



## Research Article

# Using Artificial Intelligence to Identify Tumor Microenvironment Heterogeneity in Non–Small Cell Lung Cancers

Tanner J. DuCote<sup>a</sup>, Kassandra J. Naughton<sup>a</sup>, Erika M. Skaggs, Therese J. Bocklage<sup>b,c</sup>,  
Derek B. Allison<sup>b,c</sup>, Christine F. Brainson<sup>a,b,\*</sup>

<sup>a</sup> Department of Toxicology and Cancer Biology, University of Kentucky, Lexington, Kentucky; <sup>b</sup> Markey Cancer Center, University of Kentucky, Lexington, Kentucky; <sup>c</sup> Department of Pathology and Laboratory Medicine, University of Kentucky, Lexington, Kentucky

## ARTICLE INFO

## Article history:

Received 7 February 2023

Revised 1 May 2023

Accepted 5 May 2023

Available online 12 May 2023

## Keywords:

artificial intelligence  
non–small cell lung cancer  
tumor immunology  
tumor microenvironment

## ABSTRACT

Lung cancer heterogeneity is a major barrier to effective treatments and encompasses not only the malignant epithelial cell phenotypes and genetics but also the diverse tumor-associated cell types. Current techniques used to investigate the tumor microenvironment can be time-consuming, expensive, complicated to interpret, and often involves destruction of the sample. Here we use standard hematoxylin and eosin–stained tumor sections and the HALO AI nuclear phenotyping software to characterize 6 distinct cell types (epithelial, mesenchymal, macrophage, neutrophil, lymphocyte, and plasma cells) in both murine lung cancer models and human lung cancer samples. CD3 immunohistochemistry and lymph node sections were used to validate lymphocyte calls, while F4/80 immunohistochemistry was used for macrophage validation. Consistent with numerous prior studies, we demonstrated that macrophages predominate the adenocarcinomas, whereas neutrophils predominate the squamous cell carcinomas in murine samples. In human samples, we showed a strong negative correlation between neutrophils and lymphocytes as well as between mesenchymal cells and lymphocytes and that higher percentages of mesenchymal cells correlate with poor prognosis. Taken together, we demonstrate the utility of this AI software to identify, quantify, and compare distributions of cell types on standard hematoxylin and eosin–stained slides. Given the simplicity and cost-effectiveness of this technique, it may be widely beneficial for researchers designing new therapies and clinicians working to select favorable treatments for their patients.

© 2023 United States & Canadian Academy of Pathology. Published by Elsevier Inc. All rights reserved.

## Introduction

The complexity of the tumor microenvironment (TME) is vast, with numerous cellular and structural patterns distinct in each cancer. Even within non–small cell lung cancer (NSCLC), a wide

variety of histologic variants lead to numerous possible diagnoses, each with its own preferred treatment regimen.<sup>1,2</sup> Among the different subtypes of NSCLC, adenocarcinoma (ADC) and squamous cell carcinoma (SCC) are the most commonly diagnosed. Pathological assessment of patient biopsies has long been the standard for diagnosis, and pathologists are skilled at identifying the multitude of cell types, in addition to the malignant epithelial cells, that can be found in carcinomas.<sup>3–5</sup> Although the role of the pathologist will never be replaced, a method for less highly trained researchers to quickly and accurately assess the types of cells within a histologic sample would greatly benefit the field. As

These authors contributed equally: Tanner J. DuCote and Kassandra J. Naughton.

\* Corresponding author.

E-mail address: [cfbrainson@uky.edu](mailto:cfbrainson@uky.edu) (C.F. Brainson).



ELSEVIER

NSCLC treatments move toward the use of immunotherapy as a first-line therapy for most cancers,<sup>2</sup> assessment of the TME is critical.

ADC and SCC TMEs are very heterogeneous and contain numerous types of cells, often collectively referred to as “stroma.” Commonly identified cell types in NSCLC TME include mesenchymal cells, plasma cells, macrophages, neutrophils, and lymphocytes.<sup>3,4,6,7</sup> Although cells such as CD8+ T cells in the TME can be tumor-eliminating, it is generally accepted that tumors reprogram the microenvironment to favor tumor-promoting cells. In particular, mesenchymal cells, such as cancer-associated fibroblasts, can produce various cytokines, including transforming growth factor beta, that can create an immunosuppressive TME by repressing CD8+ T cells and increasing regulatory T cells.<sup>8,9</sup> Similarly, tumor-associated neutrophils are also thought to be predominantly immunosuppressive, particularly by producing arginase and reactive oxygen species.<sup>10,11</sup> A high neutrophil-to-lymphocyte ratio typically predicts poor overall survival and poor responses to immunotherapies.<sup>12,13</sup> However, current strategies to study the neutrophil-to-lymphocyte ratio of patients with cancer involve blood samples, but these may not represent the immune microenvironment at the tumor site.

ADC and SCC differ in their histology as well as genetic profiles, leading to divergent treatment options, with ADC possessing mutations allowing for more targeted therapies.<sup>1–3</sup> Among the many genetic alterations present in NSCLCs, mutations in genes such as *KRAS*, *TP53*, and *EGFR* are some of the most common. Mutations in *KRAS* and *EGFR* are more frequent in ADCs, with *KRAS* mutated in ~32% and *EGFR* in ~27% of tumors. However, mutations in these targetable genes are much less frequent in SCCs that instead have frequent mutations in *TP53* (90%), *PIK3CA*, and *PTEN* (15%).<sup>1,14</sup> To produce large amounts of histologically and genetically similar tumors, researchers have designed genetically engineered mouse models (GEMMs) of lung cancers to mimic patient genetics and have used these models to systematically characterize lung cancer TMEs. Data have shown that murine lung adenocarcinomas tend to attract macrophages,<sup>15–17</sup> but if there are alterations in *EGFR*, *LKB1*, or in late-stage *KRAS*-driven tumors, neutrophils may be recruited.<sup>18–20</sup> In murine squamous lung cancers, neutrophils predominate and are believed to drive a transition from ADC to SCC states.<sup>15,17,21,22</sup>

Current methods to characterize the TME include flow cytometry, single-cell RNAseq (scRNAseq), mass cytometry, and other multiplexed immunohistochemistry approaches.<sup>4,6,23,24</sup> Flow cytometry and standard single-cell RNAseq are robust methods but require a relatively large amount of live starting material that must be dissociated prior to experimental assessment. Dissociating tissues can lead to the loss of fragile cell populations and erase spatial information. Spatial scRNAseq can overcome this hurdle but is expensive to implement. Other options include spatial profiling through multiplexed immunohistochemistry and imaging mass cytometry techniques such as fluidigm or spatial proteomics.<sup>25,26</sup> These are powerful approaches but require specialized equipment, a large investment of money for reagents, and trained personnel to interpret results.<sup>23</sup> Therefore, developing a TME quantification method that is more cost effective, faster to implement, and retains the sample is imperative to further research efforts and to provide clinicians with a tool to help guide treatment for their patients.

In this study, we used the HALO AI nuclear phenotyping software (Indica Labs) to identify differences in the TME in both human and murine NSCLC tissues. We investigated 5 murine models, including *Kras*<sup>G12D</sup>/*p53*-null, *Pik3ca*<sup>E545K</sup>/*p53*-null, *EGFR*<sup>T790M/L858R</sup>, *Kras*<sup>G12D</sup>/*Lkb1*-null, and *Lkb1*-null/*Pten*-null, to determine how the

presence of these mutations alters the TME. We confirmed the accuracy of the algorithm with both correlations to immunohistochemistry (IHC) staining and testing on lymphocyte-rich lymph node sections. In addition to further identifying differences in these tumors, we confirmed accumulations of macrophages in murine adenocarcinomas and neutrophils in murine squamous cell carcinomas. In human samples, we uncovered extremely strong negative correlations between both mesenchymal cells and neutrophils with lymphocytes, and higher proportions of mesenchymal cells predicted poor overall survival.

## Methods

### Mouse Models

Mouse models included were: *KRAS*<sup>G12D</sup>/*p53*-null lung adenocarcinoma,<sup>27,28</sup> *KRAS*<sup>G12D</sup>/*p53*-null/*Ezh2*-heterozygous and *KRAS*<sup>G12D</sup>/*p53*-null/*Ezh2*-null lung adenocarcinoma,<sup>29</sup> *PIK3CA*<sup>E545K</sup>/*p53*-null lung adenocarcinoma,<sup>30</sup> *EGFR*<sup>T790M/L858R</sup> lung adenocarcinoma,<sup>31</sup> *KRAS*<sup>G12D</sup>/*Lkb1*-null mixed lineage tumors,<sup>17,32</sup> and *Lkb1*-null/*Pten*-null squamous cell carcinoma.<sup>15</sup> All animal work was approved by the University of Kentucky, Dana-Farber Cancer Center, or Boston Children's Hospital Institutional Animal Care and Use Committee. Adult mice were allowed to inhale Cre- or FlpO-encoding virus to initiate autochthonous lung tumors. Mice housed at the University of Kentucky all received  $2.9 \times 10^6$  pfu of adenoviral Cre (University of Iowa). Historical banked tissues from previous studies were also used.

### Patient Samples

The patient tissue was obtained from the Markey Cancer Center Biospecimen Procurement and Translational Pathology Shared Resource Facility (BPTP SRF). A tissue microarray (TMA) was prepared from patient biopsies from de-identified excess tissue. The array consisted of 3 core biopsies from each patient, with a total of 83 adenocarcinoma (ADC), 102 squamous cell carcinoma (SCC), 14 adenosquamous and mixed histology tumors (ADSCC), and 17 poorly differentiated tumors, including 2 large cell carcinoma, 1 giant cell carcinoma, 1 pleomorphic carcinoma, and 1 sarcomatoid carcinoma.

### Histology and Immunohistochemistry

All tissues were fixed with 10% neutral-buffered formalin overnight. They were then transferred to ethanol, embedded in paraffin, and sectioned at 4  $\mu$ m. CD3 immunohistochemistry was performed using the BPTP SRF at the Markey Cancer Center. Antigen retrieval was performed in a Biocare Medical decloaking chamber at 95 °C for 30 minutes with Dako Target Retrieval Solution, pH 9 (S236784-2). Endogenous peroxidase activity was quenched with Dako peroxidase block (K800021-5). The primary antibody for CD3 (Dako, IR503 ready-to-use) was added and incubated at room temperature for 45 minutes. Amplification was performed using Dako Envision anti-rabbit horseradish peroxidase conjugated antibody for 30 minutes at room temperature (K4003). Detection was performed with DAB for 10 minutes (Dako K346711-2). Slides were then counterstained with Harris hematoxylin. F4/80 staining was also performed by the BPTP SRF. Antigen retrieval was performed on the Ventana Discovery Ultra system using the CC2 mild protocol (citrate pH 6 buffer, 32 minutes

at 91 °C). The primary antibody anti-F4/80 (Cell Signaling, #70776) was used at 1:200 for 1 hour at 37 °C. Detection was performed using Ventana OmniMap anti-Rabbit-HRP for 20 minutes, followed by Ventana DAB. Slides were then counterstained with Mayer's hematoxylin and blued with ammonia water. Dehydrated mounted slides were scanned with the Aperio slide scanner at  $\times 40$  magnification, 6 tumor-containing regions of 355 mm<sup>2</sup> were selected, and the HALO classifier Multiplex IHC v3.0.4 was used to quantify the positively stained cells within the regions.

#### HALO AI Nuclear Phenotyper Algorithm

Histology slides were scanned at  $\times 20$  to  $\times 40$  with an Aperio AT2 scanner in brightfield mode. The scanner had an Olympus  $\times 20$  objective with an optional  $\times 2$  magnifier, with a resolution of 0.5  $\mu\text{m}/\text{pixel}$  at  $\times 20$  magnification and 0.25  $\mu\text{m}/\text{pixel}$  at  $\times 40$  magnification. The format was an SVS file, a tiled TIFF, which was then compressed to JPEG. Z-stacking was not used. The images were loaded into the HALO program, and the AI nuclear phenotyper algorithm was trained for 556,930 iterations using a total of 34,427 nuclei from 52 different samples. The accuracy of the final nuclear phenotyper algorithm was verified by pathologists at the Markey Cancer Center. For each murine lung section, the tumor areas with approximately 50  $\mu\text{m}$  surrounding area were hand-annotated for analysis. For the *KRAS/Lkb1-null* tumors, adenocarcinoma and squamous cell carcinoma regions were manually sub-annotated by an experienced researcher.

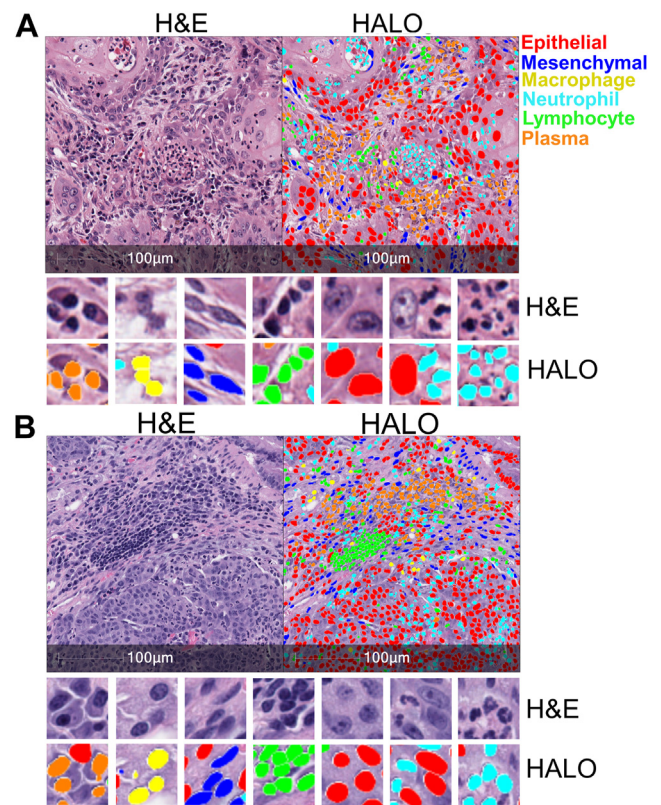
#### Statistics and Reproducibility

Statistics were performed using GraphPad Prism v9.2.0. The statistical tests and the exact sample numbers (n) are listed in the figure legends. The HALO software versions were HALO v3.5.3577.214 and HALO AI 3.5.3577. For murine data, all tumors of a given subtype present on a slide were analyzed together. For human data, results from 3 independent core biopsies were averaged. The n listed in the figures are biological replicates for individual mice or humans.

## Results

#### Training and Validation of the Nuclear Phenotyper Algorithm for Human and Murine Lung Cancer

To profile lung cancer TME components using hematoxylin and eosin (H&E)-stained sections, we developed an algorithm that could reliably identify cell types in both patient and murine samples based on nuclear morphology. We used the HALO AI nuclear phenotyping software and trained the software in iterations to identify 6 distinct cell types: epithelial cells, mesenchymal cells, plasma cells, macrophages, neutrophils, and lymphocytes. Overall, 556,930 iterations were run using 34,427 nuclei from 19 human samples and 33 murine samples. Representative images of human squamous cell carcinoma (Fig. 1A) and murine squamous cell carcinoma (Fig. 1B) show all 6 subtypes of cells identified by the algorithm. Neutrophils were identified by their polymorphonuclear or segmented nuclei. Plasma cells contained a characteristic perinuclear "hof," which is a light area adjacent to the nucleus. Tumor or epithelial cells were identified by their generally large cell size, pleomorphic nucleus, and dark nucleoli. Mesenchymal cells had elongated and spindle-shaped nuclei

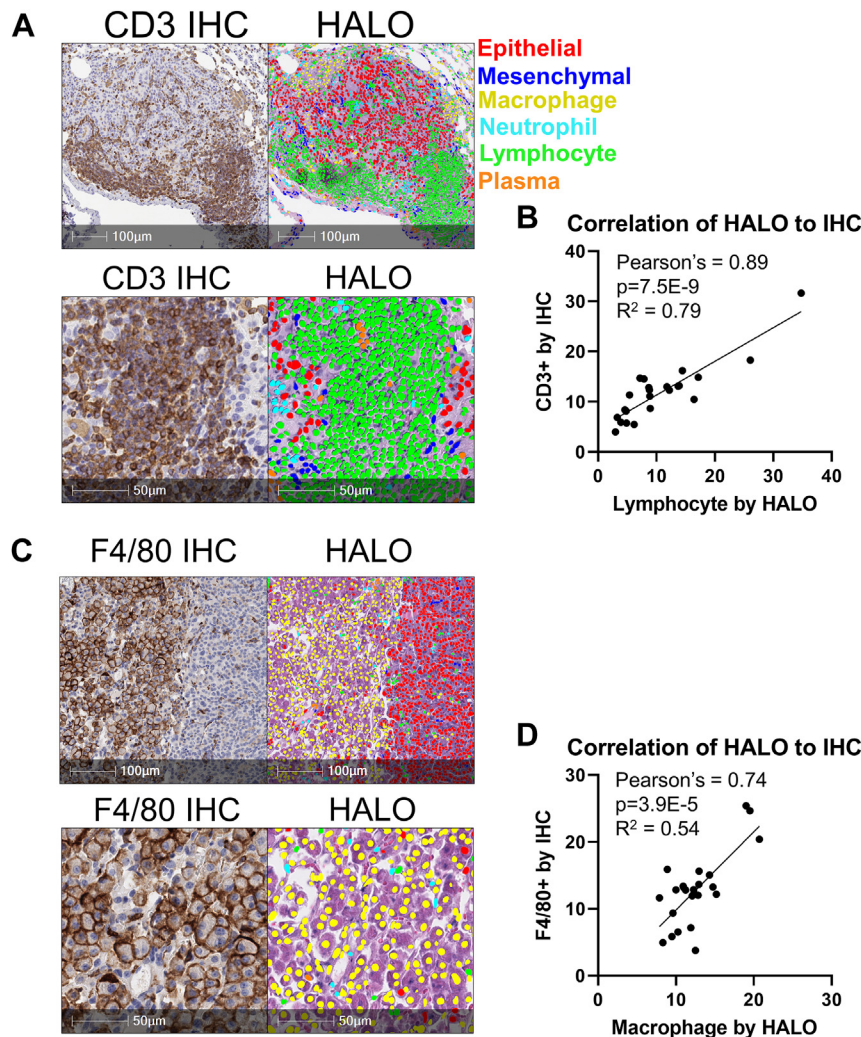


**Figure 1.**

Artificial Intelligence identifies nuclear phenotypes in human and mouse lung cancers. (A) H&E-stained sample from a patient with squamous cell carcinoma (left). The same H&E-stained image with different cell types identified by the HALO AI nuclear phenotyping algorithm overlaid with corresponding colors (right). Zoomed images show more detail of the different cell types identified in the tumor immune micro-environment. (B) H&E-stained sample from *Kras<sup>G12D</sup>/Lkb1-null* mouse with squamous cell carcinoma (left), with the cell types identified by the HALO AI algorithm (right). Cell types are color-coded as follows: epithelial/tumor cells (red), mesenchymal cells (dark blue), neutrophils (light blue), plasma cells (orange), lymphocytes (green), and macrophages (yellow). H&E, hematoxylin and eosin.

resembling endothelial cells, smooth muscle, fibroblasts, and malignant epithelial cells that had undergone an epithelial-mesenchymal transition. Lymphocytes were characterized by small cytoplasm containing a small but often dark and rounded nucleus. Lastly, macrophages were identified by their round nuclei and diffuse cytoplasmic area.

Next, we examined the concordance of the nuclear phenotyper algorithm with traditional IHC approaches. First, we compared the identification of lymphocytes using the algorithm to IHC staining of the T cell marker CD3 and observed a strong overlap in cell identification by the 2 methods (Fig. 2A). We recently reported that when one copy of the gene encoding the histone methyltransferase EZH2 is deleted in *Kras<sup>G12D</sup>/p53-null* murine tumors, there are more abundant lymphocytes in the tumors.<sup>29</sup> Using tumors that were EZH2 wild-type, heterozygous, and null, we directly compared the abundance of lymphocytes in the tumors as measured by the nuclear phenotyping algorithm and by CD3 IHC quantification (Supplementary Fig. S1A, B). With both methods, we observed that EZH2 heterozygous tumors had statistically more lymphocytes as a percentage of total cells. The concordance of the 2 methods was also strong (Fig. 2B). To further validate the ability of the algorithm to identify lymphocytes and plasma cells accurately, we examined murine lymph nodes. The lymph nodes of a murine sample can be observed with an abundance of

**Figure 2.**

The HALO AI nuclear phenotyping algorithm accurately identifies cell types. (A) Representative CD3 IHC staining of lung tumor with abundant tumor-infiltrating lymphocytes. Lymphocytes in the IHC are stained with anti-CD3 antibody and stained brown (left) with HALO AI nuclear phenotyping algorithm overlaid on the H&E-stained section (right). (B) Graph of the correlation of percentages of CD3+ cells versus lymphocytes identified by HALO AI with Pearson's correlation coefficient and R<sup>2</sup> values indicated on the chart, n = 24 mice total. (C) Representative F4/80 IHC staining of lung tumor with abundant tumor-infiltrating lymphocytes. Lymphocytes in the IHC are stained with anti-F4/80 antibody and stained brown (left) with HALO AI nuclear phenotyping algorithm overlaid on the H&E-stained section (right). (D) Graph of the correlation of percentages of F4/80+ cells versus macrophages identified by HALO AI with Pearson's correlation coefficient and R<sup>2</sup> values indicated on the chart, n = 24 mice total. H&E, hematoxylin and eosin; IHC, immunohistochemistry.

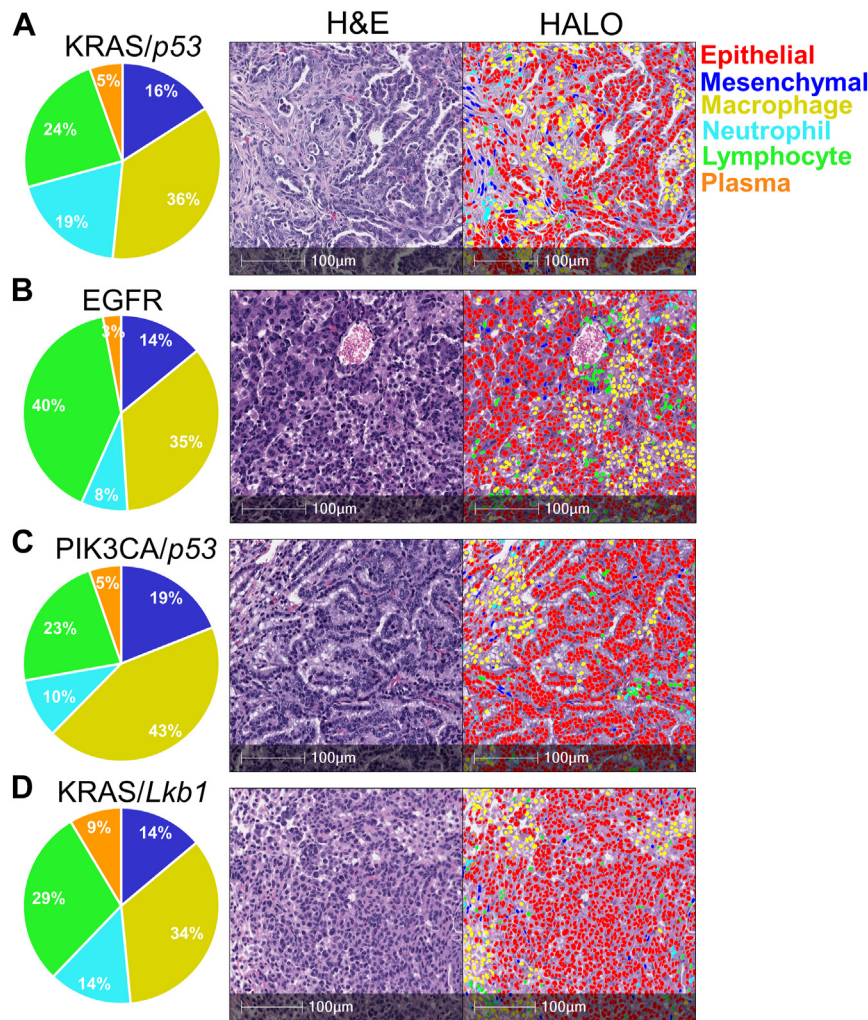
lymphocytes and lower numbers of the other cell types, which is expected for a typical lymph node (Supplementary Fig. S1C). In contrast, the nuclear phenotyper identifies numerous other cell types, including abundant epithelial cells, when a metastatic tumor is present in the lymph node tissue (Supplementary Fig. S1D). Lastly, we used F4/80 IHC, compared it to macrophages identified by the phenotyper and again observed a strong concordance between the methods (Fig. 2C, D). Taken together, these data indicate that the HALO AI nuclear phenotyper is an easy and reproducible way to identify changes in the TME.

#### Identification of Predominant Cell Types Within Lung Cancer of Genetically Engineered Mouse Models

To further investigate the capabilities of the HALO AI nuclear phenotyping algorithm, we examined H&E-stained tumors from 4

distinct lung cancer models: *KRAS*<sup>G12D</sup>/*p53*-null lung adenocarcinoma,<sup>27,28</sup> *PIK3CA*<sup>E545K</sup>/*p53*-null lung adenocarcinoma,<sup>30</sup> *EGFR*<sup>T790M/L858R</sup> lung adenocarcinoma,<sup>31</sup> and *KRAS*<sup>G12D</sup>/*Lkb1*-null mixed lineage tumors.<sup>17,32</sup> Consistent with numerous previous reports,<sup>15–17</sup> all 4 models had large populations of macrophages at the site of the tumors. When considering the other abundant populations, the *KRAS*<sup>G12D</sup>/*p53*-null model possessed the highest percentage of neutrophils relative to the other 2 models (Fig. 3A); the *PIK3CA*<sup>E545K</sup>/*p53*-null model had the highest percentage of mesenchymal cells (Fig. 3B); the *EGFR*<sup>T790M/L858R</sup> mutant model had the largest proportion of lymphocytes present (Fig. 3C); and the *KRAS*<sup>G12D</sup>/*Lkb1*-null had the most plasma cells (Fig. 3D).

In addition to the *KRAS*<sup>G12D</sup>/*Lkb1*-null mouse model generating adenocarcinomas, it is also capable of creating tumors of adenocarcinoma and fully squamous histology.<sup>17,32</sup> Again, consistent with several reports,<sup>17,21,22</sup> the algorithm showed that squamous areas from the *KRAS*<sup>G12D</sup>/*Lkb1*-null model predominantly recruit

**Figure 3.**

Diverse genotypes of murine lung adenocarcinomas have predominant macrophage infiltration. (A–D) Representative H&E-stained sections with different cell types identified by the HALO AI nuclear phenotyping algorithm. Pie charts represent percentages of non-epithelial cells. (A) KRAS<sup>G12D</sup>/p53-null lung adenocarcinomas, n = 8. (B) PIK3CA<sup>E545K</sup>/p53-null lung adenocarcinomas, n = 6. (C) EGFR<sup>T790M/L858R</sup> lung adenocarcinomas, n = 6. (D) KRAS<sup>G12D</sup>/Lkb1-null adenocarcinomas, n = 9. H&E, hematoxylin and eosin.

neutrophils (Fig. 4A). We then explored the SCC model generated through biallelic deletion of the tumor suppressors *Lkb1* and *Pten*.<sup>15</sup> Tumors from this model recruit a considerable number of neutrophils, as indicated by the presence of large pockets of polymorphonuclear cells (Fig. 4B). Likewise, when we compared immune cell composition from squamous tumor models, neutrophils predominated; however, among all the adenocarcinoma models, macrophages predominated (Fig. 4C). The distinct tropisms of the adenocarcinoma and squamous cell carcinomas were particularly evident when the 2 tumors were juxtaposed, showing macrophages recruited toward the adenocarcinoma histology and neutrophils toward the squamous tumor (Supplementary Fig. S2).

#### Characterization of Human Non–Small Cell Lung Cancer Cell Profiles

To examine the TME heterogeneity in human samples, we used a TMA generated from 216 NSCLC tumor samples. Using this TMA, we could recapitulate an accurate identification of the cells within the TME using the HALO AI nuclear phenotyping algorithm in both

ADC and SCC (Fig. 5A, B). In contrast to our observations in the mouse models of NSCLC, we did not observe any significant differences between the cell types within these tumors (Fig. 5C). The most commonly identified TME cell type was lymphocytes, in agreement with other studies.<sup>4,7</sup> The fact that the histotype did not predict cell infiltration suggests that more heterogeneous factors contribute to the recruitment of immune cells within the tumor stroma in human samples than those present in our genetically defined mouse models.

When comparing the percentages of mesenchymal cells to lymphocytes, we observed a very significant negative correlation between the 2 populations (Fig. 6A). Furthermore, we observed a strongly significant negative correlation between tumor-infiltrating neutrophils and lymphocytes (Fig. 6B). This finding may indicate that mesenchymal cells and neutrophils in the tumor stroma are phenotypically detrimental to lymphocyte survival or recruitment. To further interrogate the relationship between the immune infiltrates identified by the HALO AI nuclear phenoter, we created a correlation plot (Fig. 6C). Positive correlations were observed between mesenchymal cells and neutrophils, suggesting that the mesenchymal cells could be phenotypically attracting

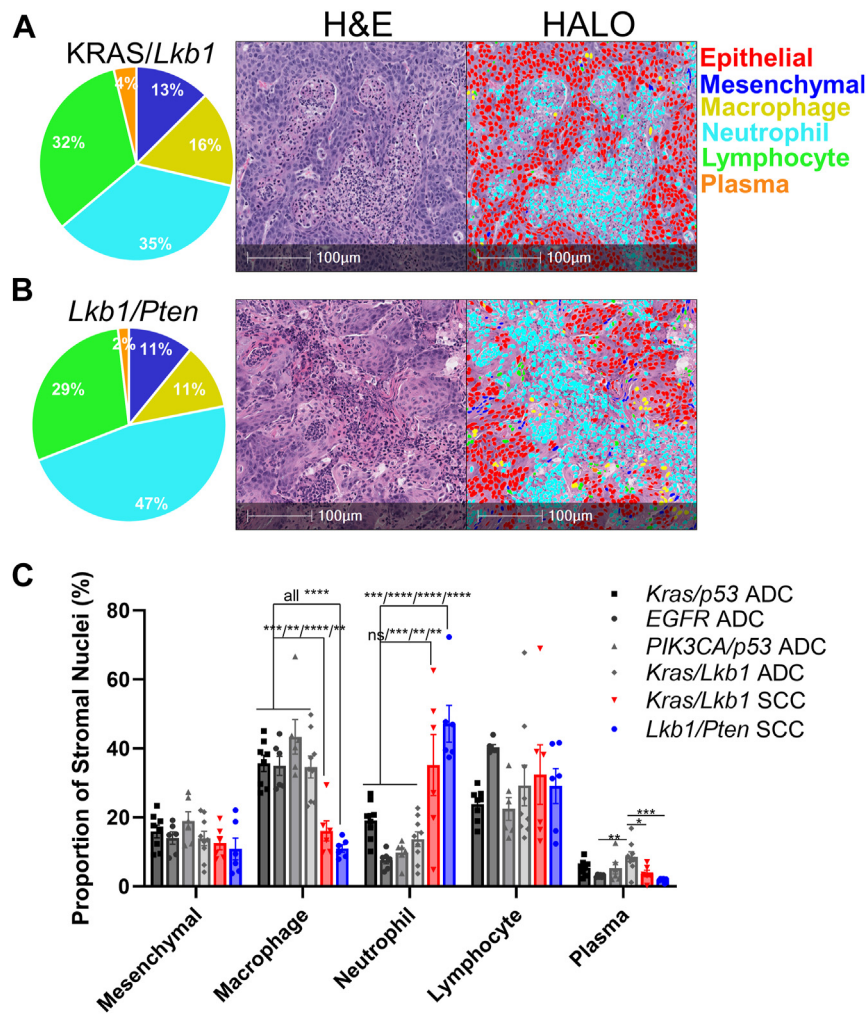


Figure 4.

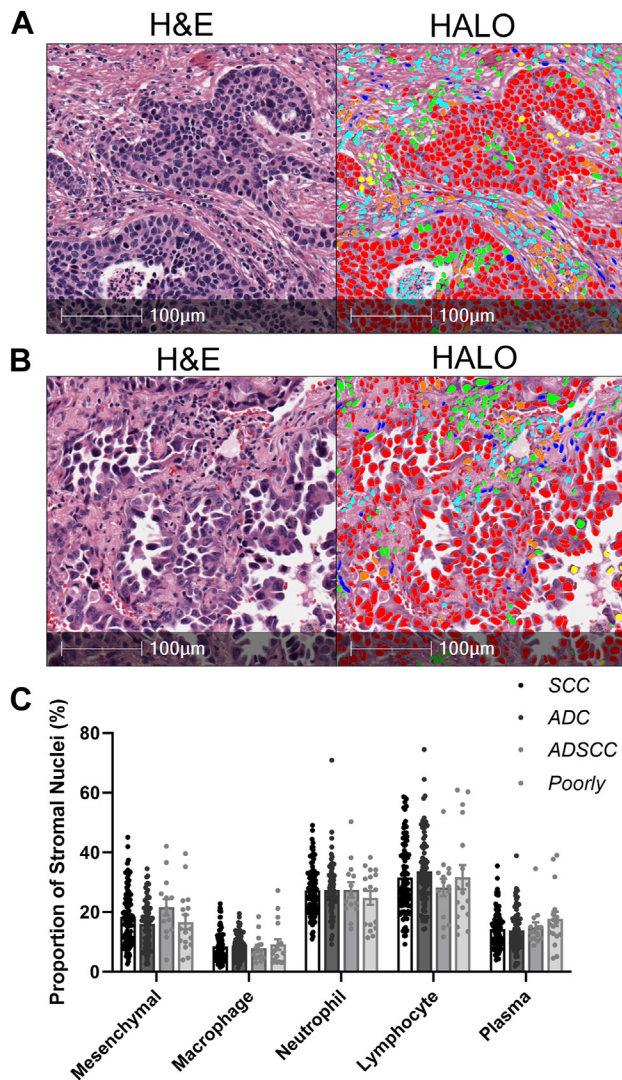
Neutrophil infiltration predominates in genetically defined murine lung squamous cell carcinomas. (A, B) Representative H&E-stained sections with different cell types identified by the HALO AI nuclear phenotyping algorithm. Pie charts represent percentages of nonepithelial cells. (A) KRAS<sup>G12D</sup>/Lkb1-null squamous cell carcinomas, n = 6. (B) Lkb1-null/Pten-null squamous cell carcinoma, n = 6. (C) The bar graph represents the proportion of nonepithelial nuclei for the indicated genetic mouse models of NSCLC, plotted as the mean ± SEM. One-way ANOVA with multiple comparisons and Holm-Sidak's multiple comparisons tests were performed. \*P < .02, \*\*P < .006, \*\*\*P < .001, and \*\*\*\*P < .0001. H&E, hematoxylin and eosin; NSCLC, non-small cell lung cancer.

neutrophils and together support a highly immunosuppressed environment. A negative correlation was also observed between macrophages and lymphocytes, but the degree of the correlation was lower.

We last sought to determine whether the abundance of any cell types within the tumors predicted poor overall survival in this cohort. We observed that patients whose tumors were classified as mesenchymal low survived longer than patients whose tumors were classified as mesenchymal high (Fig. 6D). This result was similar whether we queried all causes of death or if we limited the cohort to lung cancer-specific death (Supplementary Fig. S3A). However, with multivariate analysis, confounding variables of sex, age at diagnosis, and tumor histology reduced the significance of the correlation of mesenchymal cells with poor survival (Supplementary Fig. S3B, C, Supplementary Table S1). Although not significant, high lymphocyte abundance predicted a better prognosis, and a higher number of plasma cells predicted a worse prognosis (Supplementary Fig. S3B, C).

## Discussion

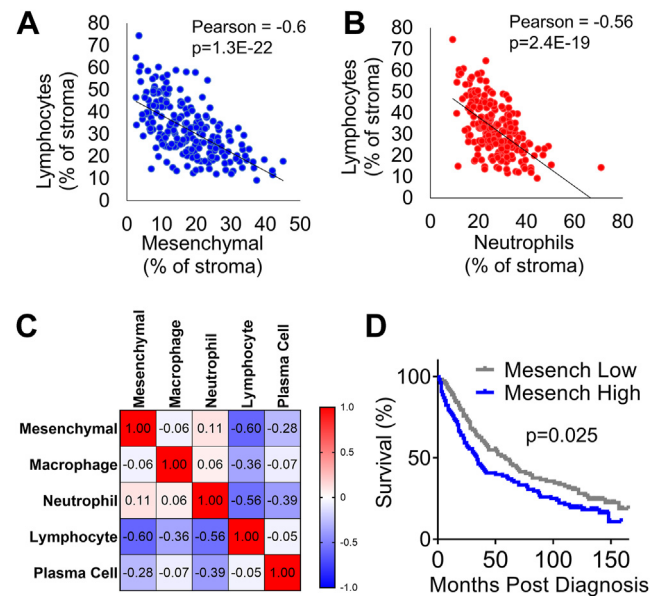
While there are many methods to investigate the TME, some are expensive, time-consuming, and often destroy the sample in the process of data generation. We present a HALO AI nuclear phenotyping algorithm to examine the TME in NSCLC. This method boasts low overall costs, quick turnaround, and the ability to retain the original sample. Using both human and mouse histology samples, we demonstrated the algorithm's ability to identify different TME cell types, validating it as a tool for researchers and a potential clinical tool for selecting appropriate treatments. Using GEMMs with known immune cell tropisms, we confirmed that the nuclear phenotyping algorithm works very well to define the microenvironments of these subtypes of NSCLC. This is one of the many ways this software can differentiate between the microenvironments of these subtypes of NSCLC. We further confirmed the efficacy of the algorithm by assessing lymph node sections and comparing the algorithm results to a more traditional IHC staining for CD3. The algorithm performed very well with



**Figure 5.** The HALO AI nuclear phenotypic analysis illustrates heterogeneity in patient NSCLC samples. (A) Representative human squamous cell carcinoma. (B) Representative human adenocarcinoma. (C) Proportions of the different cell types within the tumor microenvironment, SCC n = 102; ADC n = 83; ADSCC n = 14; poorly differentiated n = 17. ADC, adenocarcinoma; ADSCC, adenosquamous and mixed histology tumors; H&E, hematoxylin and eosin; NSCLC, non-small cell lung cancer; SCC, squamous cell carcinoma.

each of these known data sets and could provide additional information about the abundant cell types in the GEMMs examined.

In a patient sample TMA, the AI nuclear phenotypic provided a rapid and simple way to study the correlation of different cell types at the tumor site. Our data demonstrated that neutrophils and lymphocytes are highly negatively correlated with human NSCLCs. This finding is in agreement with fresh patient tissue flow cytometry, which demonstrated that neutrophils are negatively correlated with CD8<sup>+</sup> T cells.<sup>24</sup> Given that CD8<sup>+</sup> T cells are responsible for eliminating malignant epithelial cells in immunotherapy contexts, a high neutrophil/low lymphocyte tumor may not respond well to immunotherapy, and this appears to be the case in patient samples.<sup>33</sup> Furthermore, we showed a strong negative correlation between mesenchymal cells and lymphocytes, and this is an additional avenue that could be explored for immunotherapy response. Given that the



**Figure 6.** Lymphocytes negatively correlate with neutrophils and mesenchymal cells, and mesenchymal cells predict poor prognosis in human samples. (A) Correlation plot between percentages of mesenchymal cells and lymphocytes, Pearson's correlation coefficient, and P value indicated on the graph. (B) Correlation plot between percentages of neutrophils and lymphocytes, Pearson's correlation coefficient, and P value indicated on the graph. (C) The correlation matrix depicts relationships between all tumor-associated cell types. (D) Kaplan-Meier survival plot between mesenchymal low vs mesenchymal high tumors split at the median. The P value shown is for the Mantel-Cox log-rank test, n = 216 patients for all plots.

mesenchymal cell population also correlates with poor prognosis, further characterizing this heterogeneity will be an important next step. Lastly, we observed numerous plasma cells in human NSCLCs, and recent reports have demonstrated that plasma cells may have a negative correlation to survival in lung cancer patients.<sup>34,35</sup> These findings emphasize the importance of this program as a tool for researchers to understand the function of the TME and the potential of this method to help determine patient treatment strategies. This easy-to-implement approach allows for a specific understanding of the TME at the site of the tumor itself and has the potential to allow researchers to investigate further how the abundance of different cell types influences the efficacy of therapies.

The data described here do not represent the first instance where AI has been suggested as a potential prognostic technique for histopathology. In fact, AI has been predicted to become a useful, if not necessary, tool for pathologists to triage slide analysis, quantify phenotypes, and even predict genetic alterations.<sup>36,37</sup> However, there remain concerns about the ability of AI techniques to accurately differentiate between histopathologies, particularly if the algorithm is not trained on certain distinct patterns. Among proponents of AI, it is widely accepted that the rigor of these algorithms will need to be tested abundantly to prove that their efficiency is equal to that of a pathologist before they are implemented as a prognostic technique.<sup>37</sup> Therefore, the immediate implications of this technique are to allow translational research to begin adopting this technique while further clinical validation is ongoing. Although we specifically study NSCLC, we believe this artificial intelligence-based cell detection algorithm will have wider utility in many other diseases characterized by heterogeneous chronic inflammation.

## Acknowledgments

We thank Dana Napier at the BPTP Shared Resource Facilities for extensive support with the software and immunohistochemistry.

## Author Contributions

C.F.B., D.B.A., and T.J.B. conceptualized and trained the algorithm. T.J.D., K.J.N., and C.F.B. performed formal analysis. C.F.B. was responsible for funding acquisition. C.F.B. and E.M.S. performed data acquisition and processing. T.J.D., K.J.N., and C.F.B. wrote, reviewed, and edited the original manuscript.

## Data Availability

All data presented in this manuscript are available from the corresponding author upon reasonable request. The HALO AI nuclear phenotyper is available on Zenodo (<https://doi.org/10.5281/zenodo.7883644>). An ONNX (open neural network file) is available upon request with the appropriate MTA.

## Funding

This work was supported in part by a grant from the American Cancer Society (133123-RSG-19-081-01-TBG) and grants from the American Institute for Cancer Research (NIGMS P20 GM121327 and NCI R01 CA237643). This research was also supported by the Cancer Research Informatics, Biostatistics and Bioinformatics, and Biospecimen Procurement and Translational Pathology Shared Resource Facilities of the University of Kentucky, Markey Cancer Center (P30 CA177558).

## Ethics Approval and Consent to Participate

All animal work was approved by and performed in accordance with the University of Kentucky, Dana-Farber Cancer Center, or Boston Children's Hospital Institutional Animal Care and Use Committees guidelines. A TMA was prepared from patient surgical specimens. The samples were leftover clinical specimens collected under an IRB-approved protocol with informed consent or waiver. The BPTP SRF at the University of Kentucky acted as an honest broker and deidentified the tissue. The Cancer Research Informatics Shared Resource Facility provided the necessary clinical data on deidentified specimens. Because the investigators do not have patient identifying information, this is considered IRB-exempt research.

## Supplementary Material

The online version contains supplementary material available at <https://doi.org/10.1016/j.labinv.2023.100176>

## References

- Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature*. 2018;553(7689):446–454.
- Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. *Lancet*. 2021;398(10299):535–554.

- Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong K-K. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer*. 2014;14(8):535–546.
- Stankovic B, Bjørhovde HAK, Skarshaug R, et al. Immune cell composition in human non-small cell lung cancer. *Front Immunology*. 2018;9:3101.
- Kadota K, Nitadori J-i, Ujii H, et al. Prognostic impact of immune microenvironment in lung squamous cell carcinoma. *J Thorac Oncol*. 2015;10(9):1301–1310.
- Zilionis R, Engblom C, Pfirschke C, et al. Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. *Immunity*. 2019;50(5):1317–1334.e1310.
- Wang C, Yu Q, Song T, et al. The heterogeneous immune landscape between lung adenocarcinoma and squamous carcinoma revealed by single-cell RNA sequencing. *Signal Transduct Target Ther*. 2022;7(1):289.
- Koppensteiner L, Mathieson L, O'Connor RA, Akram AR. Cancer associated fibroblasts - an impediment to effective anti-cancer T cell immunity. *Front Immunol*. 2022;13, 887380.
- Baker AT, Abuwarwar MH, Poly L, Wilkins S, Fletcher AL. Cancer-associated fibroblasts and T cells: from mechanisms to outcomes. *J Immunol*. 2021;206(2):310–320.
- Aarts CEM, Hiemstra IH, Béguin EP, et al. Activated neutrophils exert myeloid-derived suppressor cell activity damaging T cells beyond repair. *Blood Adv*. 2019;3(22):3562–3574.
- Rodríguez PC, Quiceno DG, Zabaleta J, et al. Arginase I production in the tumor microenvironment by mature myeloid cells inhibits T-cell receptor expression and antigen-specific T-cell responses. *Cancer Res*. 2004;64(16):5839–5849.
- Cupp MA, Cariolou M, Tzoulaki I, Aune D, Evangelou E, Berlanga-Taylor AJ. Neutrophil to lymphocyte ratio and cancer prognosis: an umbrella review of systematic reviews and meta-analyses of observational studies. *BMC Med*. 2020;18(1):360.
- Jin J, Yang L, Liu D, Li W. Association of the neutrophil to lymphocyte ratio and clinical outcomes in patients with lung cancer receiving immunotherapy: a meta-analysis. *BMJ Open*. 2020;10(6):e035031.
- Campbell JD, Alexandrov A, Kim J, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet*. 2016;48(6):607–616.
- Xu C, Fillmore Christine M, Koyama S, et al. Loss of Lkb1 and Pten leads to lung squamous cell carcinoma with elevated PD-L1 expression. *Cancer Cell*. 2014;25(5):590–604.
- Wang D-H, Lee H-S, Yoon D, et al. Progression of EGFR-mutant lung adenocarcinoma is driven by alveolar macrophages. *Clin Cancer Res*. 2017;23(3):778–788.
- Zhang H, Fillmore Brainson C, Koyama S, et al. Lkb1 inactivation drives lung cancer lineage switching governed by Polycomb Repressive Complex 2. *Nature Commun*. 2017;8, 14922–14922.
- Koyama S, Akbay EA, Li YY, et al. STK11/LKB1 Deficiency promotes neutrophil recruitment and proinflammatory cytokine production to suppress T-cell activity in the lung tumor microenvironment. *Cancer Res*. 2016;76(5):999–1008.
- Pfirschke C, Engblom C, Gungabeesoon J, et al. Tumor-promoting Ly-6G(+) SiglecF(high) cells are mature and long-lived neutrophils. *Cell Rep*. 2020;32(12), 108164.
- Tang KH, Li S, Khodadadi-Jamayran A, et al. Combined inhibition of SHP2 and CXCR1/2 promotes antitumor T-cell response in NSCLC. *Cancer Discov*. 2022;12(1):47–61.
- Nagaraj AS, Lahtela J, Hemmes A, et al. Cell of origin links histotype spectrum to immune microenvironment diversity in non-small-cell lung cancer driven by mutant Kras and loss of Lkb1. *Cell Rep*. 2017;18(3):673–684.
- Mollaoglu G, Jones A, Wait SJ, et al. The lineage-defining transcription factors SOX2 and NKX2-1 determine lung cancer cell fate and shape the tumor immune microenvironment. *Immunity*. 2018;49(4):764–779.e769.
- Moffitt JR, Lundberg E, Heyn H. The emerging landscape of spatial profiling technologies. *Nature Rev Genet*. 2022;23(12):741–759.
- Kargl J, Busch SE, Yang GHY, et al. Neutrophils dominate the immune cell composition in non-small cell lung cancer. *Nature Commun*. 2017;8, 14381.
- Li R, Lin Y, Wang Y, et al. Characterization of the tumor immune microenvironment in lung squamous cell carcinoma using imaging mass cytometry. *Front Oncol*. 2021;11, 620989.
- van Maldegem F, Valand K, Cole M, et al. Characterisation of tumour microenvironment remodelling following oncogene inhibition in preclinical studies with imaging mass cytometry. *Nature Commun*. 2021;12(1):5906.
- Jackson EL, Willis N, Mercer K, et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev*. 2001;15(24):3243–3248.
- Jackson EL, Olive KP, Tuveson DA, et al. The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res*. 2005;65(22):10280–10288.
- Chen F, Byrd AL, Liu J, et al. Polycomb deficiency drives a FOXP2-high aggressive state targetable by epigenetic inhibitors. *Nature Commun*. 2023;14(1):336.

30. Chen F, Liu J, Song X, et al. EZH2 inhibition confers PIK3CA-driven lung tumors enhanced sensitivity to PI3K inhibition. *Cancer Lett.* 2022;524: 151–160.
31. Chen F, Liu J, Flight RM, et al. Cellular origins of EGFR-driven lung cancer cells determine sensitivity to therapy. *Adv Sci (Weinh).* 2021;8(22): e2101999.
32. Ji H, Ramsey MR, Hayes DN, et al. LKB1 modulates lung cancer differentiation and metastasis. *Nature.* 2007;448(7155):807–810.
33. Kargl J, Zhu X, Zhang H, et al. Neutrophil content predicts lymphocyte depletion and anti-PD1 treatment failure in NSCLC. *JCI Insight.* 2019;4(24).
34. Kurebayashi Y, Emoto K, Hayashi Y, et al. Comprehensive immune profiling of lung adenocarcinomas reveals four immunosubtypes with plasma cell subtype a negative indicator. *Cancer Immunol Res.* 2016;4(3):234–247.
35. Lee HE, Luo L, Kroneman T, et al. Increased plasma cells and decreased B-cells in tumor infiltrating lymphocytes are associated with worse survival in lung adenocarcinomas. *J Clin Cell Immunol.* 2020;11(1).
36. Colling R, Pitman H, Oien K, et al. Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *J Pathol.* 2019;249(2):143–150.
37. Acs B, Hartman J. Next generation pathology: artificial intelligence enhances histopathology practice. *J Pathol.* 2020;250(1):7–8.