Original Research Article

Single Cell RNA-Sequencing Reveals Metastatic and Therapeutic Signatures in Non-Small Cell Lung Cancer

Anthony Wang

Parkland High School, 2700 N Cedar Crest Blvd, Allentown, PA 18104, United States

ABSTRACT

Lung cancer is the leading cause of cancer death globally, with non-small cell lung cancer (NSCLC) accounting for most cases. This remains driven by its heterogeneity and metastatic potential. In this research, the primary objective is to understand the differential expression of metastasis-associated markers and differences in therapeutic potential, both of which are key to treatment outcomes. In this study, cell scoring was applied to a dataset of 20 treatment-naive NSCLC patients, which consisted of 14 adenocarcinoma (ADC) patients, 3 squamous cell carcinoma (SCC) patients, 1 combined small cell lung cancer (C-SCLC) patient, 1 patient with mixed adenocarcinoma and neuroendocrine carcinoma (MANEC), and 1 pulmonary chondroid hamartoma patient undergoing surgical resections. This raw dataset contained 9,001 cells and 24,873 genes. A narrative, literature-supported exploratory study was conducted in which genes expressed in fewer than 3 cells and cells with fewer than 100 genes were filtered out, resulting in 8,979 cells and 22,596 genes (Supplementary Figure S1). This allowed the researcher to determine the relative proportions of immune and cancerous cells among the 8,979 cells in total (Supplementary Figure S1). It enabled the researcher to assess the overall immune response in patients diagnosed with NSCLC. Next, immune cells were filtered out to focus on cancerous epithelial cells, resulting in a total of 3,324 cells. From these epithelial cells, known expression markers associated with NSCLC subtyping, multidrug resistance, apoptosis resistance, and altered cancer metabolism (GSE119911) were identified. In this analysis, the researcher demonstrated that the samples encompass diverse immune cell populations that play a crucial role in shaping cancer heterogeneity. Concepts of anti-apoptosis, multi-drug resistance, and metastatic potential are also explored using marker genes. Cellular subtypes and clusters defined by the selected markers, which may otherwise be missed, were characterized by the researcher using single-cell RNA sequencing. The observed results demonstrated the utility of this method for uncovering cellular diversity, which could be used for cancer outcome prediction and treatment decision-making.

Keywords: Non-Small Cell Lung Cancer; Cancer Heterogeneity; Apoptosis Evasion; Multi-Drug Resistance; Metastatic Potential; Mutations; Single-Cell RNA Sequencing

INTRODUCTION

NSCLC, or lung adenocarcinoma, is the most common form of lung cancer worldwide, accounting for 80-85 percent of all cases (1). There are three main subtypes of lung adenocarcinoma: ADC (40% of lung cancers), SCC (25-30% of lung cancers), and large cell carcinoma (10% of lung cancers). If metastasis occurs, most of them

Corresponding author: Anthony Wang, E-mail: anthonywang1215@gmail.com.

Copyright: © 2026 Anthony Wang. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Accepted May 29, 2026
https://doi.org/10.70251/HYJR2348.43329345

die after the 18th month of diagnosis. Another type of lung cancer is small-cell lung cancer, which occurs in a smaller proportion of the population. NSCLC occurs when abnormal cells form and multiply in the lungs. These cells are usually larger and grow more slowly than those in small-cell lung cancer. During metastasis, NSCLC can spread to other parts of the body, such as the adrenal glands, bones, brain, liver, lymph nodes, and skin. Specifically, in the advanced stage of this disease, more than 70% of the patients experience metastatic spread. Treatments can vary depending on the patients' conditions. For instance, if the cancer cells are confined to the lung, surgery might be the first-line treatment, in which a healthcare provider removes the tumor and the surrounding cells. Other forms of treatment that are used include chemotherapy (the use of drugs to attack lung cancer), immunotherapy (the use of certain drugs to boost the immune system so that it can recognize/destroy cancer cells), targeted therapy (the use of drugs to attack specific types of lung cancer), and radiation therapy (the use of X-rays to kill cancer by aiming them at lung cancer wherever it is on the body).

NSCLC usually arises from mutations in DNA that encode genes responsible for cell-cycle regulation. These mutations usually occur at cell division checkpoints, mechanisms that ensure cells divide in a controlled manner (2). The three main cell cycle checkpoints are the G1/S checkpoint, the G2/M checkpoint, and the spindle checkpoint.

In the G1/S checkpoint, regulation of the cell cycle occurs through the hypophosphorylation / phosphorylation of the tumor-suppressing retinoblastoma protein (Rb) along with its interaction with the E2 factor (E2F) family of transcription factors (2). When any of these processes involve gene deletion, overexpression, or point mutations, they can affect the balance between phosphorylation and hypophosphorylation of the Rb protein, which in turn can lead to increased cell proliferation. At the G2/M checkpoint, the transition into mitosis often depends on the accumulation of a key protein called Cyclin B. The G2/M checkpoint is often responsible for checking problems in DNA replication or damage. In situations of DNA damage/stress, a regulatory factor such as protein 53 (p53) will enforce a G2 arrest until the damage is repaired.

Cancerous cells, despite abnormalities within their DNA damage and replication machinery, continue to divide, leading to further mutations and genomic instability. In the spindle checkpoint, a quality-control mechanism ensures that chromosomes are correctly distributed and specifically attached to the mitotic spindle. This should result in cells receiving equal numbers of chromosomes and prevent any chromosome missegregation/misalignment. A common feature of cancer cells, therefore, is a faulty chromosome number and partitioning, which can lead to aneuploidy. Therefore, due to the loss of control of cell division, many cells will divide rapidly, leading to cancer heterogeneity.

Cancer heterogeneity can be interpatient, intratumoral, and intertumoral (3). Intratumoral heterogeneity refers to genetic and phenotypic diversity within a single tumor and is of great importance for treatment resistance, disease trajectory, metastasis potential, and disease recurrence. Subpopulations of tumors may differ from one another in cell morphology, genetic makeup, metabolism, and proliferation rate, among other characteristics. In this case, genetic mutations can drive cancer cell heterogeneity by accumulating over cell division cycles. Cells within a cancer can also undergo even further changes, such as environmental selection and pressure. For example, tumor cells can interact with either immune cells, the extracellular matrix, or other cells. This can further lead to the differentiation of cell subpopulations within a tumor. This ultimately increases genetic diversity in the population, allowing the tumor to resist drugs or other treatments. The resistant population can thus continue to divide and grow rapidly over time. This heterogeneity present in cancer can often have big implications in clinical decision-making. Some clinical decisions often use single-core biopsy samples, which are taken from cells, tissues, or bodily fluids, to diagnose advanced cancer. This can lead to several problems because the various tumoral subpopulations are often not detected in the process. It is often hard to tell how the tumor subpopulation will diversify and expand throughout time. Often, a combinatorial approach can be helpful in addressing this by integrating various drugs with multi-omics to target multiple pathways and eliminate the cancer cell population as much as possible. Multi-omics can be used to analyze tumor samples, characterizing differences between cells within the same tumor (intramolecular heterogeneity) or molecular variation across different tumor samples (intermolecular heterogeneity). It can also be used to identify the growth and evolution of tumor subpopulations over time.

Furthermore, cancer cells can accumulate mutations that enable their migration to various organs and tissues, further from the original cancer site—a process called metastasis. Metastatic potential refers to the capacity of

a tumor, often malignant, to spread from the primary site to other parts of the body through the bloodstream or lymphatic system (4).

In this study, the researcher analyzed a single-cell RNA sequencing dataset from patients with NSCLC across stages I-IV and sought to characterize cell heterogeneity using three main factors: metastatic potential, multidrug resistance (MDR), and apoptosis evasion markers. Through a literature review, the researcher identified specific markers associated with each of the factors above. Hence, this work presents an exploratory study of a publicly available dataset, combined with an extensive review of differentially expressed genes identified in cell subpopulations, characterizing alternative trajectories observed in heterogeneous NSCLC.

METHODS AND MATERIALS

The dataset used originated from a single-cell RNA sequencing study of 20 treatment-naive NSCLC patients, available on the publicly available Gene Expression Omnibus (GEO) platform under accession GSE119911 (5). The samples originated from a diverse population of NSCLC patients: 14 ADC patients, 3 SCC patients, 1 C-SCLC patient, 1 patient with MANEC, and 1 patient with pulmonary chondroid hamartoma undergoing surgical resection.

The raw dataset, comprising 9,001 cells and 24,873 genes, was analyzed using the scanpy package. Genes expressed in fewer than 3 cells and cells with less than 100 genes were filtered, resulting in 8,979 cells and 22,596 genes (Supplementary Figure S1). The dataset was then normalized and log-transformed to remove technical noise and stabilize the data structure, so that downstream analyses reflect true biological differences between cells rather than artifacts of sequencing depth. The top 2,000 most variable genes were selected for downstream analysis (Supplementary Figure S2).

Following the initial filtering, a principal component analysis (PCA) was applied to further reduce the dataset's dimensionality, and the top 50 principal components were used for a cell-nearest-neighbors search with the default 15 neighbors and Euclidean distance. Using this graph, the researcher carried out Leiden clustering to identify discrete subgroups for manual cell annotation. The clustering quality was assessed across different granularity settings, and a resolution of 0.1 was selected for its good visual balance between cluster separation and granularity (Supplementary Figure S3). Quality control measurements were also re-assessed to ensure that the detected data-structure results reflect genuine biological signals (Supplementary Figure S3). Automated cell annotation was carried out using the Immune_All_Low model from the celltypist package (6). The model allows the researcher to distinguish between cancerous cells and their immune microenvironment in detail. To better characterize the properties of the tumor cells, cells annotated as epithelial were selected for further analysis, and the neighborhood search and Leiden clustering were repeated. Manual cell annotations were performed by calculating the Mann-Whitney U-test for differentially expressed genes within each cluster versus all others.

Metastasis is a major challenge in NSCLC mortality. Therefore, cells were further scored for their metastatic potential using five markers, which have been described as strong predictors - phosphoinositide-3-kinase regulatory subunit 1 (*PIK3R1),* Family Tyrosine Kinase *(FYN),* Thrombospondin-1 *(THBS1),* Spectrin Alpha, Non-Erythrocytic 1 *(SPTAN1),* Secreted Phosphoprotein 1 *(SPP1)* (7). The *PIK3R1* gene encodes 5 protein domains. In cancerous cells, including those in NSCLC, mutations have been detected across all 5 protein domains. Deletion mutations are common in the *PIK3R1* gene, thereby promoting tumor development through uncontrolled activation of the PI3K signaling pathway (8). The uncontrolled activation of the PI3K signaling pathway can trigger Epithelial-mesenchymal transition (EMT), which causes cancer cells to acquire migratory features and detach, spreading across various tissues/organs throughout the body. The *FYN* gene is part of the Src family protein kinases (SFKs) and is located on chromosome 6 (9). This gene can significantly contribute toward NSCLC through cell proliferation and inhibition of apoptosis (10). It promotes cell proliferation by targeting microRNA-125a-3p, which induces cell capture, thereby inhibiting cell proliferation (11). It promotes inhibition of apoptosis by blocking ERK1/2, reducing caspase activation, and interfering with pathways such as the Akt/mTOR/S6K axis, ultimately leading to cell survival. *FYN* increases the expression of EMT transcription factors and decreases the expression of epithelial cells (9). This can cause cancerous cells to spread throughout the body by traversing the circulatory and lymphatic systems of the body. The *THBS1* gene was found to be lowly expressed in NSCLC (12). It is located on chromosome 15 in humans. Low expression of this gene can inhibit tumor suppression and cell migration, ultimately leading to a poor prognosis (12). Inhibition

of the *THBS1* gene can promote metastasis in NSCLC by suppressing immune responses, including T cells, altering the tumor microenvironment to favor cancer cells, and inhibiting angiogenesis (13). The *SPTAN1* gene, located on chromosome 9, plays an important role in maintaining the stability of the plasma membrane/organelles, thereby ensuring polarity and cell stability (14). It also plays minor roles in DNA repair, cell adhesion, and cell cycle control (14). Overexpression or underexpression of this gene can disrupt the functions listed above and increase the risk of cancer development. *SPTAN1* was found to be expressed at lower levels in lung cancer due to suppression by microRNA-128-3p, which affects its ability to repair DNA. Therefore, this has implications for mutations in the genes that can ultimately lead to cancer progression (15). In addition, low expression levels of the *SPTAN1* gene can weaken cell-to-cell adhesion, which in turn can allow tumor cells to detach and spread throughout the body through the cardiovascular and lymphatic systems (16). The *SPP1* (secreted phosphoprotein 1) gene, located on the 4th chromosome, is a multifunctional secreted phosphorylated glycoprotein (17). Increased SPP1 expression has been associated with poor prognosis in various cancers, including NSCLC (17). *SPP1* contributed to anti-cancer drug resistance in NSCLC (17). The *SPP1* gene can bind to cell-surface receptors such as CD44 and integrins, thereby boosting cell motility, which in turn can increase invasion. The score was overlaid on the (Uniform Manifold Approximation and Projection) UMAP plot to distinguish clusters with high potential for establishing metastasis.

Furthermore, NSCLC has been shown to have a high potential to develop resistance to multiple therapies (18). This is known as multi-drug resistance (MDR). Two markers largely responsible for MDR in NSCLC: ATP-Binding Cassette Subfamily B Member 1 (*ABCB1)* and ATP-Binding Cassette Subfamily B Member 2 (*ABCG2)* were identified. *ABCB1* marker refers to the gene (*MDR 1*) and its protein product, P-glycoprotein (P-gp). Overexpression of these genes due to genetic mutation has been identified in a wide range of multidrug-resistant cancers. *ABCG2* is a significant marker for stem cells and cancer stem cells and is a crucial efflux pump. It seems to have a particularly significant impact on the survival of patients with lung cancer and on immunotherapy response, particularly regarding immune cell infiltration. Both of these markers have been identified widely in multidrug-resistant cancer but have had a significant impact on survival. The percentage of cancerous cells expressing one or both of those markers was identified. Cell clusters with higher or lower proportions of MDR markers were also identified.

NSCLC has also been shown to have a high potential to evade apoptosis. The researcher therefore investigated the expression of the apoptosis evasion markers B-cell lymphoma 2 (*BCL2)*, Survivin (*BIRC5)*, and Myeloid Cell Leukemia (*MCL1)* (19). The BCL2 and MCL1 proteins bind to and neutralize pro-apoptotic proteins such as BCL-2-associated X protein (BAX), BCL-2 homologous antagonist/killer (BAK), BCL-2-like protein 11 (BIM), and p53 upregulated modulator of apoptosis (PUMA), which block the activation of executioner caspases, thus keeping the cell alive (20). Furthermore, Survivin promotes tumor growth and cell division. It does this by interfering with caspase-9, thereby preventing entry into the intrinsic pathway of apoptosis by binding to Smac/DIABLO. MCL1 has also been implicated in resistance to chemotherapeutic agents, pointing to the multiple functions of these proteins in cancer pathways (21).

RESULTS

Histologic subtype cancer cells of NSCLC - adenocarcinoma, squamous cell carcinoma, neuroendocrine carcinoma

There are three main subtypes of NSCLC: ADC, SCC, and neuroendocrine carcinoma (NEC). As shown in the heat maps in Figure 1, the researcher clearly identified ADC as the most common NSCLC subtype, with 61.4% of epithelial cells expressing the three markers: Napsin A Aspartic Peptidase (*NAPSA),* Keratin 7 *(KRT7), and* NK2 Homeobox 1 (*NKX2-1*) (Figure 1A and Supplementary Figure S4). The second most common subtype of NSCLC was SCC, with 24.7% of the epithelial cells expressing the three markers: Tumor Protein P63 (*TP63),* Keratin 5 *(KRT5), and* Keratin 6A *(KRT6A)* (Figure 1B and Supplementary Figure S4). The least common subtype of NSCLC was NEC, with 13.1% of the epithelial cells expressing the three markers: Synaptophysin (*SYP),* Neural Cell Adhesion Molecule 1 *(NCAM1),* and Chromogranin B *(CHGB)* (Figure 1C and Supplementary Figure S4). The researcher next sought to explore the distribution of these markers across clusters (Figure 1). All clusters showed high expression of the ADC markers, with only clusters 0, 3, and 8 showing lower expression. Markers of SCC were more clearly defined across clusters, with cluster 12 showing high *KRT6A* expression. Similarly, clusters 4 and 8 were strongly defined by *CHGB* expression.

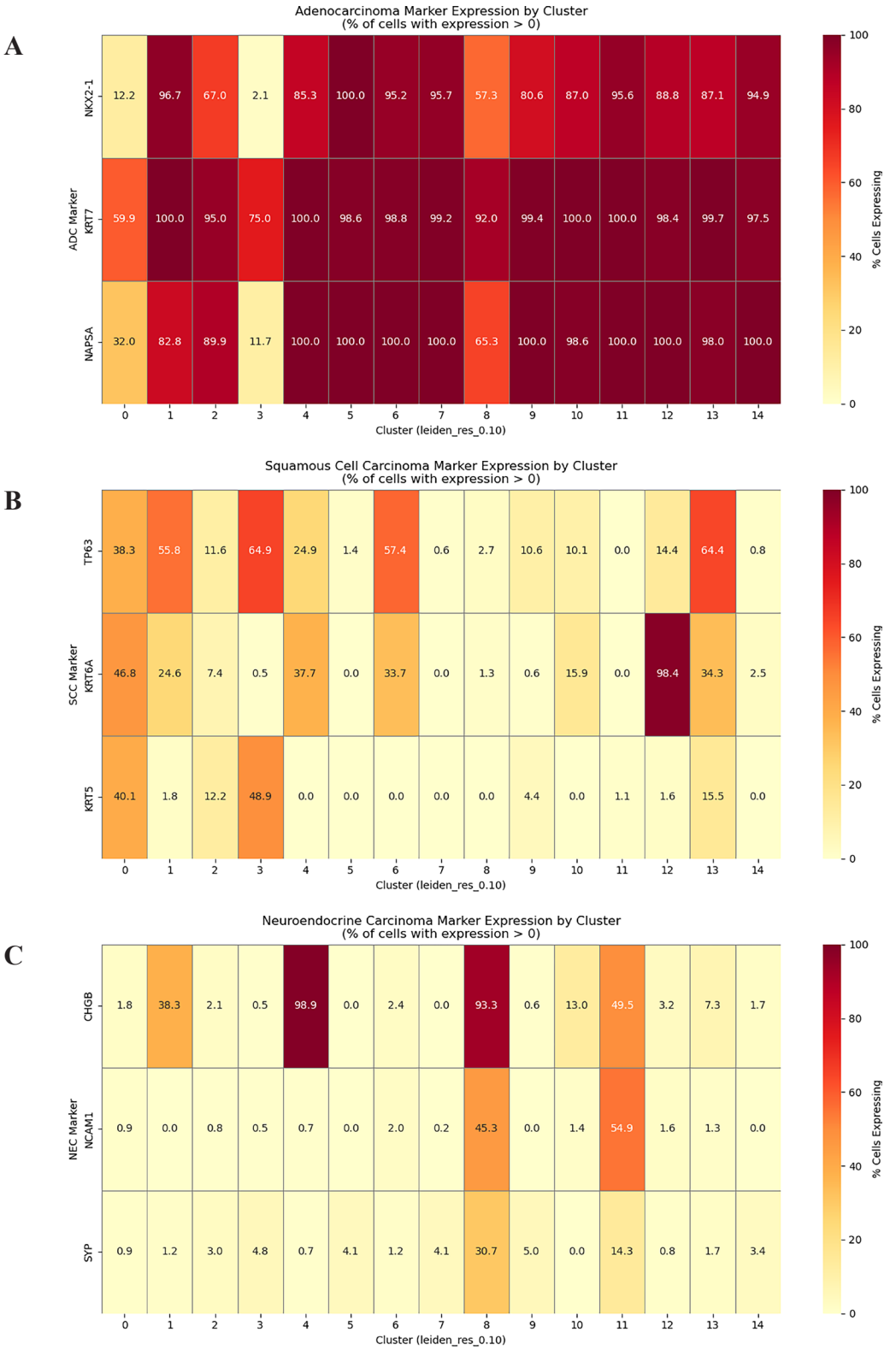


Figure 1. *A) Percentage of cells that express the three markers commonly found in ADC based on each cluster. B) Percentage of cells that express the three markers commonly found in SCC based on each cluster. C) Percentage of cells that express the three markers commonly found in NEC based on each cluster.*

Cell type annotations

Cell heterogeneity largely underlies cancer heterogeneity, which can be described as differences among cancer cells within a tumor (intratumoral heterogeneity) or between tumors from different patients (intertumoral heterogeneity). In this analysis, the researcher used marker genes to annotate cell types within the NSCLC samples. The researcher identified the immune composition inside these cancers (Figure 2). The presence of the varying immune cells defines immune responses, further shaping cancer cell heterogeneity (Figure 2). The identified immune cells can be divided into two categories: innate immunity and adaptive immunity. The innate immune cells represented in Figure 2 include CD16+ natural killer (NK) cells, alveolar macrophages, macrophages, DCs, mast cells, and non-classical monocytes. The adaptive immune cells shown in Figure 2 include the following: double-positive thymocytes, mast cells, plasma cells, regulatory T cells, rcm/naive helper T cells, tem/trm cytotoxic T cells, and pDC cells. The non-immune cells include epithelial cells and plasma cells.

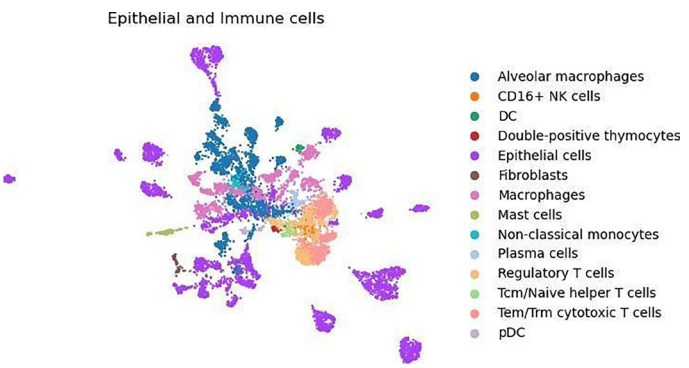


Figure 2. *UMAP of cell types, including cancer-associated fibroblasts and immune cells, based on annotation using cell markers from CellTypist.*

When cancer cells emerge, the innate immune system is usually the first to be activated, as it can respond to a broad range of mutated cells. It is the first line of defense. In many cases, a more specific adaptive immune system is required, with mechanisms developed to target specific cancer antigens. Once exposed to these antigens, immune cells in the adaptive immune system develop immunological memory, so a secondary exposure is likely to elicit a faster response. The innate cells and their respective functions are explained in the following examples (22). Macrophages' function is associated with cancer cell clearance through phagocytosis, as they engulf particles and transport them to the extracellular matrix, where lysozymes break them down. Regulatory T cells are a component of the adaptive immune system and do not directly attack cancerous cells; instead, they direct other immune cells, such as cytotoxic T cells, against them. NK cells are a type of white blood cell, specifically lymphocytes. They can directly destroy these cancerous cells and also communicate with other immune cells by releasing cytokines to fight them off. DC cells act as antigen-presenting cells by capturing cancer cells, processing their antigens, and presenting these antigens to T cells to activate adaptive immune responses. Mast cells release chemicals like histamine, which can dilate blood vessels, bringing in other immune cells to directly target cancer cells. Non-classical monocytes are the first line of defense because they are often anti-inflammatory, thereby maintaining vascular homeostasis. They recognize the cancerous cells and recruit other immune cells to target them.

The adaptive cells and their respective functions are explained in the following examples. Double-positive thymocytes are immature T cells that express both CD4 and CD8. Once they undergo positive or negative selection, they can mature into either CD4 or CD8 receptors. Plasma cells are B lymphocytes that secrete antibodies (23). They are mainly responsible for humoral immunity. These B-lymphocytes, once exposed to a particular cancer antigen, can develop immunological memory and respond to that antigen much faster upon secondary exposure. Regulatory T cells are a component of the adaptive immune system and do not directly attack cancerous cells; instead, they direct other immune cells, such as cytotoxic T cells and B cells, to them. Regulatory T cells ensure that various immune cells do not directly attack self-antigens; only antigens from cancerous cells are targeted. Naïve helper T cells are CD4 receptor-expressing cells that developed in the thymus but have not yet encountered a specific antigen. Once they encounter a specific antigen, they differentiate into effector cells, which, in turn, can coordinate various immune responses. Tem/trm cytotoxic T cells are long-lived immune cells that provide varying levels of protection against pathogens and tumors. TEM cells (effector memory) circulate throughout the blood and tissues, providing fast and rapid action. TRM cells (tissue-resident memory T cells) reside in specific tissues and provide a rapid, localized immune response. pDC cells fight cancer by activating other immune cells, such as T and NK cells. Because distinct immune cells target

different cancer cells, this directly reflects a complex heterogeneous cellular environment in which some cancer cells are often eliminated. In contrast, others become resistant and can divide and multiply.

Cell clustering and differential expression analysis

Following the identification of immune cells and cancer cells annotated as epithelial cells, the immune cells were filtered out to further focus the analysis on the properties of the 3,324 cancerous cells. Unsupervised clustering at a Leiden resolution of 0.10 was performed, yielding 15 clusters (Figure 3A). Using differential expression analysis, the researcher identified markers that define these clusters, summarised in Figure 3B. All clusters, except cluster 0, were clearly defined by their marker genes. Cluster 0 showed a mix of cell types, indicating it is the bin cluster identified by the unsupervised Leiden algorithm.

As shown in Figure 3B, the metabolic plasticity of NSCLC is a primary driver of its resilience to physiological stress and therapeutic intervention, as evidenced by Clusters 4, 5, and 7. Cluster 7 functions as a bioenergetic powerhouse, characterized by the overexpression of Cytochrome c Oxidase Subunit 6C (*COX6C)*, which catalyzes the terminal step of the electron transport chain, thereby maximizing ATP production via oxidative phosphorylation (24). This metabolic efficiency is complemented by Lysosomal Associated Protein Transmembrane 4 Beta (*LAPTM4B*), a lysosomal protein that activates the Nuclear Factor Erythroid 2-Related Factor 2 (*NRF2*) signaling pathway to bolster antioxidant defenses and inhibit apoptosis.

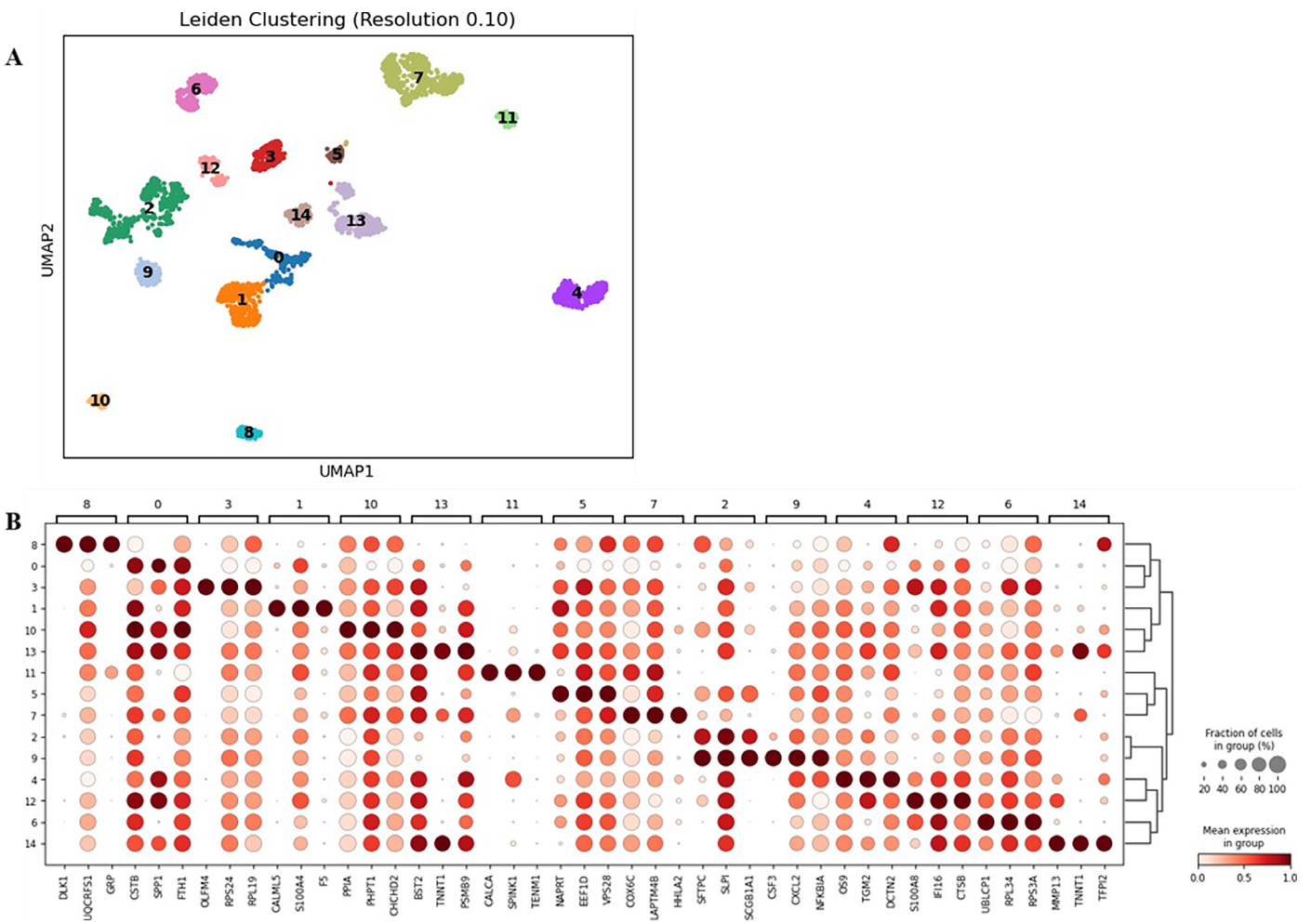


Figure 3. *Unsupervised clustering and differential expression analysis of the NSCLC samples. A) Leiden clustering identified 15 distinct clusters at a resolution of 0.1. B) The top cluster-specific differentially expressed genes.*

Similarly, Cluster 5 maintains essential metabolic flux via Nicotinate Phosphoribosyltransferase (*NAPRT)*, which preserves NAD+ levels required for the activity of survival-linked enzymes such as ADP-ribose transferases (25). This theme extends to hypoxia adaptation in Cluster 4, where Osteosarcoma Amplified 9 (*OS9)* facilitates survival in oxygen-depleted environments by binding and degrading HIF-1 subunits. Furthermore, Cluster 4 highlights a critical bridge between metabolism and chemoresistance: the expression of Transglutaminase 2 (*TGM2)* is a determinant of cisplatin sensitivity, with non-methylated, overexpressed states correlating with significant drug resistance (26). Collectively, these clusters define a phenotype capable of navigating the harsh metabolic demands of the tumor microenvironment.

The transition from localized malignancy to systemic metastasis is orchestrated through the structural and phenotypic shifts identified in Clusters 3, 12, and 14. Cluster 14 is prominently defined by Metallopeptidase 13 (*MMP13)* and Troponin T (*TNNT1)*; the former remodels the extracellular matrix (ECM) to facilitate bone marrow microinvolvement, while the latter overactivates the Wnt/β-catenin pathway to drive the EMT (27). This structural interference is mirrored in Cluster 12, where Cathepsin B (*CTSB*) protease activity degrades collagen and elastin, providing a physical pathway for invasion (28). At the same time, S100 Calcium Binding Protein A8 (*S100A8)* enhances metastatic potential through calcium-mediated signaling (29). Cluster 3 further augments this invasive phenotype via Ribosomal Protein L19 (*RPL19)* overexpression, which triggers the erbB-2 signaling pathway, a known facilitator of aggressive metastatic spread (30). By integrating internal signaling shifts with active ECM degradation, these clusters collectively transform the tumor's physical architecture, thereby supporting systemic dissemination and poor clinical outcomes.

Underlying the rapid expansion of NSCLC is a comprehensive loss of regulatory checkpoints, as demonstrated in Clusters 2, 8, 10, and 12. Cluster 8 exploits non-canonical Notch signaling via Delta-Like Non-Canonical Notch Ligand (*DLK1)*, which serves as both a predictor of recurrence and a target for radioimmunotherapy, alongside Gastrin-Releasing Peptide (GRP), which bypasses apoptotic triggers to maintain proliferation (31). In Cluster 10, the co-amplification of Epidermal Growth Factor Receptor (EGFR) with Coiled-coil-Helix-Coiled-Coil-Helix Domain Containing 2 (*CHCHD2)* results in hyperactive growth signaling, further accelerated by Phosphohistidine Phosphatase 1 (*PHPT1)* activity (32). The collapse of tumor suppression is most pronounced in Cluster 12, where the downregulation of Interferon Gamma Inducible Protein 16 (*IFI16)* suppresses p53, effectively bypassing G2-stage cell-cycle arrest (33). This loss of homeostatic control is complemented by findings in Cluster 2, in which the downregulation of the alveolar marker Surfactant Protein C (*SFTPC)* removes a natural brake on the Wnt/β-catenin pathway (34). Together, these clusters illustrate how the hijacking of developmental pathways and the silencing of tumor suppressors create an environment of unchecked cellular growth.

The clinical trajectory of NSCLC is deeply influenced by the establishment of an immunosuppressive niche, as seen in Clusters 1, 9, 11, and 13. Cluster 9 facilitates this through the secretion of Colony Stimulating Factor 3 (*CSF3)* (G-CSF) and C-X-C Motif Chemokine Ligand 2 (*CXCL2)*, which recruit neutrophils and white blood cells to cultivate a pro-inflammatory yet immunosuppressive milieu via ERK/MAPK and PI3K/AKT signaling (35). While Cluster 1 retains some defensive capacity through *CALML5* and its co-localization with cytotoxic immune activity, this is often overwhelmed by the pro-metastatic influence of S100 Calcium Binding Protein A4 (*S100A4)*. Strategic evasion is further refined in Clusters 7 and 11, where markers such as HERV-H LTR-Associating Protein 2 (*HHLA2)* and Calcitonin-Related Polypeptide Alpha (*CALCA)* create an immune shield, preventing T-cell infiltration and neutralizing natural killer (NK) cell responses by inhibiting TNF-α production. However, Cluster 13 presents a potential therapeutic window: the expression of Proteasome 20S Subunit Beta 9 (*PSMB9)* is essential for MHC Class I antigen presentation, making it a crucial biomarker to guide T cell-based immunotherapy and overcome the immune siloing characteristic of these clusters (36).

In conclusion, this unsupervised cell annotation of NSCLC reveals a highly heterogeneous landscape characterized by metabolic adaptation and immune subversion. A prominent trend across clusters is the hijacking of developmental pathways, such as Wnt/β-catenin and Notch, which transition from normal pulmonary maintenance to driving EMT and systemic metastasis. Furthermore, the data highlight a clear shift toward an immunosuppressive microenvironment, in which markers such as *HHLA2*, Secretory Leukocyte Protease Inhibitor (*SLPI)*, and *CALCA* shield malignant cells from cytotoxic activity. These findings demonstrate the need to move beyond broad treatments toward

personalized strategies that simultaneously target metabolic vulnerabilities, such as NAD+ biosynthesis, and restore immune visibility through biomarkers like Proteasome 20S Subunit Beta 9 (*PSMB9).*

Supervised annotation of clusters with metastatic potential

Metastasis occurs when cancer cells spread from the original site to other areas of the body by traveling through the lymphatic or circulatory systems (1). While a majority of these cancer cells end up dying as they traverse through the vessels, some can make their way through. Once they arrive at the new site, they continue to grow and divide in favorable microenvironments, thus producing a metastatic tumor. These metastatic tumors can also be referred to as malignant (1). For NSCLC, cancerous cells often form a primary site in the lung tissue. Cancer cells can then spread toward metastatic secondary sites, which could include the adrenal gland, bone, brain, and liver (37). Using metastatic markers *PIK3R1, FYN, THBS1, SPTAN1, and SPP1*, the researcher calculated a metastasis score and overlaid it on the UMAP of cells in Figure 4. Clusters 4, 0, and 10 showed the highest metastatic score, suggesting that cells in these clusters are evolving toward a high potential to establish metastasis in secondary sites. Meanwhile, clusters with the lowest metastatic potential included clusters 6 and 9. This meant that the cancer cells were less likely to grow and spread to other parts of the body.

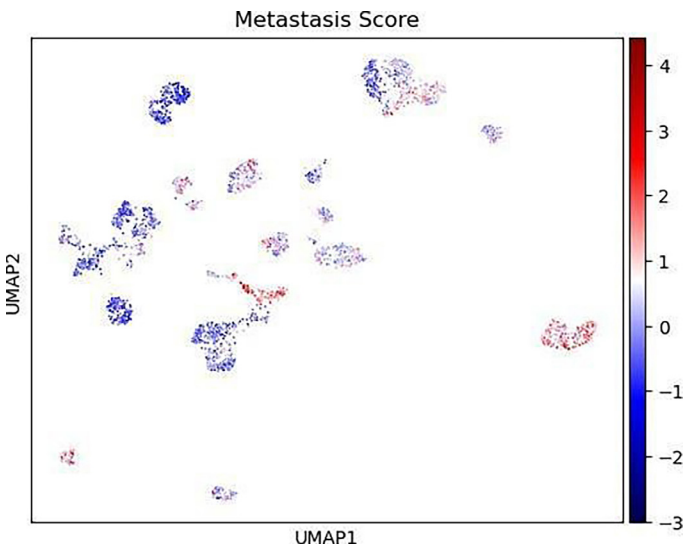


Figure 4. *UMAP of metastasis scores overlaid on the 3,324 cancerous cells.*

Clusters with multi-drug resistance potential

Next, a detailed literature review was conducted to identify previously characterized MDR markers, including *ABCB1* and *ABCG2*. The *ABCB1* marker refers to the gene (MDR 1) and its protein product, P-glycoprotein (P-gp). Overexpression of these genes due to genetic mutation has been identified in a wide range of multidrug-resistant cancers. Furthermore, *ABCG2* is a crucial efflux pump and a significant marker of stem cells and cancer stem cells. It has a particularly significant impact on the survival of patients with lung cancer and on the effect of immunotherapy related to immune cell infiltration. Using the dimensionality-reduced data from Figure 5B for visual reference, along with the bar chart and heatmap from Figure 5A, the researcher found that, of the 3,324 cancerous cells, approximately 35 highly expressed the *ABCB1* marker (1.05% of all cancer cells). In comparison, 173 cells highly expressed the *ABCG2* marker (5.20% of total cancer cells). In addition, 3 cells contained both markers (0.09% of total cancer cells). When analyzed by cluster in Figure 5A, clusters 7 and 13 were identified as having the highest proportions of MDR+ cells, with cluster 7 containing around 17 percent of MDR+ cells within its cancer population and cluster 13 containing around 12 percent of MDR+ cells within its cancer population.

Clusters with apoptosis evasion potential

The researcher next sought to study markers of apoptosis evasion in NSCLC. Markers of this characteristic were identified from the 3,324 cancerous cells and included *BCL2*, *BIRC5* (Survivin), and *MCL1.* Apoptosis is a crucial genetic mechanism to regulate cell death. It is also responsible for maintaining normal development and homeostasis in various human cells. Overall, there are two major apoptotic pathways: exogenous and endogenous (38). In an exogenous pathway, the stimulus is activated outside the cell, causing apoptosis, whereas in an endogenous pathway, internal mechanisms within the cell activate apoptosis. Evasion of either of these apoptosis pathways can lead to the development of cancer cells, a process called apoptosis evasion. The BCL2 protein (B-cell lymphoma 2) contains both of the pro and anti-apoptotic activities (38). The anti-apoptotic activities of this BCL2 protein can allow cells to permanently evade apoptosis, thereby enabling malignant clone formation. Within the subfamily of anti-apoptotic proteins, the members include BCL-2, BCL-XL, MCL-1, BCL-W, and BFL-1 (38). This family of

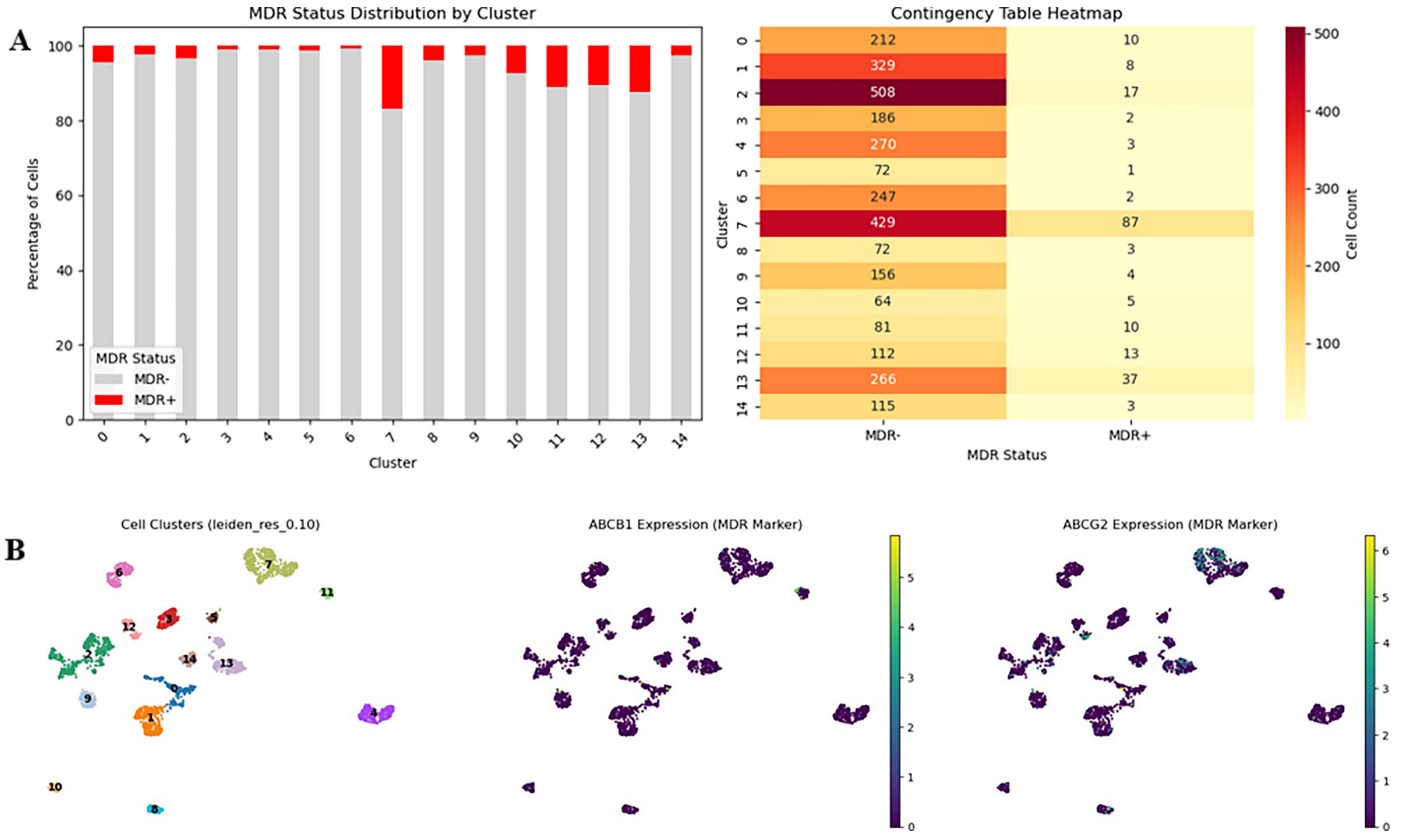


Figure 5. *Investigation of the MDR-linked markers in NSCLC. A) MDR (+ or -) status distributed across 15 cell clusters given through a visual representation and with a table. B) UMAP plot overlaid with each multi-drug resistance marker across the clusters.*

proteins generally blocks apoptosis by inhibiting caspase activity and altering mitochondrial membrane permeability. Survivin is the smallest inhibitor of the apoptosis protein family. Yet it has a unique structural identity. It is overexpressed in malignant tissue cells, disrupts intracellular signaling during apoptosis, and enables various cancer cells to adapt to unfavorable conditions (38). MCL1 is a protein that is part of the larger subfamily of anti-apoptotic BCL2 family members. They largely inhibit apoptosis by binding and inactivating pro-apoptotic proteins such as BAK/BAX and by regulating the mitochondrial pathway of apoptosis (38). Like the survivin marker, this marker is often overexpressed in various cancer cells. Because of this overexpression, it can lead to tumor progression, which ultimately leads to cancer cell heterogeneity as more tumor cells divide and become exposed to different subenvironments in the body. As shown in Figure 6A, the *BCL2* marker was expressed within 8.51 percent of the cancer cell population. The *BIRC5* marker was detected in 36.91% of the cancer cell population. The *MCL1* marker was detected in 90.76% of the cancer cell population. Cell clusters such as 1 and 3 had a high apoptosis evasion score, indicating rapid division and growth of tumor cells. On the other hand, a low apoptosis evasion score was observed for cell clusters such as 6 and 9, indicating few cells expressing mechanisms that limit apoptosis. The *MCL1* marker has the highest mean expression in Clusters 7 and 14. The *BIRC5* marker has the highest mean expression in Clusters 8 and 3. The *BCL2* marker has the highest mean expression in Clusters 8 and 11 (Figure 6B). Overall, this shows that a high percentage of the 3,324 cancerous cells express mechanisms that limit apoptosis, allowing rapid division and growth of tumor cells, which in turn could lead to heterogeneity.

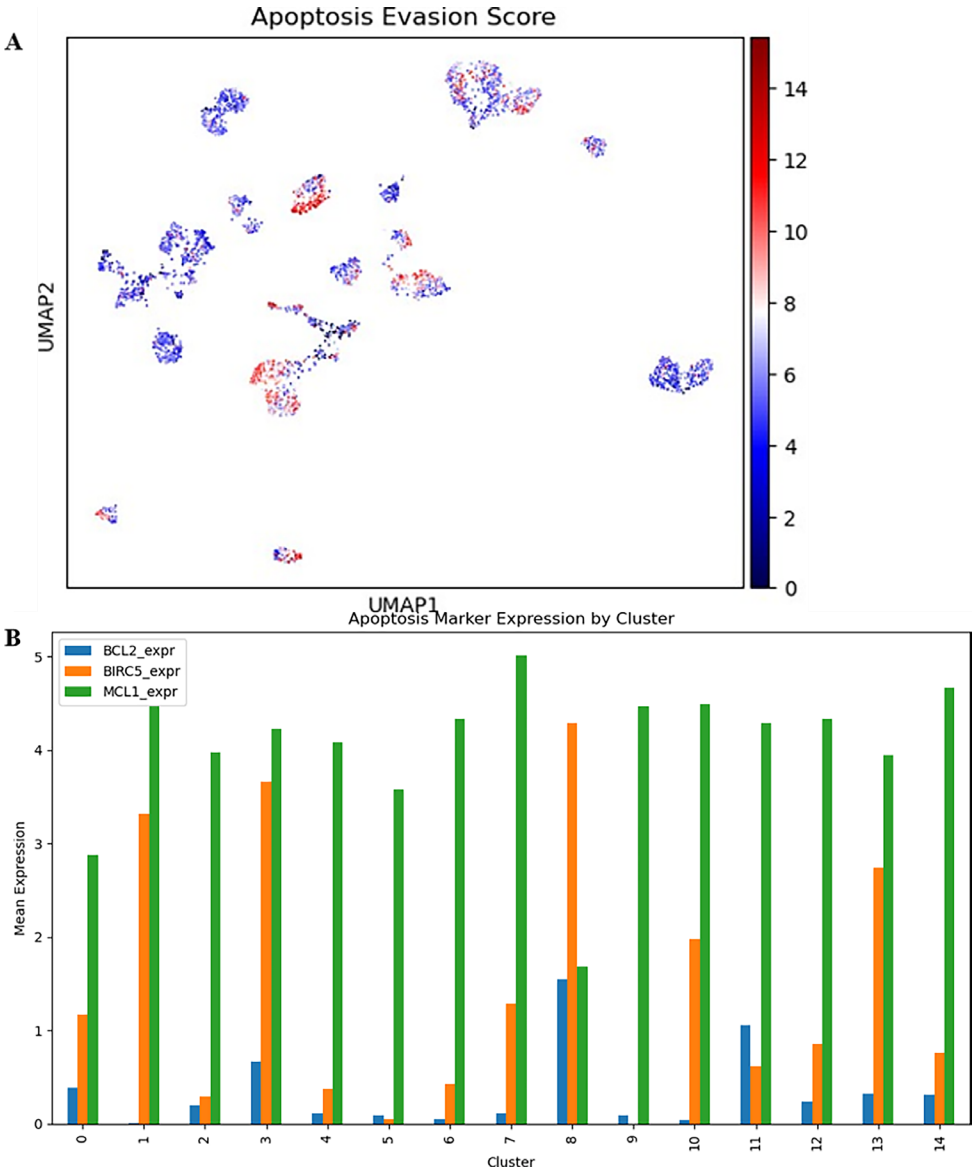


Figure 6. *Investigation of anti-apoptosis markers in NSCLC. A) Apoptosis Evasion score given from a scale of 0 to 14 and a color gradient (blue to red) representation. B) Apoptosis Marker Mean Expression (0 to 5) distributed across 15 clusters.*

DISCUSSION

In this analysis, the researcher analyzed a sample of NSCLC cases and identified cancer heterogeneity associated with metastatic potential, multidrug resistance, or evasion of apoptosis. The researcher identified distinct clusters associated with overexpression of metastasis-linked markers, suggesting that cell populations from the primary NSCLC site could evolve to form metastases at secondary sites. The studied markers *PIK3R1, FYN, THBS1, SPTAN1, and SPP1* have been described in multiple NSCLC studies. The *PIK3R1* gene has been linked to poor survival in patients, primarily due to loss of expression, which can induce carcinogenesis. Studies have shown that it was the 11th most frequently mutated gene among 4,429 tumors across 20 tumor types. An increase in *THBS1* gene expression has been linked to poor survival, with studies largely conducted in mice indicating that it reduced survival from 160 days to 149 days. Thus, suggesting the role of the *THBS1* gene in suppressing tumor growth. Studies have shown that inhibiting the *FYN* gene increases survival rates in

patients. Thus, innovative medicines have been developed to target the protein kinase encoded by the gene. Higher *SPTAN1 gene* expression has been linked to longer patient survival. It has been currently used as an important biomarker to refine therapeutic conditions across various cancers. Finally, overexpression of the *SPP1* gene has been associated with unfavorable survival, and its expression is a reliable prognostic indicator in NSCLC, thus making it a potential immunotherapy target.

Generally, clusters with the highest metastatic potential indicate that the various cancerous cells within each cluster have a high capacity for primary tumor cells to traverse the circulatory and lymphatic systems and establish malignant tumor sites at secondary sites throughout the body. Regarding multidrug resistance, the researcher examined two markers (*ABCB1* and *ABCG2*) that significantly affected the survival of patients with lung cancer, as well as the effects of various treatments on cancerous lung cells. Overall, around 208 cells expressed one of the two markers: 173 *ABCG2* and 35 *ABCB1*. Additionally, 3 cells expressed both markers.

The researcher was also able to identify cluster-specific cells, which contained the highest proportion of MDR markers; they were clusters 7 and 13. Finally, regarding apoptosis evasion, the researcher examined three markers (*BCL2*, *IREB5* (Survivin), and *MCL1*). The three markers can affect the intracellular process of apoptosis (endogenous pathway) or inhibit the activity of enzymes such as caspases. This will then allow cancerous cells to continue dividing and multiplying rapidly throughout the body without undergoing cell death. Overall, the *BCL2* marker was detected in 8.51 percent of cancerous cells, with the highest expression in clusters 8 and 11. The *Survivin* marker was identified throughout 36.91 percent of the cancer cell population, with the highest expression in clusters 3 and 8. The *MCL1* marker was detected in 90.76 percent of cancer cells, with the highest expression in clusters 7 and 14. MDR is the common reason why several types of treatment have been very ineffective against it. This means that standard treatments such as chemotherapy might not be effective in patients who express one of the markers listed above. Therefore, current methods have served as alternative approaches to treatment, and they are listed as follows: immunoprevention, microparticles, and nanomedicine. Immunoprevention aims to prevent cancer development and can effectively eradicate cancer cells in the early stages. Immunoprevention strategies include vaccines, immunostimulators, and antibodies. Microparticles involve the use of nanoparticles for the targeted delivery of anticancer drugs to the body, along with clustered regularly interspaced short palindromic repeats (CRISPR)-associated technologies. Nanomedicine can achieve high efficacy in treating MDR cancer by enabling control of biodistribution, efficient drug release within resistant cancer cells, and effective delivery of anticancer drugs. The combinations of treatments above are alternative approaches for patients with MDR. Therefore, clinicians should use a combinatorial approach to treatment, administering these treatments simultaneously to overwhelm various resistance pathways. Immunotherapy can also be a great tool for clinicians, as it can significantly boost the patient's immune system to target cancer cells effectively.

When the factors are broadly generalized, the researcher can explain intratumoral heterogeneity and its significance for the prognosis and treatment of NSCLC. The genotype and phenotype diversity among these cancer cells can have a profound impact on treatment. While some forms of treatment eliminate most cancerous cells, this genetic diversity will ensure that a proportion of the cancerous cell population remains resistant to that particular treatment. Therefore, to target a large proportion of these diverse cancer cells, combination therapies are often the best-suited choice (39). The purpose of these therapies is often to target multiple pathways simultaneously while preventing the selection of resistant populations.

One of the major limitations of the dataset included the sample size and the cohort compositions. The dataset included only tumor and adjacent normal tissues from 20 patients (5). While they analyzed over 9,000 cells from these patients using single-cell RNA sequencing, the findings cannot be generalized to the global population because of the genetic and phenotypic diversity of cancer cells across populations. While these 20 patients were diagnosed at various stages of NSCLC, from I to IV, the researcher cannot tell whether each stage was represented in equal proportions. Some of the stages mentioned above might have experienced overrepresentation, and others underrepresentation. This might limit the ability to draw general conclusions regarding metastasis. For future experimentation, the researcher would seek a dataset that includes both ethnicity and cancer stage to limit potential bias. The researcher wants to find an equal proportion across all ethnicities, and, within each ethnicity, an equal proportion of cancer stages (I to IV). This would effectively limit the sample size and cohort composition while also providing a much larger sample size. This would allow the

researcher to analyze ethnicity and its major implications in cancer heterogeneity. Essentially, the researcher would like to determine whether or not ethnicity is a factor contributing to distinct cancer heterogeneities.

Another key limitation of the dataset might be in its methodology. The researchers used the Single-Cell Tagged Reverse Transcription (STRT-Seq) technique, which enables profiling of the transcriptome (the complete set of RNA molecules) of cells from different cell types and tissues (5). RNA molecules often have 5' and 3' ends, which indicate the directionality of a specific strand. The 5' end often consists of a phosphate group attached to the carbon molecule on the ribose sugar. In contrast, the 3' end consists of a free hydroxyl group(-OH) attached to the carbon molecule on the ribose sugar. Various methods of single-cell sequencing read either full-length from the 5' to the 3' end, the 5' end only, or the 3' end only. Regarding transcription coverage, the STRT-Seq technique does not capture the full length of the RNA molecule; it only targets the 5' end of the transcript (40). Thus, it is at a disadvantage relative to other single-cell RNA techniques in terms of gene allele expression, RNA editing identification, and detection of lowly expressed genes/transcripts (40). This might limit the ability to study transcription factors or rare signaling molecules. To address this, the researcher can look for a dataset that uses a multi-omic approach, such as single-cell RNA sequencing or other omics, such as epigenetics. This will significantly enhance efficiency by enabling the integration of diverse data for cell mapping, allowing detailed study of various genes and their transcription factors, providing richer cell-type data, and improving overall heterogeneity analysis (40).

CONCLUSION

In summary, the research question of metastatic and therapeutic signatures in NSCLC was addressed by exploring intratumoral heterogeneity. Through a literature review, markers of MDR, apoptosis evasion, and metastatic potential were identified and used to characterize a publicly available scRNA-seq dataset from patients at various stages of NSCLC. Clusters of cells that overexpressed the metastasis-associated markers *PIK3R1, FYN, THBS1, SPTAN1, and SPP1* were identified, suggesting a high likelihood of secondary site establishment, which is associated with poor prognosis. Similarly, MDR markers *ABCB1* and *ABCG2* were linked to distinct cell populations, suggesting that these markers should be used to guide treatment decisions. Finally, the researcher performed unsupervised clustering, identified differentially expressed markers, and commented on other cellular trajectories that could define treatment outcomes. This exploratory work is a valuable overview of a publicly available dataset, providing detailed insights into the identified altered pathways and their roles in NSCLC.

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest related to this work.

REFERENCES

1. Cleveland Clinic. Non-small cell lung cancer (NSCLC): causes, symptoms and treatment [Internet]. Cleveland (OH): Cleveland Clinic; 2022. Available from: https://my.clevelandclinic.org/health/diseases/6203-non-small-cell-lung-cancer (accessed on 2025-11-30)
2. Molinari M. Cell cycle checkpoints and their inactivation in human cancer. *Cell Prolif*. 2000; 33 (5): 261-274. https://doi.org/10.1046/j.1365-2184.2000.00191.x
3. MacDonald WJ, Purcell C, Pinho-Schwermann M, Stubbs NM, Srinivasan PR, El-Deiry WS. Heterogeneity in cancer. *Cancers (Basel)*. 2025; 17 (3): 441. https://doi.org/10.3390/cancers17030441
4. Neophytou CM, Panagi M, Stylianopoulos T, Papageorgis P. The role of tumor microenvironment in cancer metastasis: molecular mechanisms and therapeutic opportunities. *Cancers (Basel)*. 2021; 13 (9): 2053. doi:10.3390/cancers13092053
5. Li Q, Wang R, Yang Z, Li W, Yang J, Wang Z, *et al*. Molecular profiling of human non-small cell lung cancer by single-cell RNA-seq. *Genome Med*. 2022; 14 (1): 87. doi:10.1186/s13073-022-01093-2
6. Domínguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, *et al*. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*. 2022; 376 (6594): eabl5197. doi:10.1126/science.abl5197
7. Sun R, Meng X, Wang W, Liu B, Lv X, Yuan J, *et al*. Five genes may predict metastasis in non-small cell lung cancer. *Oncol Lett*. 2019; 18 (2): 1723-1732. doi:10.3892/ol.2019.10489
8. Thorpe LM, Spangle JM, Ohlson CE, Cheng H, Roberts TM, Cantley LC, *et al*. PI3K-p110alpha mediates the oncogenic activity induced by loss of the novel tumor suppressor PI3K-p85alpha. *Proc Natl Acad Sci U S A*. 2017; 114 (27): 7095-7100. doi:10.1073/pnas.1704706114
9. Peng S, Fu Y. FYN: emerging biological roles and

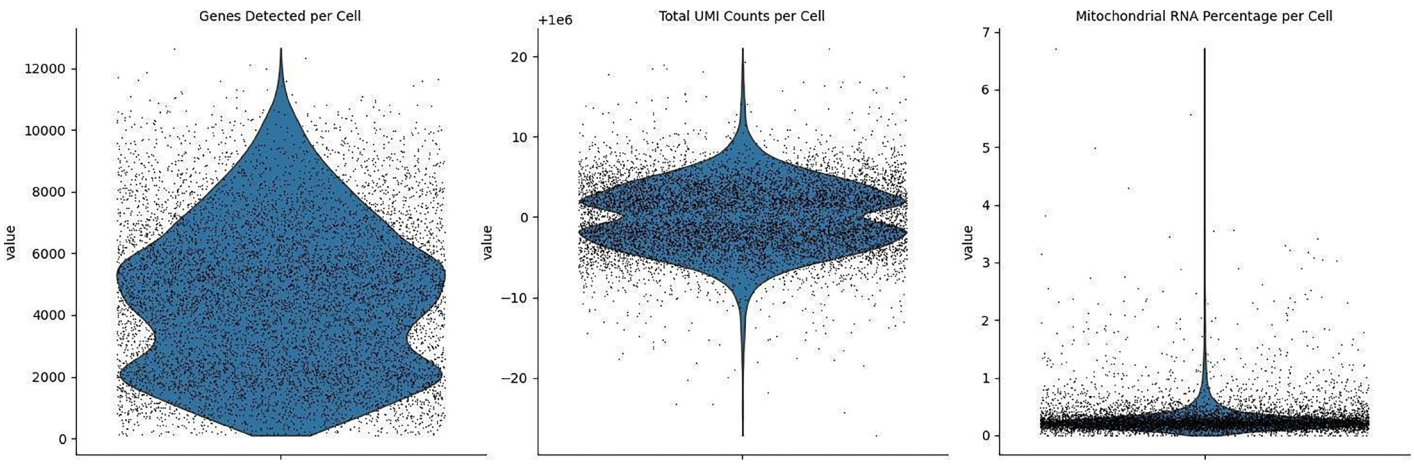
potential therapeutic targets in cancer. *J Transl Med*. 2023; 21 (1): 89. doi:10.1186/s12967-023-03902-1

10. Gururajan M, Cavassani KA, Sievert M, Duan P, Lichterman J, Huang JM, *et al*. SRC family kinase FYN promotes the neuroendocrine phenotype and visceral metastasis in advanced prostate cancer. *Oncotarget.* 2015; 6 (42): 44072-44083. doi:10.18632/oncotarget.6398
11. Ninio-Many L, Grossman H, Shomron N, Chuderland D, Shalgi R. microRNA-125a-3p reduces cell proliferation and migration by targeting FYN. *J Cell Sci*. 2013; 126 (Pt 13): 2867-2876. doi:10.1242/jcs.123414
12. Weng TY, Wang CY, Hung YH, Chen WC, Chen YL, Lai MD. Differential expression pattern of THBS1 and THBS2 in lung cancer: clinical outcome and a systematic analysis of microarray databases. *PLoS One*. 2016; 11 (8): e0161007. doi:10.1371/journal.pone.0161007
13. Lopez-Dee Z, Pidcock K, Gutierrez LS. Thrombospondin-1: multiple paths to inflammation. *Mediators Inflamm*. 2011; 2011: 296069. doi:10.1155/2011/296069
14. Schrecker C, Behrens S, Schönherr R, *et al*. SPTAN1 expression predicts treatment and survival outcomes in colorectal cancer. *Cancers (Basel)*. 2021; 13 (14): 3638. doi:10.3390/cancers13143638
15. Zhang R, Liu C, Niu Y, *et al*. microRNA-128-3p regulates mitomycin C-induced DNA damage response in lung cancer cells through repressing SPTAN1. *Oncotarget.* 2017; 8 (35): 58098-58107. doi:10.18632/oncotarget.19407
16. Machnicka B, Grochowalska R, Bogusławska DM, *et al*. Spectrin-based skeleton as an actor in cell signaling. *Cell Mol Life Sci*. 2012; 69 (2): 191-201. doi:10.1007/s00018-011-0804-5
17. Matsubara E, Komohara Y, Esumi S, Shinchi Y, Ishizuka S, Mito R, *et al*. SPP1 derived from macrophages is associated with a worse clinical course and chemo-resistance in lung adenocarcinoma. *Cancers (Basel)*. 2022; 14 (18): 4374. doi:10.3390/cancers14184374
18. Meador CB, Hata AN. Acquired resistance to targeted therapies in NSCLC: updates and evolving insights. *Pharmacol Ther.* 2020; 210: 107522. doi:10.1016/j.pharmthera.2020.107522
19. Gkogkou C, Frangia K, Saif MW, Trigidou R, Syrigos K. Necrosis and apoptotic index as prognostic factors in non-small cell lung carcinoma: a review. *Springerplus*. 2014; 3: 120. doi:10.1186/2193-1801-3-120
20. Zhang H, Guttikonda S, Roberts L, Uziel T, Semizarov D, Elmore SW, *et al*. MCL-1 is critical for survival in a subgroup of non-small-cell lung cancer cell lines. *Oncogene*. 2010; 30 (16): 1963-1968. doi:10.1038/onc.2010.577
21. Sancho M, Leiva D, Lucendo E, Orzáez M. Understanding MCL1: from cellular function and regulation to pharmacological inhibition. *FEBS J*. 2021; 289 (20): 6209-6234. doi:10.1111/febs.16138
22. Aristizábal B, González Á. Innate immune system [Internet]. Bogotá: El Rosario University Press; 2013. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459455/ (accessed on 2026-1-2)
23. Alberts B, Johnson A, Lewis J. The adaptive immune system [Internet]. New York: Garland Science; 2002. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21070/ (accessed on 2025-12-15)
24. Liu S, Shao F, Wang Y, Zhang Y, Yu H, Zhang N, *et al*. COX6C expression driven by copy amplification of 8q22.2 regulates cell proliferation via mediation of mitosis by ROS-AMPK signaling in lung adenocarcinoma. *Cell Death Dis*. 2024; 15 (1): 55. doi:10.1038/s41419-024-06411-3
25. Cole J, Guiot MC, Gravel M, Bernier C, Shore GC, Roulston A. Novel NAPRT specific antibody identifies small cell lung cancer and neuronal cancers as promising clinical indications for a NAMPT inhibitor/niacin co-administration strategy. *Oncotarget*. 2017; 8 (44): 77846-77859. doi:10.18632/oncotarget.18822
26. Park KS, Kim HK, Lee JH, Choi YB, Park SY, Yang SH, *et al*. Transglutaminase 2 as a cisplatin resistance marker in non-small cell lung cancer. *J Cancer Res Clin Oncol.* 2010; 136 (4): 493-502. doi:10.1007/s00432-009-0683-3
27. Zhang X, Deng Q, Wan X, Zhao J, Zheng X, Wang H, *et al*. Pan-cancer analysis reveals the associations between MMP13 high expression and carcinogenesis and its value as a serum diagnostic marker. *Aging (Albany NY).* 2023; 15 (6): 2115-2135. doi:10.18632/aging.204599
28. Gong F, Peng X, Luo C, Shen G, Zhao C, Zou L, *et al*. Cathepsin B as a potential prognostic and therapeutic marker for human lung squamous cell carcinoma. *Mol Cancer*. 2013; 12: 125. doi:10.1186/1476-4598-12-125
29. Koh HM, An HJ, Ko GH, Lee JH, Lee JS, Kim DC, *et al*. Prognostic role of S100A8 and S100A9 protein expressions in non-small cell carcinoma of the lung. J Pathol Transl Med. 2019;53(1):13-22. doi:10.4132/jptm.2018.11.26
30. 30. Kuroda K, Takenoyama M, Baba T, Shigematsu Y, Shiota H, Ichiki Y, *et al*. Identification of ribosomal protein L19 as a novel tumor antigen recognized by autologous cytotoxic T lymphocytes in lung adenocarcinoma. *Cancer Sci*. 2010; 101 (1): 46-53. doi:10.1111/j.1349-7006.2009.01352.x
31. Takagi H, Zhao S, Muto S, Yokouchi H, Nishihara H, Harada T, *et al*. Delta-like 1 homolog (DLK1) as a possible therapeutic target and its application to

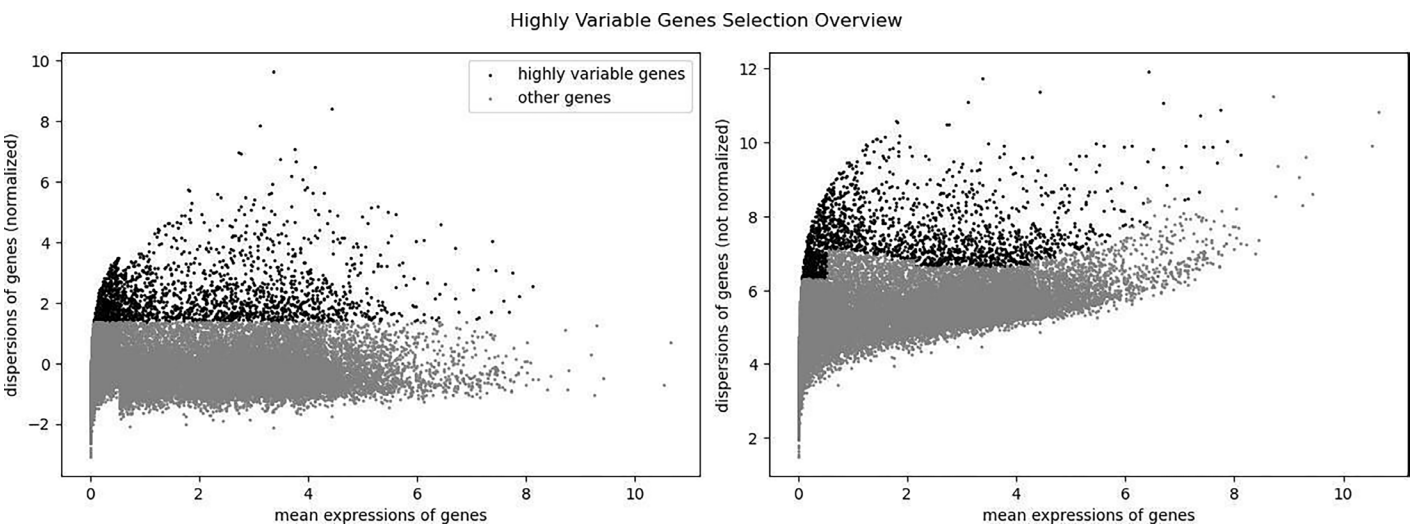
radioimmunotherapy using 125I-labelled anti-DLK1 antibody in lung cancer models. *Lung Cancer*. 2021; 153: 134-142. doi:10.1016/j.lungcan.2021.01.012

32. Wei Y, Vellanki RN, Coyaud É, Ignatchenko V, Li L, Krieger JR, *et al*. CHCHD2 is coamplified with EGFR in NSCLC and regulates mitochondrial function and cell migration. *Mol Cancer Res*. 2015; 13 (7): 1119-1129. doi:10.1158/1541-7786.MCR-14-0535
33. Li D, Xie L, Qiao Z, Mai S, Zhu J, Zhang F, *et al*. STING-mediated degradation of IFI16 negatively regulates apoptosis by inhibiting p53 phosphorylation at serine 392. *J Biol Chem*. 2021; 297 (2): 100930. doi:10.1016/j.jbc.2021.100930
34. Zhang Q, An N, Liu Y, Zhu Y, Pan W, Gu P, *et al*. Alveolar type 2 cell marker gene SFTPC inhibits epithelial-to-mesenchymal transition by upregulating SOX7 and suppressing WNT/β-catenin pathway in non-small cell lung cancer. *Front Oncol*. 2024; 14: 1460753. doi:10.3389/fonc.2024.1460753
35. Lv Y, Chen C, Han M, Tian C, Song F, Feng S, *et al*. CXCL2: a key player in the tumor microenvironment and inflammatory diseases. *Cancer Cell Int*. 2025; 25 (1): 155. doi:10.1186/s12935-025-03719-1
36. Ma X, Zhu Q, Wu Z, Han W. PSMB9 orchestrates tumor immune landscape and serves as a potent biomarker for prognosis and T cell-based immunotherapy response. *Curr Issues Mol Biol*. 2025; 47 (9): 712. https://doi.org/10.3390/cimb47090712
37. National Cancer Institute. Metastatic cancer: when cancer spreads [Internet]. National Cancer Institute; 2025. Available from: https://www.cancer.gov/types/metastatic-cancer (accessed on 2025-12-21)
38. Qian S, Wei Z, Yang W, Huang J, Yang Y, Wang J. The role of BCL-2 family proteins in regulating apoptosis and cancer therapy. *Front Oncol*. 2022; 12: 985363. doi:10.3389/fonc.2022.985363
39. Yao Y, Fareed R, Zafar A, Saleem K, Huang T, Duan Y, *et al*. State-of-the-art combination treatment strategies for advanced stage non-small cell lung cancer. *Front Oncol*. 2022; 12: 1039472. doi:10.3389/fonc.2022.1039472
40. Chen G, Ning B, Shi T. Single-cell RNA-seq technologies and related computational data analysis. *Front Genet*. 2019; 10: 317. doi:10.3389/fgene.2019.00317

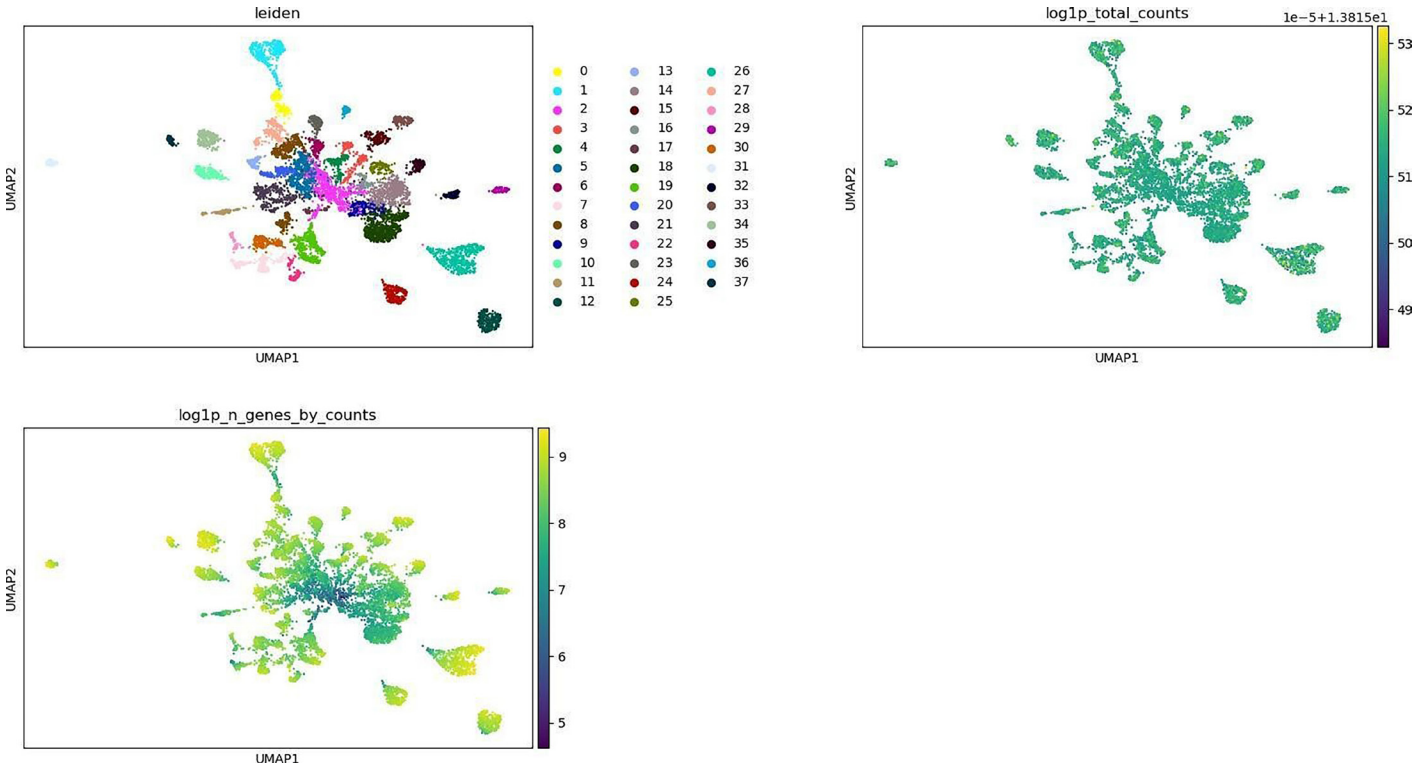
SUPPLEMENTARY MATERIALS



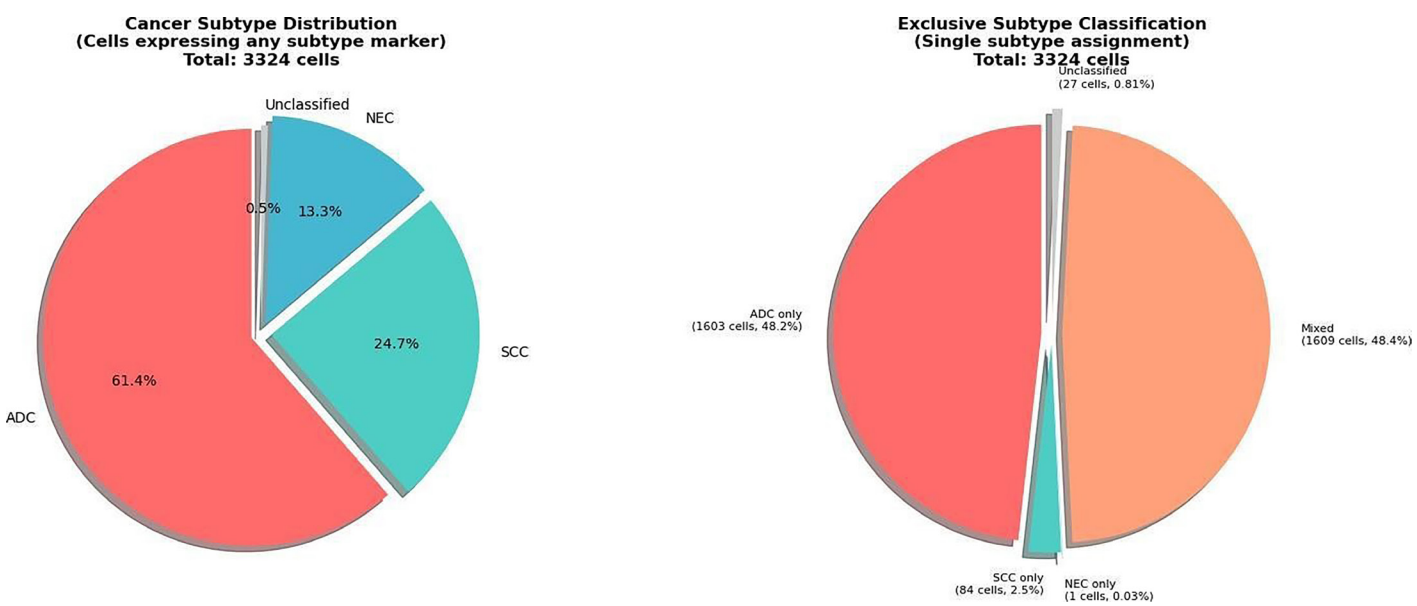
Supplementary Figure 1. *A QC Metric was created that passes a specific gene population and then calculates the proportion of gene counts. The QC Metric can then be visualized as a violin Plot with the following variables: the number of genes in the matrix, the total counts per cell, and the percentage of counts in mitochondrial genes.*



Supplementary Figure 2. *To graph these results, the count depth was first used, then scaled with a log1p transformation for normalization. Then, feature selection was used to identify the genes that are highly variable and most informative. In this case, 2,000 of the genes were selected that represented the above qualities and characteristics. Notice how the dispersion of genes in the normalized graph shows less variability compared to the dispersion of genes that were not normalized.*



Supplementary Figure 3. *A reassessment of the filtering strategy, visualizing various QC metrics using UMAP. The reassessment was performed after normalization, as demonstrated by a logarithmic transformation.*



Supplementary Figure 4. *Representation of the three main subtypes of cancer cells across the 3,324 cancerous cells, a) classification of subtypes based on subtype markers, and b) distribution based on each marker or mixtures of markers.*