

# Scaling comparative lung cancer studies using Chromium Single Cell Gene Expression Flex

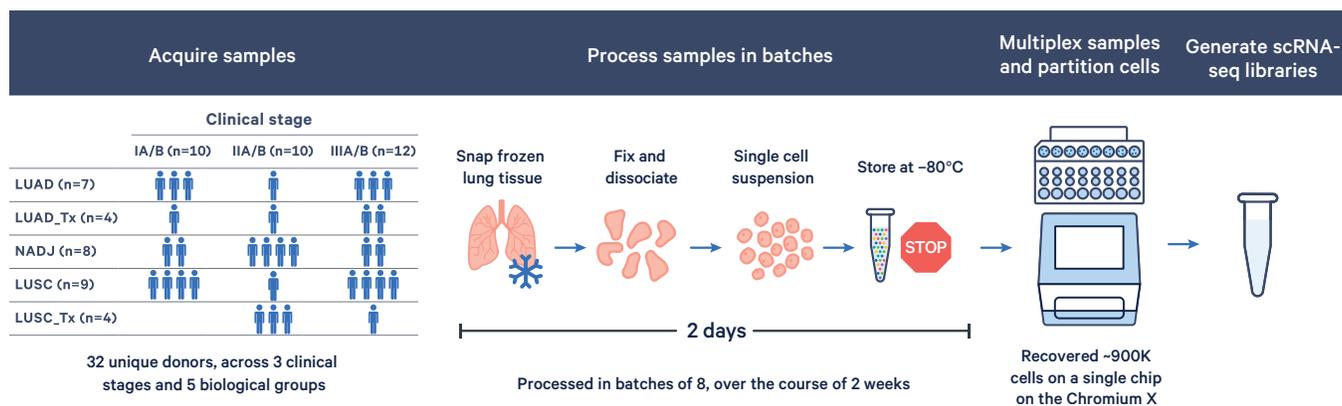
## Abstract

Lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) are both subtypes of non-small cell lung cancer (NSCLC), yet typically present a wide disparity in patient prognosis. Understanding the molecular underpinnings of these different outcomes requires transcriptomic studies at the single cell level.

Here, we used Chromium Single Cell Gene Expression Flex to perform a comparative study of LUAD and LUSC, profiling 32 samples, across three clinical stages. NSCLC and normal adjacent (NADJ) samples were fixed and dissociated in batches of eight over multiple days. Batched samples were multiplexed and partitioned on a single chip, for a total of ~900K cells in one run. Pseudobulk differential gene expression analysis identified tumor-specific enrichment of several markers correlated with poor prognosis. Changes in cell-type distribution revealed differences in specific cell proportions across sample types. While expected cell types were consistently present, analysis revealed deregulation of canonical cell markers across biological groups.

## Highlights

- Flexible sample prep workflow with multiple, optional stopping points enables storage and batching of samples
- Built-in multiplexing capability allows profiling of clinical samples at scale, for up to 1M cells on a single chip
- Robust fixation protocol expands sample access for multi-site translational studies without compromising data quality
- Comparative analysis identifies NSCLC tumor-specific cell types that had been previously annotated as rare cell types



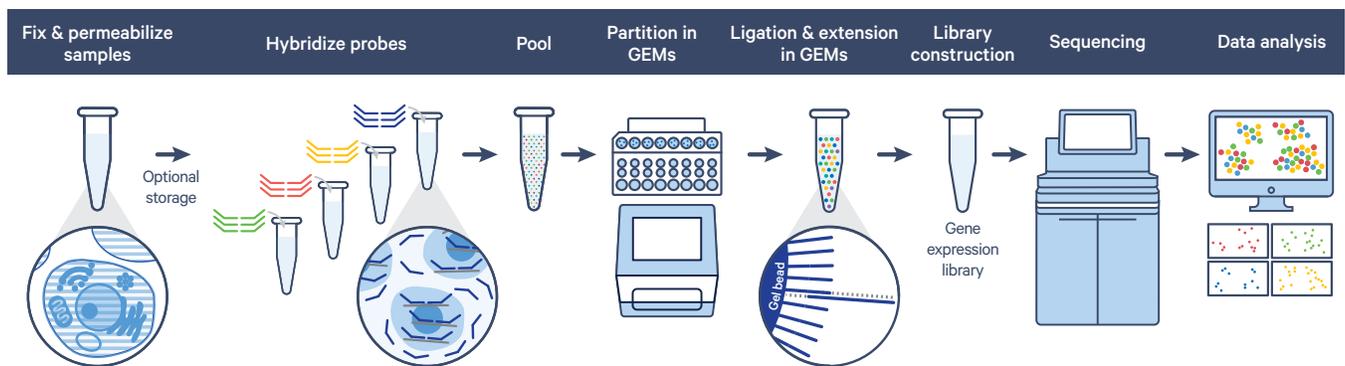
**Figure 1. Experimental design.** Fresh frozen NSCLC and NADJ tissues were obtained from 32 unique donors. The NSCLC tumor subtypes are displayed by row (LUAD, LUSC, and treated LUAD and LUSC [LUAD\_Tx and LUSC\_Tx, respectively]), and clinical stages are indicated by column. Samples were processed (fixed, dissociated, and stored) in batches of eight over the course of two weeks. Samples were barcoded, pooled, and then run on a single Chromium chip for cell partitioning prior to library generation.

## Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide and, of these diagnoses, 85% (1) are attributed to non-small cell lung cancer (NSCLC). Lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) make up the majority of NSCLC cases, yet those diagnosed with LUAD typically have a poorer prognosis, impacting smokers and non-smokers alike. While comparative transcriptomic single cell studies have aimed to uncover the differences underlying the two lung cancer subtypes, these studies have mostly focused on specific cell populations or late-stage tumors (2, 3, 4).

LUAD and LUSC are both histological subtypes of NSCLC, yet they are considered to have different cells of origin (5). LUAD tumors are often found more distally in the lung and are thought to originate from the alveolar progenitor cells, alveolar type 2 (AT2) cells. LUSC is typically found more proximally and is believed to originate from basal cells, which are the progenitors that give rise to the airway epithelium. Given the different cells of origin for each tumor subtype, it is important to identify the molecular changes that take place at the single cell level to further understand disease progression in LUAD versus LUSC.

In this Application Note, we used the Chromium Single Cell Gene Expression Flex assay to profile ~900K cells from 32 human lung donors (tumor subtypes and clinical stages are summarized in Figure 1). The Flex workflow enabled samples to be processed in batches, stored at  $-80^{\circ}\text{C}$ , and then pooled before running on a single Chromium chip (Figure 1). A pseudobulk differential gene expression analysis comparing the five different biological groups (NADJ, treated and treatment-naïve LUSC and LUAD) revealed a tumor-specific enrichment of several markers associated with poor tumor prognosis. Additionally, many canonical cell type-specific markers were differentially expressed across the biological groups. To understand if these gene expression differences were driven by changes in cell-type proportions, we performed automated cell typing and quantified the changes in cell-type distribution. While all of the expected lung cell types were present, there were differences in specific cell proportions between the sample types. However, within a specific cell population, many of the cell-type markers appeared to be deregulated across the biological groups.



**Figure 2. Chromium Single Cell Gene Expression Flex workflow overview.** Fixed samples were hybridized with barcoded probes, pooled, then partitioned in the Chromium X instrument. Resulting GEMs were transferred and incubated for probe ligation and addition of a 10x GEM Barcode. Gene expression libraries were constructed during subsequent steps, then sequenced on an Illumina NovaSeq system and analyzed using Cell Ranger.

## Methods

### Sample preparation

Fresh frozen, human NSCLC and NADJ samples (30–84 mg each) of RIN scores 7 or greater were acquired from BioIVT ASTERAND®. Samples were fixed and dissociated in batches of eight across a two-week time span. Each sample was fixed for 20 hours, then dissociated into single cell suspensions as described in the Tissue Fixation & Dissociation for Chromium Fixed RNA Profiling Demonstrated Protocol (CG000553). Fixed cell suspensions were stored at –80°C following the Long-Term Storage guidance.

### Hybridization and pooling

One million cells per sample were added to each individual probe hybridization reaction and incubated at 42°C for 17 hours. The following day, samples were stained with ViaStain AO/PI staining solution and counted using the Cellaca MX High-throughput Automated Cell Counter. Barcoded samples were pooled using the Fixed RNA Profiling for Multiplexed Samples Pooling Workbook (CG000565), washed, and resuspended at a concentration of 10,300–11,500 cells/μl.

### Single cell partitioning, library construction, and sequencing

Each 16-plex sample was loaded onto four individual chip lanes, targeting a cell recovery of 128,000 cells per lane. A total of 897,773 cells were partitioned on a single Chromium chip. Fixed RNA gene expression libraries were constructed per the User Guide (CG000527) and sequenced on an Illumina NovaSeq 6000 system using paired-end reads with a 28 bp (R1), 10 bp (i7), 10 bp (i5), and 90 bp (R2) configuration at a targeted sequencing depth of ~20,000 read pairs per cell.

## Analysis

### Primary data processing

Cell Ranger v7.1 was used to process the Gene Expression Flex assay data. A single instance of cellranger multi was used to process the full dataset.

### Cell-type annotation

Automated cell-type annotation was performed using Azimuth (6). In the first step, the human lung cell atlas reference (7) was used (available from the Azimuth webpage). Then, we created a lung cancer-specific reference based on published data (3) and re-annotated the cells manually.

16-plex sample	Number of reads (M)	Estimated # of cells	Valid barcodes	Mean reads per cell	Confidently mapped reads in cells	Reads confidently mapped to probe set	Sequencing saturation
hLung 1–16, Rep 1	3,491	120,835	92.7%	28,896	90.0%	90.6%	74.1%
hLung 1–16, Rep 2	2,489	130,781	95.9%	19,039	91.6%	91.4%	68.5%
hLung 1–16, Rep 3	3,005	130,713	92.6%	22,991	92.1%	90.7%	68.6%
hLung 1–16, Rep 4	2,499	129,071	93.0%	19,366	89.9%	90.8%	64.9%
hLung 17–32, Rep 1	2,841	95,360	94.4%	29,796	92.4%	87.8%	71.7%
hLung 17–32, Rep 2	2,456	96,673	94.4%	25,413	92.0%	87.9%	67.3%
hLung 17–32, Rep 3	3,259	102,252	94.3%	31,882	92.2%	87.8%	72.9%
hLung 17–32, Rep 4	2,921	92,048	94.4%	31,741	92.8%	87.8%	72.7%

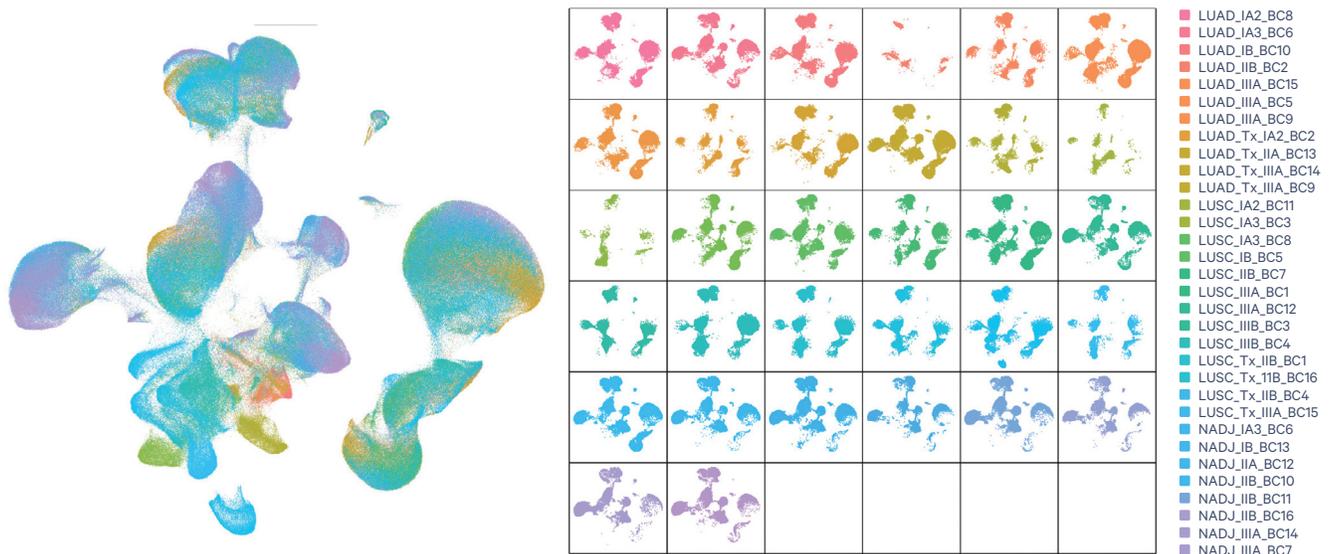
**Table 1. Summary of library-level sequencing metrics.** Human NSCLC and NADJ samples are labeled as "hLung," followed by multiplex pool notation and replicate identifier. Number of reads is reported in millions (M).

## Results

### Profiling ~900K cells on a single chip with Flex

In order to examine a variety of NSCLC tumors at scale, we took advantage of the Chromium Single Cell Gene Expression Flex assay and its built-in multiplexing capability to profile hundreds of thousands of cells from 32 unique human donors on a single chip (Figure 2). In total, 897,773 cells were recovered; 254,429 cells from

NADJ, 168,978 cells from LUAD, 123,642 cells from LUAD\_Tx, 225,536 from LUSC, and 125,148 cells from LUSC\_Tx. Library-level metrics are detailed in Table 1. UMAP projections of samples split by donor are shown in Figure 3.

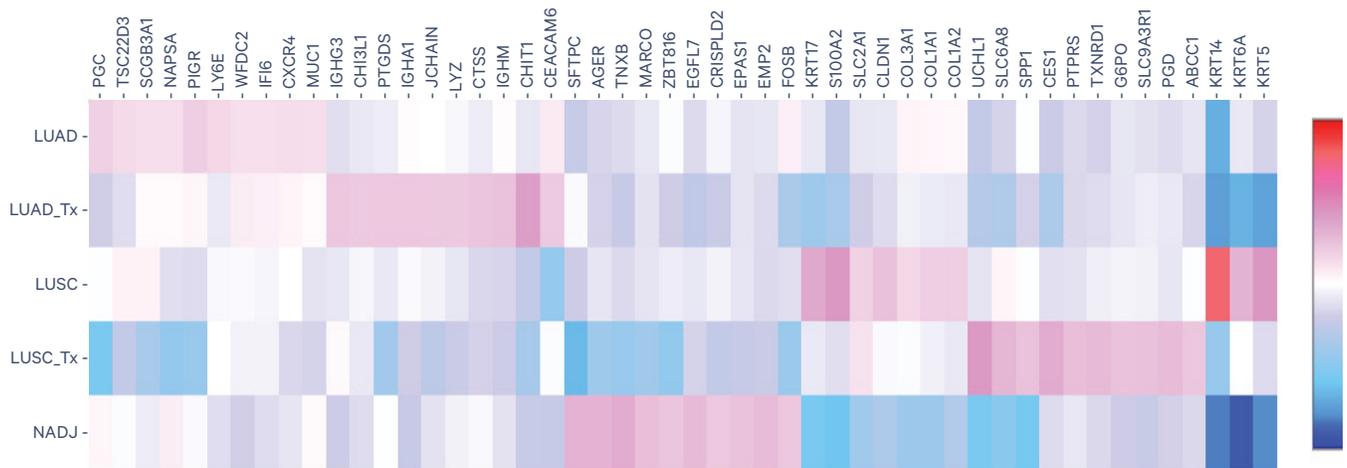


**Figure 3. UMAP projection of ~900K NSCLC dataset, color-coded and split by donor.** Samples are labeled with biological group (sample type and treatment status), clinical stage (IA–IIIA), and probe barcode (BC1–BC16).

### Top differentially expressed genes are associated with poor prognosis

A pseudobulk differential gene expression analysis comparing the five biological groups revealed an enrichment of several markers known to be associated with tumor progression and poor prognosis (Fig. 4). Considered to have tumor suppressor properties, markers for AT1 cells *AGER* and *EMP2* (8, 9) showed decreased expression in tumors relative to NADJ samples. Similarly, AT2 marker *SFTPC* and macrophage marker *MARCO* have previously been described as potentially useful diagnostic biomarkers in the detection of LUAD (11), and these showed decreased expression in tumor samples relative to NADJ tissues. *MUC1* and *CEACAM6*, markers associated with poor tumor prognosis (12, 13), showed increased expression levels, specifically in LUAD tissues.

In addition to the enrichment of markers associated with poor tumor prognosis, several of the top differentially expressed genes are also used as cell-type markers, suggesting possible differences in the recovery of specific cell populations across the biological samples. For example, LUAD tumors had high expression of alveolar and secretory cell markers whereas LUAD\_Tx tumors showed an enrichment of markers associated with innate and adaptive immune cells (Fig. 4). LUSC tumors displayed an upregulation in basal cell markers, while LUSC\_Tx tumors showed an increase in rare and tumor cell markers. Given that several of the top differentially expressed genes are specific to distinct cell populations, we asked whether these differences were being driven by changes in overall cell composition across biological groups.



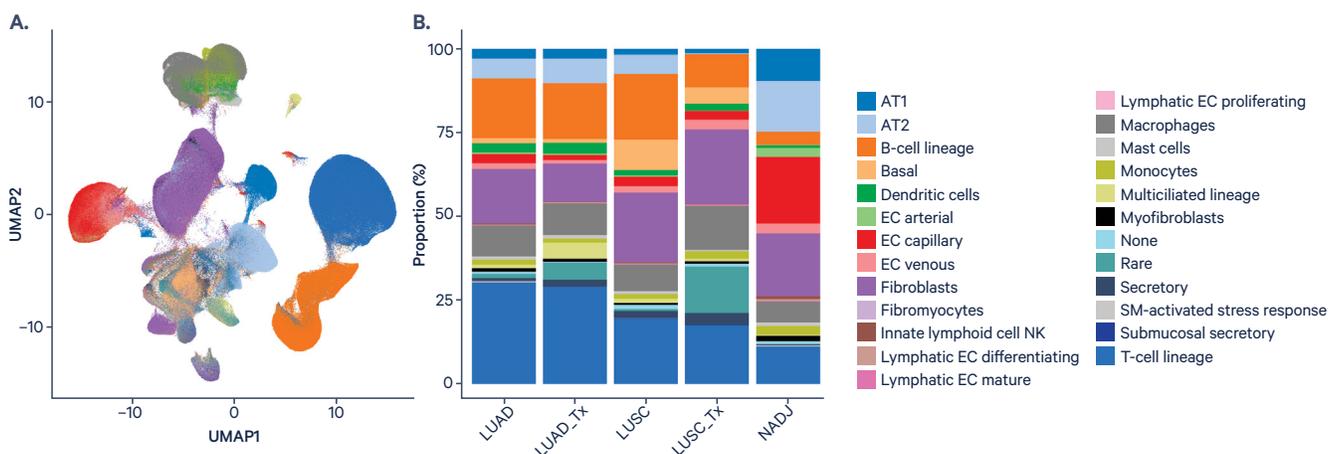
**Figure 4. Heatmap showing distinct gene signatures for each of the five biological groups.**

### Deregulation of cell-specific markers in tumor samples

Using Azimuth’s healthy lung reference, automated cell-type annotation was performed and changes in specific cell proportions were quantified across the different biological groups. Notably, there was a dramatic increase in “rare” and unknown cell types within the tumor samples. To better understand what these “rare” cell types might be, cells were manually reannotated using a reference specific to lung cancer (3). Approximately 50% of those cells initially classified as “rare” were reannotated as tumor cells, highlighting the importance of using a relevant reference for cell annotation.

As suggested by the earlier differential gene expression analysis, we observed a decrease in AT2 and AT1 cells in tumors relative to NADJ samples and an increase in basal cells, particularly for treated and treatment-naïve LUSC tumors (Fig. 5).

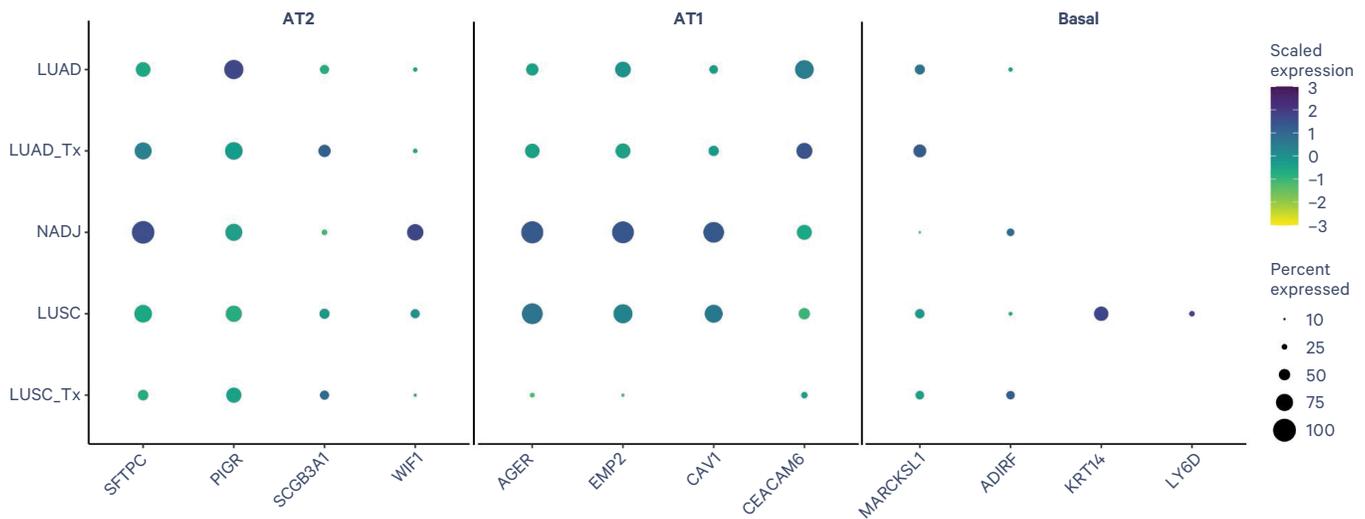
While differences in the distribution of cell populations were observed across sample types, more detailed comparisons also revealed differential gene expression in particular cell types across the biological groups. For example, while AT2 and AT1 cells were present in lower proportions in tumors relative to NADJ tissues, the tumor AT2 and AT1 cells show a deregulation of many canonical



**Figure 5. Breakdown of cell types in normal and lung cancer samples. A.** UMAP projection showing cell annotation. **B.** Distribution of different cell populations across the five biological groups.

cell markers (e.g., *AGER*, *EMP2*, and *SFTPC*; 8, 9, 11) compared to the same cell types from NADJ samples (Fig. 6). *WIF1*, a marker expressed in AT2 cells (14), was also downregulated outside of NADJ AT2 cells and has been found to induce apoptosis in NSCLC cells. Interestingly, club cell markers *PIGR* and *SCGB3A1* showed increased expression levels in tumor AT2 cells, and an “alveolar/club-like” signature has been previously described in lung adenocarcinoma (15). Further analysis showed that

expression of carcinoembryonic antigen-related cell adhesion molecule (*CEACAM6*) is low in NADJ AT2 and AT1 cells, but is elevated in AT2 (data not shown) and AT1 cells from treatment-naïve LUAD tissues. *CEACAM6* is a marker for epithelial cancers, and previous studies have suggested that it prevents anoikis, or cell death due to loss of cell adhesion. Patients with high *CEACAM6* expression have an overall poorer prognosis (13).



**Figure 6. Dot plot showing differential expression of cell-type markers in AT1, AT2, and basal cells.** Several AT1- and AT2-specific markers show decreased expression in the tumor samples whereas several basal cell markers show increased expression in tumors relative to NADJ samples. Tumor AT2 cells also showed an upregulation of club and secretory cell-type markers.

Basal cells also displayed a difference in cell-type proportion across the different biological groups, with an expanded population in the treated and treatment-naïve LUSC samples, relative to the NADJ tissues. However, a comparison of basal cell gene signatures across sample types revealed elevated expression levels of known basal cell markers (such as *ADIRF*) specifically in the treated LUSC tumor samples (Fig. 6). *KRT14*, a marker for transit amplifying parabasal cells, and *LY6D*, a marker for differentiating basal cells, were highly expressed in treatment-naïve LUSC samples. Interestingly, *MARCKSL1*, a marker for submucosal secretory cells (16), was found to be upregulated in both AT2 cells (data not shown) and LUAD basal cells.

*MARCKSL1* has previously been described as promoting epithelial-to-mesenchymal transition in lung adenocarcinoma cells.

Taken together, these data highlight the importance of using a single cell analysis approach when examining tumor subtypes considered to have different cells of origin. This is further supported by observations of gene deregulation in several cell populations across tumor subtypes that would otherwise be masked in a bulk analysis. Additionally, the complexities of tumor development and disease prognosis are underscored by the presence of more AT2 and AT1 cells in LUAD relative to LUSC, and more basal cells in LUSC and LUSC\_Tx relative to NADJ and LUAD tumors.

## Conclusion

Investigations into the cellular mechanisms at work behind cancer progression benefit from the use of complementary technologies. While pseudobulk differential gene expression analysis of human lung tissue revealed several known markers associated with poor tumor prognosis, single cell data was able to provide the resolution needed to determine which cell type was responsible for that expression pattern. Furthermore, comparison of the gene expression profiles belonging to specific cell types revealed deregulation of canonical cell-type markers present in known cell types across biological groups.

Cell typing atlases are typically based on profiles of healthy tissues, making it challenging to correctly annotate cell types that may be specific to certain disease states. Through manual reannotation of cell types found in the NSCLC dataset, we found that more than half of the 3% of cells that had been originally classified as rare cell types were reclassified as tumor cells. Moving forward, the ability to profile larger numbers

of cells from various disease states can improve cell-type classification and enable the construction of more complete references to help better understand the cellular dynamics that drive disease onset and progression.

In this Application Note, we have shown not only the importance of conducting comparative studies on different disease states, but also the utility of flexible single cell chemistry that eases the workload for these large-scale, multi-site experiments through sample batching. Using Single Cell Gene Expression Flex, we were able to fix and dissociate 32 human lung samples across three clinical stages, in batches, over the course of multiple days. Batched samples were then multiplexed to profile ~900K individual cells, representing five biological groups, in a single run. The resulting analysis provided novel insights into the complex biology that drives differential disease progression and patient outcome between two NSCLC subtypes.

## References

- Herbst R S, Morgensztern D & Boshoff C. The biology and management of non-small cell lung cancer. *Nature* 553: 446–454 (2018). <https://doi.org/10.1038/nature25183>
- Wang C, et al. The heterogeneous immune landscape between lung adenocarcinoma and squamous carcinoma revealed by single-cell RNA sequencing. *Signal transduction and targeted therapy*, 7(1): 289 (2022). <https://doi.org/10.1038/s41392-022-01130-8>
- Lambrechts D, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med* 24: 1277–1289 (2018). <http://doi.org/10.1038/s41591-018-0096-5>
- Wu F, et al. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat Commun* 12(1): 2540 (2021). <https://doi.org/10.1038/s41467-021-22801-0>
- Ferone G, et al. Cells of origin of lung cancers: lessons from mouse studies. *Genes Dev* 34(15-16): 1017–1032 (2020). <https://doi.org/10.1101/gad.338228.120>
- Hao Y, et al. Integrated analysis of multimodal single-cell data. *Cell* 184(13): 3573–3587.e29 (2021). <https://doi.org/10.1016/j.cell.2021.04.048>
- Sikkema L, et al. An integrated cell atlas of the lung in health and disease. *Nat Med* 29(6): 1563–1577 (2023). <https://doi.org/10.1038/s41591-023-02327-2>
- Wu S, et al. RAGE may act as a tumour suppressor to regulate lung cancer development. *Gene* 651: 86–93 (2018). <https://doi.org/10.1016/j.gene.2018.02.009>
- Ma Y, et al. Epithelial membrane protein 2 suppresses non-small cell lung cancer cell growth by inhibition of MAPK pathway. *Int J Mol Sci* 22(6): 2944 (2021). <https://doi.org/10.3390/ijms22062944>

10. Shang J, et al. Identification of lung adenocarcinoma specific dysregulated genes with diagnostic and prognostic value across 27 TCGA cancer types. *Oncotarget* 8(50): 87292–87306 (2017). <https://doi.org/10.18632/oncotarget.19823>
11. Xu F, Liu F, Zhao H, An G & Feng G. Prognostic significance of mucin antigen MUC1 in various human epithelial cancers: a meta-analysis. *Medicine*, 94(50): e2286 (2015). <https://doi.org/10.1097/MD.0000000000002286>
12. Kim E, et al. Overexpression of CEACAM6 activates Src-FAK signaling and inhibits anoikis, through homophilic interactions in lung adenocarcinomas. *Transl Oncol* 20: 101402 (2022). <https://doi.org/10.1016/j.tranon.2022.101402>
13. Luo X, et al. Wnt inhibitory factor-1-mediated autophagy inhibits Wnt/ $\beta$ -catenin signaling by downregulating dishevelled-2 expression in non-small cell lung cancer cells. *Int J Oncol* 53(2): 904–914 (2018). <https://doi.org/10.3892/ijo.2018.4442>
14. Bischoff P, et al. Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. *Oncogene* 40: 6748–6758 (2021). <https://doi.org/10.1038/s41388-021-02054-3>
15. Liang W, et al. MARCKSL1 promotes the proliferation, migration and invasion of lung adenocarcinoma cells. *Oncol Lett* 19(3): 2272–2280 (2020). <https://doi.org/10.3892/ol.2020.11313>

## Resources

Explore the Chromium Single Cell Gene Expression Flex data from these samples further by downloading the following datasets:

<https://www.10xgenomics.com/resources/datasets/aggregate-of-900k-human-non-small-cell-lung-cancer-and-normal-adjacent-cells-multiplexed-samples-16-probe-barcodes-1-standard>

## Contact us

[10xgenomics.com](https://www.10xgenomics.com) | [info@10xgenomics.com](mailto:info@10xgenomics.com)