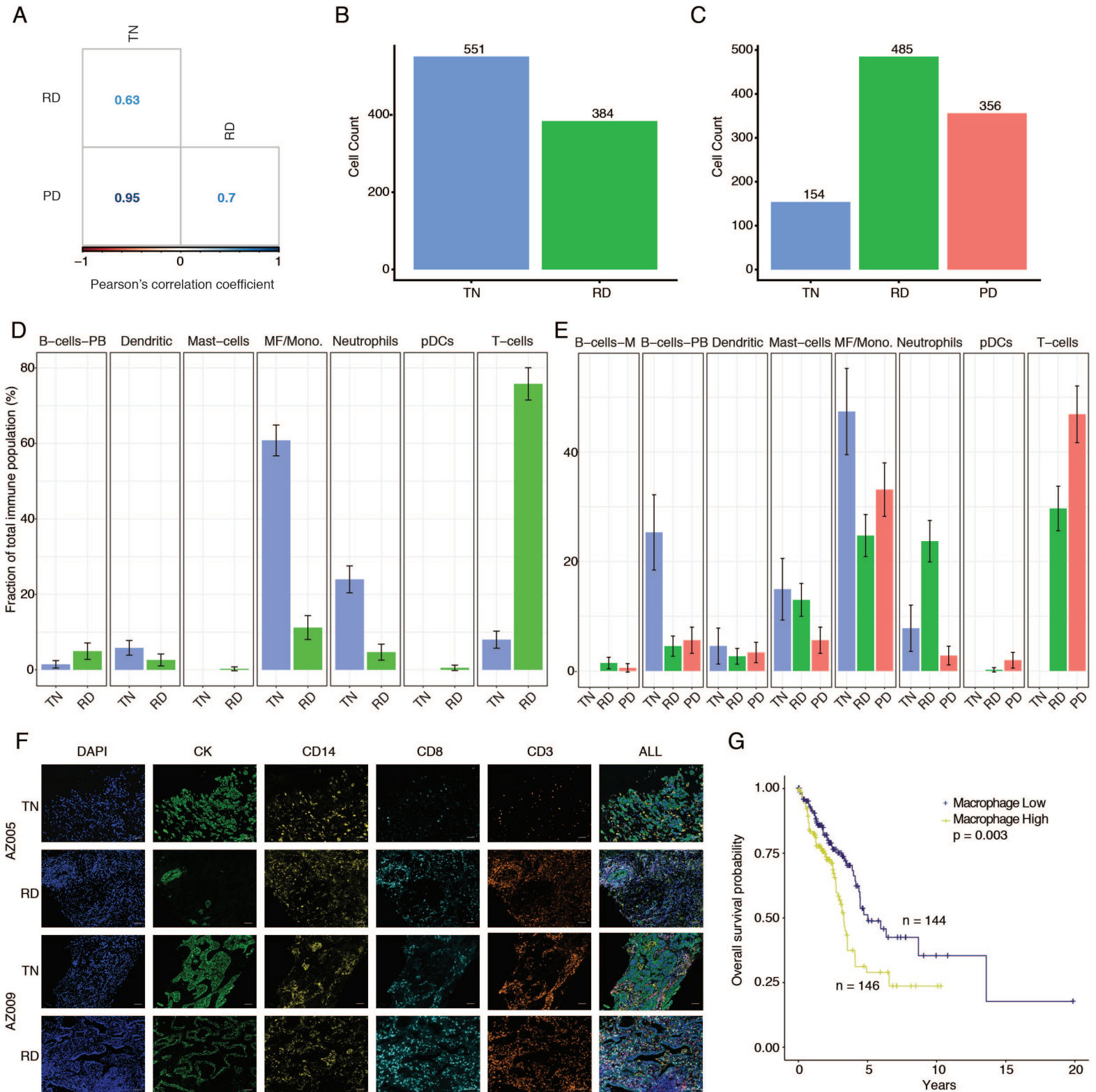
DRAFT





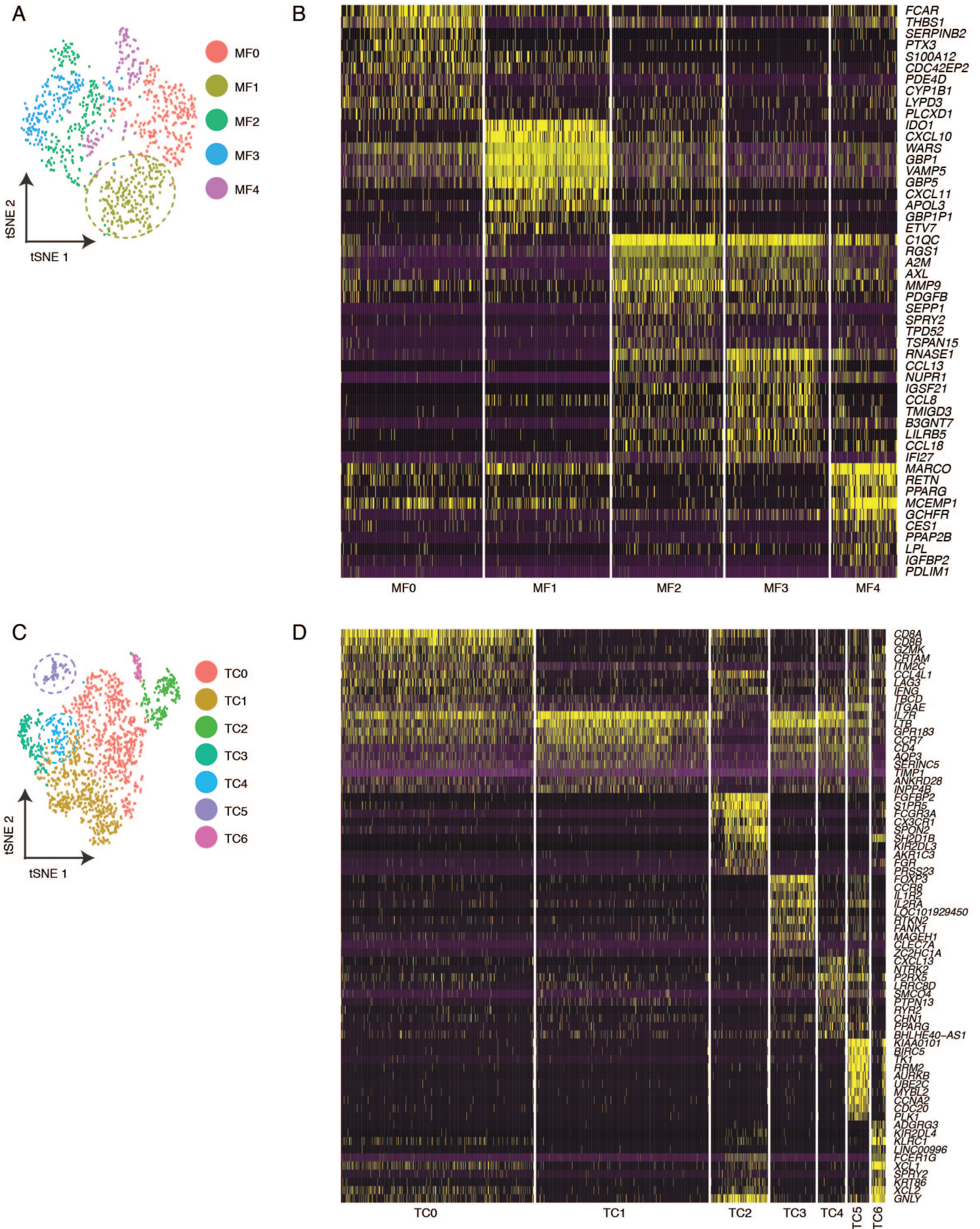
**Fig S4 related to Figure 3** (A, B, C) Longitudinal timeline of patient treatment (A) Chest CT scan at each clinical evaluation timepoint (B) Biopsy timepoint with procedural CT scan (C) HE from treatment naïve and progression timepoints demonstrating adenocarcinoma and squamous cell carcinoma, respectively, scale bar indicates 50  $\mu\text{m}$  (D) Heatmap of mutation state in COSMIC tier 1 mutated genes, where color represents mutated reads/all reads across treatment. Mutated genes as identified by clinical NGS are indicated by “X” (E-I) Boxplots of pathway signature changes (alveolar, plasminogen activating, SERPINE1, gap junction and squamous histology, respectively) across treatment timepoints.

DRAFT



**Fig S5 related to Figure 4** (A) Pairwise Pearson correlations between each treatment group's immune cell compositions which corresponds to the fraction of each immune cell type's abundance in the total immune cell population (B) Total immune cells for each biopsy of patient TH266. (C) Total immune cells for each biopsy of patient TH226. (D) Fraction of each immune cell sub-type for the two biopsies of patient TH266. Error bars indicate the 95% confidence interval for the calculated relative frequencies. (E) Fraction of each immune cell sub-type for the three biopsies of patient TH226. Error bars indicate the 95% confidence interval for the calculated relative frequencies. (F) Representative in situ immunofluorescence images from two patients with matched samples at different treatment timepoints, demonstrating fractional changes in the immune populations of macrophages and T-cells. Scale bars correspond to 50  $\mu\text{m}$  (G) Kaplan-Meier plot of deconvoluted TCGA lung adenocarcinoma data showing the relation between OS and the fraction of macrophages for each patient. Patients were stratified by high and low macrophage fraction.

DRAFT



**Fig S6 related to Figure 5** (A) t-SNE plot of all lung-derived macrophage cells. The cluster of cells enriched at PD (MF1) is highlighted (B) Heatmap showing the expression level of the top 10 differentially expressed genes for each macrophage cluster (C) t-SNE plot of all lung-derived T-cells. The clusters of cells enriched at PD (TC4 and TC5) are highlighted (D) Heatmap showing the expression level of the top 10 differentially expressed genes for each T-cell cluster.

DRAFT