

## Computational immune profiling in lung adenocarcinoma reveals reproducible prognostic associations with implications for immunotherapy

Frederick S. Varn<sup>a</sup>, Laura J. Tafe<sup>b,c,d</sup>, Christopher I. Amos<sup>d,e</sup>, and Chao Cheng<sup>a,d,e</sup>

<sup>a</sup>Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA; <sup>b</sup>Department of Pathology and Laboratory Medicine, Geisel School of Medicine at Dartmouth, Hanover, NH, USA; <sup>c</sup>Department of Pathology and Laboratory Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA; <sup>d</sup>Norris Cotton Cancer Center, Lebanon, NH, USA; <sup>e</sup>Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Hanover, NH, USA

### ABSTRACT

Non-small cell lung cancer is one of the leading causes of cancer-related death in the world. Lung adenocarcinoma, the most common type of non-small cell lung cancer, has been well characterized as having a dense lymphocytic infiltrate, suggesting that the immune system plays an active role in shaping this cancer's growth and development. Despite these findings, our understanding of how this infiltrate affects patient prognosis and its association with lung adenocarcinoma-specific clinical factors remains limited. To address these questions, we inferred the infiltration level of six distinct immune cell types from a series of four lung adenocarcinoma gene expression datasets. We found that naive B cell, CD8+ T cell, and myeloid cell-derived expression signals of immune infiltration were significantly predictive of patient survival in multiple independent datasets, with B cell and CD8+ T cell infiltration associated with prolonged prognosis and myeloid cell infiltration associated with shorter survival. These associations remained significant even after accounting for additional clinical variables. Patients stratified by smoking status exhibited decreased CD8+ T cell infiltration and altered prognostic associations, suggesting potential immunosuppressive mechanisms in smokers. Survival analyses accounting for immune checkpoint gene expression and cellular immune infiltrate indicated checkpoint protein-specific modulatory effects on CD8+ T cell and B cell function that may be associated with patient sensitivity to immunotherapy. Together, these analyses identified reproducible associations that can be used to better characterize the role of immune infiltration in lung adenocarcinoma and demonstrate the utility in using computational approaches to systematically characterize tissue-specific tumor-immune interactions.

### ARTICLE HISTORY

Received 25 August 2017  
Revised 13 December 2017  
Accepted 15 January 2018

### KEYWORDS


genomics; immunotherapy; Immunology; lung adenocarcinoma; survival

### Introduction

The immune system plays a broad role in shaping tumor growth and development. The presence of immune infiltrate in the tumor microenvironment is closely correlated with patient prognosis in numerous cancer types, with infiltrate from some cells, such as cytotoxic CD8+ T cells commonly linked to prolonged survival, while other cells, such as immunosuppressive T-regulatory cells and certain myeloid cells, associated with a shorter survival time.<sup>1</sup> Tumors can affect the behavior of different cell types to inhibit an otherwise effective adaptive immune response by recruiting immunosuppressive cells to the microenvironment or expressing immune checkpoint proteins.<sup>2</sup> There has thus been great interest in targeting these inhibitory mechanisms to treat cancer. Blockade of immune checkpoint proteins, including PD-1/PD-L1 and CTLA-4 has shown promise in several cancer types, reducing tumor burden and prolonging patient survival.<sup>3</sup> However, the success of these approaches has been tempered by a highly heterogeneous response rate across the patient population, as well as between tissues, indicating the need for tissue-specific studies into how the immune system interacts with the tumor microenvironment.

Lung adenocarcinoma, a type of non-small cell lung cancer (NSCLC), is the most common cancer of the lung and is one of the leading causes of cancer-related death in the United States.<sup>4</sup> Compared to other cancer types, lung adenocarcinoma has been well-characterized as having a high level of immune infiltration.<sup>5,6</sup> This infiltration makes it a strong candidate for immunotherapeutic approaches. However, early studies measuring the effectiveness of anti-PD-1 in NSCLC have revealed response rates ranging from 18% to 45%.<sup>7,8</sup> These highly heterogeneous response rates underscore the need for more a more thorough understanding of the tumor-immune interactions specific to lung adenocarcinoma. Early attempts to functionally characterize these interactions through survival analyses have had mixed results leading to inconsistent conclusions. Generally, it has been found that an intense lymphocytic infiltrate is associated with prolonged prognosis in NSCLC.<sup>9</sup> However at the cellular level, CD8+ and CD4+ T cells have been linked to both a protective effect and no association with prognosis,<sup>10</sup> while B cell infiltration has been associated with prolonged survival,<sup>11</sup> poor survival,<sup>12</sup> and has been found to have no association.<sup>11</sup> Beyond adaptive immune cells,

**CONTACT** Chao Cheng  [chao.cheng@dartmouth.edu](mailto:chao.cheng@dartmouth.edu)  HB7937, Rubin 701, Lebanon NH, 03756.

 Supplemental data for this article can be accessed on the [publisher's website](#).

© 2018 Frederick S. Varn, Laura J. Tafe, Christopher I. Amos, and Chao Cheng. Published with license by Taylor & Francis Group, LLC  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

infiltration from several innate cells has been studied, with macrophages linked to poor survival<sup>13</sup> and mature dendritic cells tied to prolonged survival.<sup>14</sup> Systematic analyses spanning multiple independent lung adenocarcinoma datasets may allow for a greater consensus understanding of each cell's role on prognosis in lung adenocarcinoma and NSCLC.

Recently, several computational approaches have made it possible to systematically infer immune infiltration levels in large-scale cancer patient datasets using gene expression information.<sup>5, 15-19</sup> These methods vary in their approaches and outputs, with some providing relative immune cell infiltration levels that can be compared between patients and others determining the composition of the immune cell infiltrate present in each patient's tumor. We have developed a method that infers immune infiltration from tumor gene expression data by examining the distribution of immune cell-specific genes throughout a patient's ranked gene expression profile.<sup>6</sup> This method calculates transcriptome-wide specificity weights for different immune cell types by comparing the differential expression level of each gene in a given cell's transcriptome to that of all other cells in a reference immune cell gene expression matrix. The dataset chosen for this approach, developed by the Immunological Genome Project (ImmGen) contains over 200 murine immune cell gene expression profiles capturing multiple steps in development throughout the hematopoietic hierarchy.<sup>20</sup> The comprehensive nature of this dataset makes it ideal for determining immune cell-specificity weights, as it contains a diverse collection of immune cell phenotypes found in the tumor microenvironment, as well as a series of highly proliferative or dedifferentiated progenitor cells that can be used to normalize against tumor-specific gene expression programs. Furthermore, each of the profiles in this dataset have been collected under the same protocol using the same microarray platform, removing the batch effects and platform-specific artifacts that could potentially confound infiltration analyses. While transcriptional differences exist between murine and human immune cells, previous studies have shown that murine and human transcriptomes share a high degree of global similarity, and that the expression of lineage-specific genes in analogous cell types is conserved between species.<sup>21</sup> In concordance with this finding, ImmGen-based infiltration scores have been found to be closely associated with scores derived using human reference immune cell profiles.<sup>22</sup>

In this study, we further refine this approach to perform a tissue-specific analysis with the goal of clarifying the function of different immune cells in the context of lung adenocarcinoma. To accomplish this, we pool the ImmGen dataset into six consensus immune cell signatures that have been optimized through a series of benchmarking analyses to detect infiltration signals from lung adenocarcinoma gene expression data. We then correlate the resulting immune cell infiltration scores with patient survival in four independent lung adenocarcinoma datasets to identify reproducible associations and characterize the most relevant cell types for patient prognosis. To further understand the prognostic role of different immune infiltrates, we use multivariate Cox regression to identify the cells that are independent prognostic factors. In addition to the survival analyses, we examine the effect smoking has on immune infiltration and function by comparing the infiltration scores of

different cell types between smokers and non-smokers. Finally, we conclude our analysis by integrating immune infiltration and immune checkpoint gene expression information in a multi-class survival analysis to determine how the prognostic effect of immune cell infiltration is affected by the presence of immune checkpoint proteins. Together, our results further validate the role immune infiltration plays on tumor growth and development in lung adenocarcinoma, while providing novel insights into the behavior of different immune cells in the context of immune checkpoint gene expression.

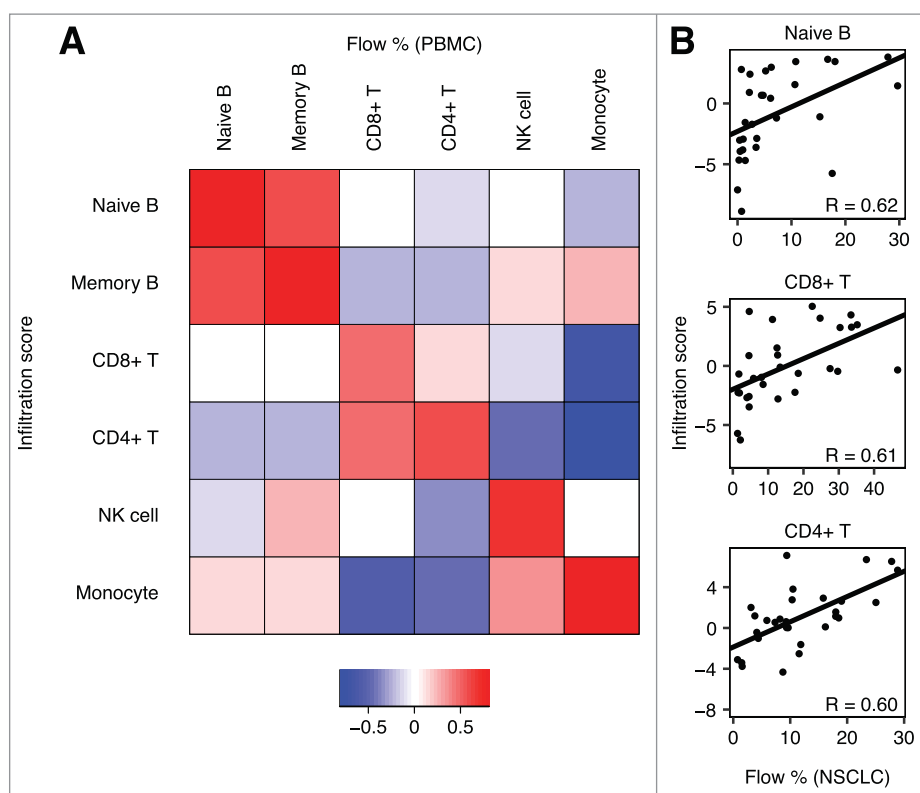
## Results

### *Consensus immune signatures detect immune infiltration in lung adenocarcinoma*

Our method traditionally uses a series of over 200 transcriptome-wide gene-specificity signatures derived from the ImmGen dataset to calculate immune infiltration scores from bulk tumor gene expression data. Many of these signatures are redundant as they come from genetically similar cell types. Thus, it is possible to further improve the output of this method by reducing the dataset down to a smaller number of immune signatures representing distinct cell types. To accomplish this, we collapsed the ImmGen dataset into a series of consensus immune cell signatures made up of the ImmGen signatures that were best associated with flow cytometry measurements of naive B cells, memory B cells, CD8+ T cells, CD4+ T cells, NK cells, and myeloid cells in peripheral blood mononuclear cell (PBMC) mixture and NSCLC tumor settings. In PBMC mixtures, the scores produced by each consensus signature were highly correlated with the flow cytometry fraction of the cell types they represented (mean SCC = 0.63), and lowly correlated with the fraction of the cell types they did not represent (mean SCC = -0.06), indicating high sensitivity and specificity (Fig. 1A). Furthermore, in NSCLC tumors, most of the consensus signatures were strongly associated with the flow cytometry fraction of their respective cell type, indicating they were not confounded by lung adenocarcinoma-specific gene expression signals (Fig. 1B). We thus reasoned that these six consensus signatures could offer high fidelity estimates of immune infiltration from lung adenocarcinoma gene expression data.

### *Immune infiltration in lung adenocarcinoma predicts patient prognosis*

We applied these six signatures to calculate immune infiltration scores in a lung adenocarcinoma dataset generated by Okayama et al (n = 226).<sup>23</sup> In this dataset, expression-based measures of tumor purity as calculated by the ESTIMATE algorithm<sup>16</sup> were negatively associated with immune infiltration, for all cell types, indicating that our infiltration scores were picking up expression signals from the tumor microenvironment (SCC range = -0.22 to -0.51). To examine the cellular immune infiltration patterns in these patients, we correlated each immune cell's infiltration scores against one another. Infiltration scores from naive B,



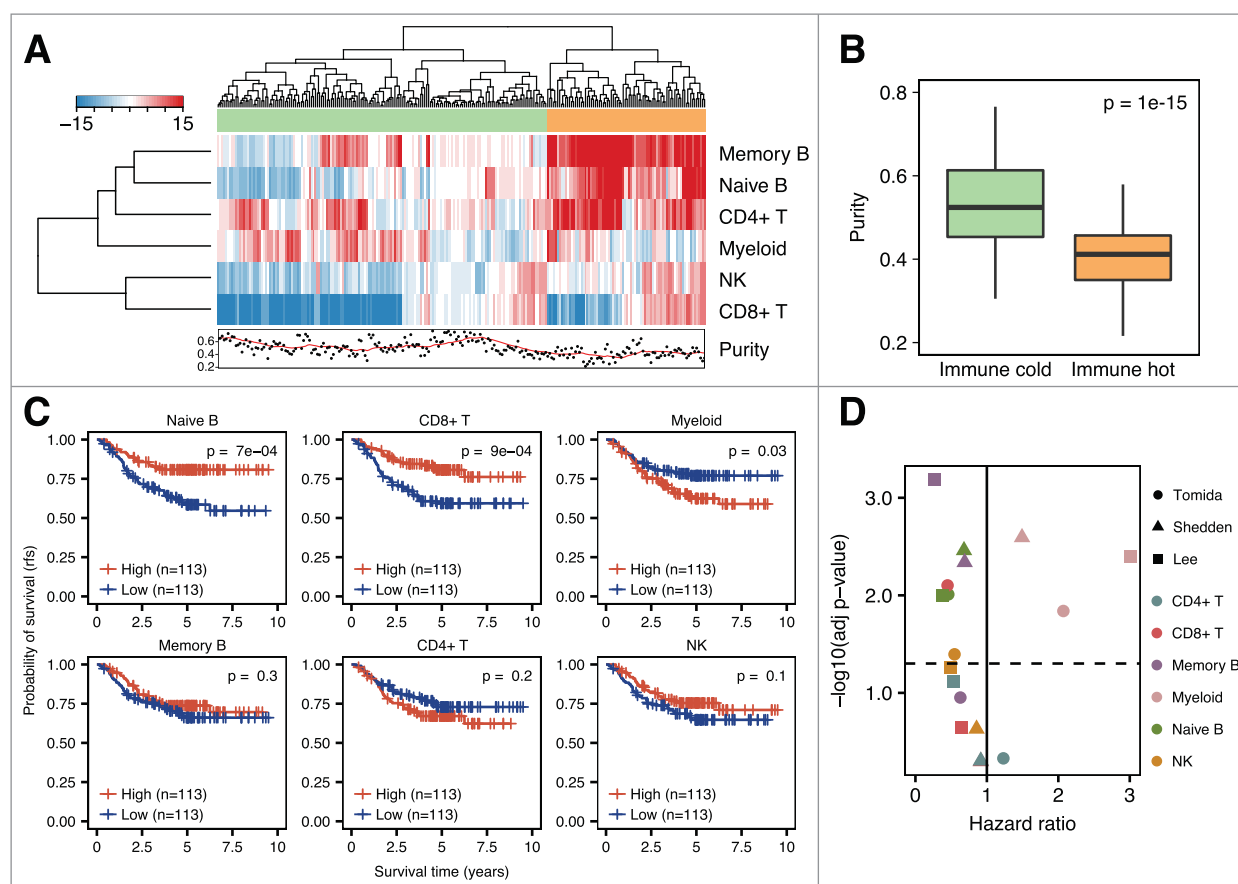
**Figure 1.** Validation of consensus immune signatures. a Heatmap depicting the correlation (Spearman) between infiltration scores calculated from the six consensus immune signatures and flow cytometry percentages for all cell types in 20 peripheral blood mononuclear cell mixtures. b Scatterplots of infiltration scores calculated from the naive B cell, CD8+ T cell, and CD4+ T cell consensus immune signatures and flow cytometry fractions from a series of 29 NSCLC tumors. The flow cytometry percentages measure the fraction live cells that were CD19+ B cells, CD8+ T cells, and CD4+ T cells, respectively. For all analyses, the Spearman correlation coefficient is displayed.

memory B, CD8+ T, CD4+T, and NK cells primarily exhibited positive correlations with one another, suggesting patterns of co-infiltration, while myeloid cells were negatively associated with the other cell types (Supplementary Fig. S1). To further examine these patterns, we hierarchically clustered the patients using their infiltration scores (Fig. 2A). Splitting the patients into two groups based on their infiltration patterns revealed an immune-hot and immune-cold cluster, where patients in the immune-hot cluster had significantly lower tumor purity values compared to those in the immune-cold cluster, indicating higher levels of immune infiltration ( $p = 1e-15$ ; Wilcoxon sum-rank test; Fig. 2B). This clustering pattern mirrored findings from a previous study<sup>24</sup> and supported the idea that lung adenocarcinoma patients can be broadly classified into two distinct immunophenotypes.

To investigate how the individual immune cells affect patient survival, we performed two class-survival comparisons by stratifying patients into high- and low-infiltration groups based on the median infiltration score of each immune cell type (Fig. 2C). Interestingly, we found that patients with high naive B and CD8+ T cell infiltration experienced significantly longer relapse-free survival times compared with their low-infiltration counterparts (log-rank  $p = 7e-4$  and  $9e-4$ , respectively). Conversely, patients with high myeloid infiltration exhibited significantly shorter relapse-free survival times compared to low myeloid infiltrate patients (log-rank  $p = 0.03$ ). Infiltration from the

remaining cell types did not show significant survival associations. These results were further validated using univariate Cox proportional hazards models that used either high/low infiltrate classification or continuous infiltration score as the variable (Supplementary Table S1). To confirm that these associations were not specific to one dataset, we applied our method to three additional lung adenocarcinoma datasets by Tomida et al ( $n = 117$ ),<sup>25</sup> Shedden et al ( $n = 442$ ),<sup>26</sup> and Lee et al ( $n = 63$ )<sup>27</sup> and split samples into high and low infiltration groups based on their median infiltration scores for each cell type. We then calculated each cell's association with patient survival by fitting high/low classification for each cell type in a univariate Cox proportional hazards model (Fig. 2D). Naive B cells and CD8+ T cells were significantly associated with prolonged patient survival in at least one of the additional datasets tested, while myeloid cells were associated with worse prognosis in all three of the additional datasets. In addition to these reproducible associations, we found that memory B cells were associated with improved prognosis in two out of the three additional datasets, while NK cells were associated with improved prognosis in one of the three datasets.

The survival associations we observed for CD8+ T cells, naive B cells, and myeloid cells, while reproducible, did not account for clinical factors that could potentially confound our results. For instance, when stratifying patients by different clinical factors, we found that infiltration levels for all three cell types varied depending on a patient's *EGFR*



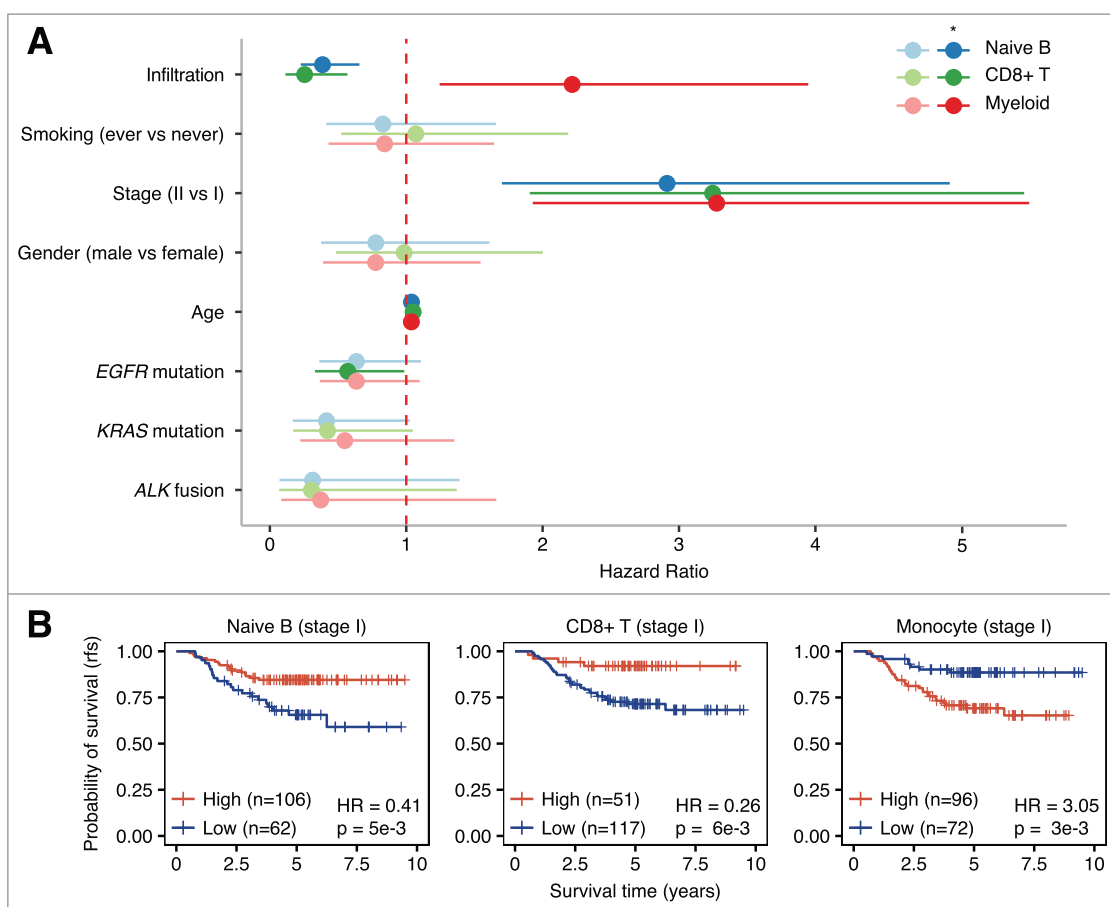
**Figure 2.** Patterns of immune infiltration in lung adenocarcinoma. a Heatmap depicting the infiltration scores for six immune cell types in the Okayama et al dataset. Bottom sidebar indicates the tumor purity level for each sample, as calculated using ESTIMATE. Red line is the rolling average purity level for the last 20 samples going left to right. Top sidebar indicates the grouping of the samples based on hierarchical clustering. b Boxplot indicating the difference in tumor purity level between each two groups of patients determined from hierarchical clustering. P-value was calculated using the two-sided Wilcoxon sum-rank test. c Kaplan-Meier plots depicting the relapse-free survival (rfs) distributions for patients with high (red) and low (blue) immune infiltration scores for the noted cell types. In Kaplan-Meier plots, p-values were calculated using the log-rank test and vertical hash marks indicate censored data. d Volcano plot depicting the  $-\log_{10}$  adjusted p-value and the hazard ratio from Cox proportional hazards models inputted with high/low infiltration score classification for the noted cell types. Colors indicate cell types used in the model and shapes indicate the dataset the analysis was performed in.

mutation status, (Supplementary Table S2). To determine whether each cell's prognostic associations were independent of these variables, we applied three multivariate Cox proportional hazards models to the Okayama et al dataset that included the high/low infiltrate classification for a given cell type, smoking status, tumor stage, gender, age, *EGFR* mutation status, *KRAS* mutation status, and *ALK* fusion status as covariates (Fig. 3A). After adjusting for these covariates, infiltration from naive B cells, CD8+ T cells, and myeloid cells each remained significant ( $p = 5e-4$ ,  $8e-4$ , and  $7e-3$ ; HR = 0.39, 0.25, and 2.22, respectively). We followed up this analysis by performing two-class survival comparisons between high- and low- infiltration groups in stage I lung adenocarcinoma patients, as these tumors have been associated with high recurrence rates following surgical resection (Fig. 3B).<sup>28, 29</sup> For all three cell types, infiltration was significantly associated with relapse-free survival, with naive B cell and CD8+ T cell infiltration associated with prolonged patient survival (log-rank  $p = 5e-3$  and  $6e-3$ , HR = 0.41, 0.26, respectively) and myeloid infiltration with shorter survival (log-rank  $p = 3e-3$ , HR = 3.05). These results suggested that the composition of the tumor micro-environment may be a useful indicator in predicting

recurrence and determining treatment strategies in early-stage lung adenocarcinomas.

### Smoking status is associated with reduced immune infiltrate

Smoking has been associated with a poor prognosis in lung adenocarcinoma,<sup>30</sup> but its effects on immune infiltration levels remain unclear. To characterize how immune infiltration differed between ever-smokers and never-smokers, we compared the immune infiltration scores between the two groups in the Okayama, Tomida, and Shedden et al datasets (Fig. 4A; Supplementary Table S3). In smokers, CD8+ T cell and NK cell infiltration were significantly lower in two of the three independent datasets, while naive B cells were significantly lower in one dataset and myeloid cells were significantly higher in one dataset. In the Shedden et al dataset there were no immune cells whose infiltration scores significantly differed based on smoking status. We next tested how smoking can affect immune cell function by analyzing patient survival distributions after double stratification by smoking status and high/low immune infiltrate (Fig. 4B). In ever-smokers, increased



**Figure 3.** Multivariate analysis of immune infiltration-survival associations. a Forest plot depicting the hazard ratios and p-values for three different multivariate Cox proportional hazards models fit to the Okayama et al dataset. Colors indicate the immune cell inputted into the model. Darker colors indicate significant associations in the model ( $p < 0.05$ ), while lighter colors indicate insignificant associations ( $p > 0.05$ ). Points indicate the hazard ratio, with lines depicting 95% confidence interval. b Kaplan-Meier plots depicting the relapse-free survival (rfs) distributions of stage I lung adenocarcinoma patients from the Okayama et al dataset with high (red) and low (blue) immune infiltration scores for the noted cell types. P-values were calculated using the log-rank test, hazard ratios (HR) were calculated using a univariate Cox proportional hazards model using high and low infiltration classification, and vertical hash marks indicate censored data.

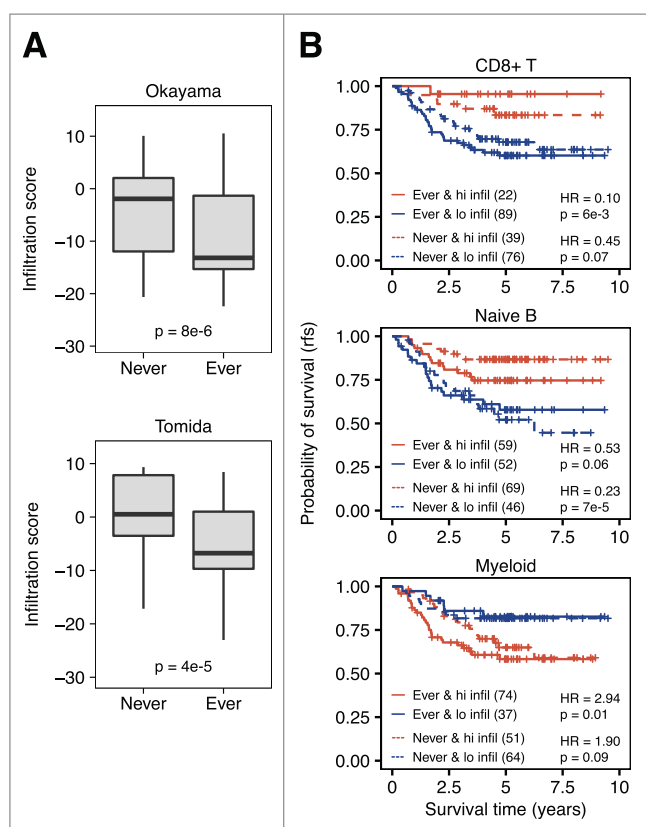
CD8+ T cell infiltration resulted in significantly longer relapse-free survival time (log-rank  $p = 6e-3$ , HR = 0.10), while the trend in never-smokers was protective but insignificant despite a comparable sample size. Myeloid cell infiltration was associated with significantly poorer relapse-free survival in ever-smokers (log-rank  $p = 0.01$ , HR = 2.94), but there was again no significant difference in prognosis in never-smokers. These trends were reversed for naive B cells, where never-smokers with high infiltrate had significantly longer relapse-free survival time than patients with low infiltrate (log-rank  $p = 7e-5$ , HR = 0.53) and infiltration in ever-smokers was protective but insignificant. Together, these findings suggest that smoking may affect both the level of infiltrate and the extent to which the infiltrate can properly function.

### Prognostic analyses of immune cell infiltration and checkpoint gene expression suggest immunomodulatory interactions

Through our survival analyses, we found that CD8+ T cells were associated with prolonged patient survival, even when considering other clinical variables. However, the expression of the immune checkpoint proteins CTLA-4, PD-1, and

PD-L1 in the tumor microenvironment are known to be involved in immunosuppression, which could potentially lead to poorer patient outcome.<sup>31</sup> We thus examined whether the expression of these proteins was associated with noticeable survival differences in patients with similar levels of CD8+ T cell infiltration. Using the Okayama et al dataset, we compared the survival distributions of four groups of patients stratified based on their levels of CD8+ T cell infiltration and their expression of either *CTLA4* or *PDCD1*, which encode CTLA-4 and PD-1, respectively (Fig. 5), as well as their expression of *CD274*, which encodes the PD-1 ligand, PD-L1 (Supplementary Fig. S2). In all analyses, at least one group of patients had significantly better survival than the other groups (log-rank  $p = 3e-3$ , 0.01, and 0.01, for *CTLA4*, *PDCD1*, and *CD274* respectively). Interestingly, patients with high CD8+ T cell infiltration and low *CTLA4* expression trended toward prolonged survival relative to patients with high CD8+ T cell infiltration and high *CTLA4* expression (log-rank  $p = 0.06$ ). This was not the case when stratifying based on *PDCD1* or *CD274* expression. To examine whether other cell types could be affected by immune checkpoint gene expression, we repeated this analysis using infiltration scores from the other five cell types. This revealed that patients stratified





**Figure 4.** Interactions between immune infiltration score and smoking status. a Boxplots depicting the difference in infiltration between never-smokers (Never) and ever-smokers (Ever) in two datasets. P-values were calculated using a two-sided Wilcoxon sum-rank test. b Kaplan-Meier plots depicting the relapse-free survival (rfs) distributions in patients from the Okayama et al dataset stratified by high (red) and low (blue) immune infiltration scores for the noted cell types as well as ever-smoking (solid line) and never-smoking (dotted line). P-values were calculated using the log-rank test, hazard ratios (HR) were calculated using a univariate Cox proportional hazards model using high and low infiltration classification, and vertical hash marks indicate censored data.

based on naive B cell infiltration and immune checkpoint gene expression exhibited similar patterns to those of CD8+ T cells (Fig. 5, Supplementary Fig. S2). Together, these results indicate that while patient survival is primarily driven by immune infiltration, the prognostic effect of this infiltration may be modulated by the proteins encoded by immune checkpoint inhibitor genes such as *CTLA4*.

## Discussion

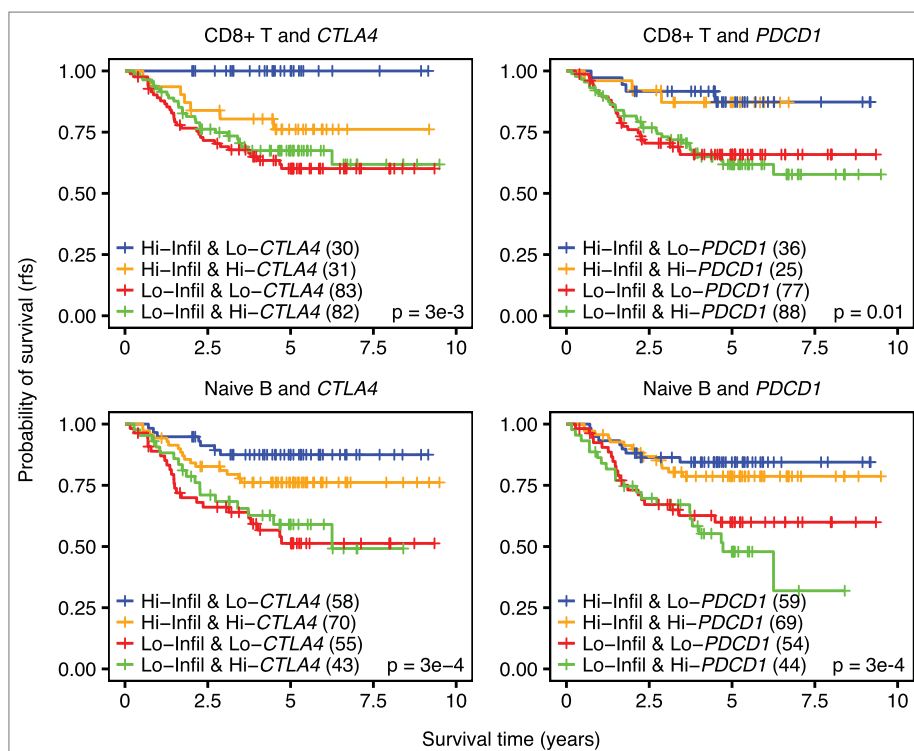
Immune infiltration in lung adenocarcinoma is a useful prognostic factor and may be a potential biomarker of immunotherapy response. Genomics-based approaches to inferring immune infiltration enable high-throughput immune profiling over many samples, increasing the power to detect these associations. Here, we have applied our computational method to multiple lung adenocarcinoma datasets to better define the role different immune cells play on patient survival. Using this method, we found reproducible associations linking infiltration from naive B cells and CD8+ T cells to prolonged relapse-free survival and infiltration from myeloid cells to decreased relapse-free survival. In addition, we identified potential interactions between smoking behavior, immune

infiltration, and patient survival, and found associations suggesting that the expression of some immune checkpoint genes may modulate the prognostic effects of certain immune cell infiltrates. Together, our results demonstrate the utility of using computational approaches to define the tumor microenvironment and introduce new potential methods for identifying immune-based biomarkers.

When inferring immune infiltration, our method uses the ImmGen dataset as a reference to determine the immune-related genes for each cell type. This dataset contains gene expression profiles from over 200 different murine hematopoietic cells and as a result, our method outputs a set of immune infiltration scores for each one of these reference cell types. Many of these scores detect redundant signals in the microenvironment, creating a need for us to simplify the output of our method. Furthermore, some of these signatures could be detecting lung tumor cell-specific expression signals, as the ImmGen dataset does not include cancer cells to normalize against. To address these issues, we created a series of consensus immune signatures representing naive B cells, memory B cells, CD8+ T cells, CD4+ T cells, NK cells, and myeloid cells that sensitively and specifically captured gene expression signals of the cells they represented in PBMC mixture experiments and were validated in a gold-standard dataset consisting of NSCLC tumors profiled by flow cytometry.

Previous efforts to characterize the prognostic effect of immune infiltration in lung adenocarcinoma have revolved around using immunohistochemistry (IHC) and flow cytometry.<sup>10-14, 24, 32, 33</sup> These approaches have substantially advanced our understanding of the lung adenocarcinoma immune response, especially regarding how the spatial distribution of certain immune cells in the tumor can affect patient prognosis. However, these approaches are time consuming and the tissue biopsies necessary for performing these analyses are not as readily available compared to other types of data. By using gene expression data to study immune infiltrate, we were able to quickly identify reproducible prognostic associations for many different immune cell types. Our analysis revealed that naive B cell and CD8+ T cell infiltration was associated with prolonged patient survival, while myeloid cell infiltration was predictive of shorter survival. These associations remained significant even after adjusting for multiple covariates, including stage, smoking status, *EGFR* and *KRAS* mutations, and *ALK* fusions. Furthermore, we found that in early-stage tumors specifically, the prognostic associations from all three cell types mirrored what was found in the dataset as a whole. These associations are in line with a previous study reporting that reduced CD8+ T cell infiltration and altered myeloid cell activity are observed beginning in stage I lung adenocarcinoma tumors<sup>34</sup> and suggest that patients with tumors containing high myeloid content or low lymphocytic infiltrate may need to be treated more aggressively. Understanding the degree to which these cells are functioning in the lung adenocarcinoma microenvironment throughout development may be able to provide additional insights into how the tumor and immune cells shape other as they evolve.

The effects of certain clinical factors, such as oncogenic mutation status and smoking behavior, on immune cell infiltration and function are currently under investigation. Our study found that *EGFR* mutations were associated with differing



**Figure 5.** Effect of immune checkpoint gene expression and immune infiltration on patient survival. Kaplan-Meier plots depicting relapse-free survival (rfs) distributions between patients with high CD8+ T cell infiltration (top) or naive B cell infiltration (bottom) scores and low checkpoint gene expression (blue), high infiltration and high expression (orange), low infiltration and low gene expression (red), and low infiltration and high gene expression (green). High/low immune infiltration cutoffs were based on an infiltration score greater than or less than 0, while high/low immune checkpoint gene expression cutoffs were made using the median. P-values were calculated using the log-rank test and vertical hash marks indicate censored data.

levels of CD8+ T cell, naive B cell, and myeloid cell infiltration, which conflicted with a smaller multi-parametric immune profiling study of 51 NSCLC tumors.<sup>24</sup> This same study also reported no association between smoking behavior and immune infiltration, while we found evidence in two independent datasets that ever-smokers exhibit significantly lower levels of CD8+ T cell infiltration compared to never-smokers. These discrepancies highlight the need for systematic analyses spanning multiple datasets as dataset-dependent factors such as sample size and demographic makeup may influence the discovery of new associations. Our study's findings regarding smoking behavior are especially noteworthy as we found that the prognostic associations for CD8+ T cell and myeloid cells were only significant in ever-smokers compared to never-smokers despite the sample size being similar between the two groups. Furthermore, B cell infiltration was only associated with survival in never-smokers, mirroring the findings of a histology-based study examining the relationship between immune infiltration and smoking behavior.<sup>35</sup> If smoking behavior truly does influence immune cell function, it will be important to take this factor into account in future studies of lung adenocarcinoma.

Immune checkpoint blockade therapy has shown promise in several cancer types, including NSCLC. However, only a small subset of NSCLC patients has proven to be responsive to these therapies.<sup>7, 8, 36</sup> To better characterize how these immune checkpoint proteins interact with cellular immune infiltrate, we compared the survival distributions of patients stratified into groups based on their levels

of infiltration and immune checkpoint gene expression. When examining this relationship using the gene encoding CTLA-4, we found that patients with high CD8+ T cell or naive B cell infiltration and low checkpoint gene expression had the best survival time of the four groups, including those with high CD8+ T cell infiltrate and high gene expression. These results suggest that CTLA-4 abundance may be associated with an immunosuppressive phenotype that modulates CD8+ T cell or naive B cell function, and that the patients most likely to respond to anti-CTLA-4 therapy are those with high CD8+ T cell infiltration and high checkpoint gene expression. Interestingly, these relationships were not present for either cell type when stratifying patients based on expression of the genes encoding PD-1 and PD-L1. These associations mirror studies in melanoma, which have noted that CTLA-4 expression and cytolytic cell infiltrate is associated with anti-CTLA-4 response,<sup>37</sup> while responders to anti-PD-1 therapy do not exhibit differential expression of PD-1, PD-L1, or genes associated with CD8+ T cell infiltrate.<sup>38</sup> However, these findings conflict with studies linking IHC-based PD-L1 levels to anti-PD-1 response.<sup>39, 40</sup> The discrepancy between our findings and those using PD-L1 IHC suggests that the location of the cells expressing PD-L1 is an important factor in determining response rate, as expression-based methods cannot provide this type of information. Alternatively, these inconsistencies could be due to a poor correlation between gene expression and protein abundance measures. Going forward, it will be important to better characterize

how the abundance and location of these checkpoint proteins affect immune cell function as this understanding may aid the development of biomarkers predicting immunotherapy response in lung adenocarcinoma.

In conclusion, we have presented a computational analysis that further characterizes the prognostic landscape of immune infiltration in lung adenocarcinoma. Using high-throughput gene expression-based analyses, we inferred infiltration levels for six distinct immune cell types and found that infiltration from three, naive B cells, CD8+ T cells, and myeloid cells, was significantly associated with patient survival, even when adjusting for several covariates. Using these infiltration scores, we showed that smoking can decrease immune infiltration levels and may modulate the prognostic effect of certain cell types. Lastly, we characterized how immune checkpoint gene expression modulates the prognostic effect of CD8+ T cells and naive B cells in the tumor microenvironment. Our results present new biomarkers predicting patient prognosis and provide insights into potential biomarkers of immunotherapy response. As computational approaches continue to mature and datasets detailing immunotherapy response are released, we are hopeful that high-throughput immune inference approaches can be used to improve precision medicine in a variety of cancer types.

## Methods

### Datasets

Lung cancer datasets by Okayama et al,<sup>23</sup> Tomida et al,<sup>25</sup> Shedden et al,<sup>26</sup> and Lee et al<sup>27</sup> were downloaded from the Gene Expression Omnibus (GEO) under accession numbers GSE31210, GSE13213, GSE68465, and GSE8894, respectively. PBMC gene expression data and associated flow cytometry data were obtained from GEO under accession number GSE65133 and a prior publication.<sup>17</sup> NSCLC Nanostring gene expression data and associated flow cytometry data were obtained from GEO under accession number GSE84797 and a prior publication.<sup>24</sup> Hematopoietic gene expression data from ImmGen was obtained in its raw form (.CEL files, Affymetrix Mouse Gene 1.0 ST Array) from GEO under accession number GSE15907 in October 2015. Raw data was background corrected using Robust Microarray Analysis and then quantile normalized. Probesets for hematopoietic profiles were fitted to a multichip linear model using the R 'affy' library's "expresso" function.<sup>41</sup> The probe with the highest average intensity across all cell types was used for each gene, and each murine transcript was matched to human transcript using the gene symbol.

### Calculation of immune infiltration score

To calculate immune infiltration scores, we used our previously developed binding association with sorted expression (BASE) algorithm.<sup>42</sup> To infer immune infiltration on a group of patients, this algorithm requires two types of data, immune cell gene-specificity weight profiles and the patients' gene expression data. Immune cell weight profiles are created from the normalized ImmGen gene expression profiles and represent the differential up- or down-regulated state of each gene in each ImmGen cell type's expression profile relative to the

rest of the ImmGen dataset. Weight profile calculation has been described previously.<sup>6</sup> BASE orders a given patient's gene expression profile from high to low and then uses each ImmGen cell's weight profiles to weigh the patient's gene expression values. BASE then calculates two running sums, one representing the cumulative distribution of the patient's weighted gene expression values (foreground function) and another representing the cumulative distribution of the patient's complementary weighted (1-weight) gene expression values (background function). In the presence of a high amount of infiltrate from a given cell type, the foreground function increases quickly early on, as the highly-expressed genes in a patient's profile tend to also be the ones with high weights for a given immune cell, before plateauing later in the patient's ranked profile, while the background function does the opposite. The maximal absolute difference of these two functions represents immune infiltration level and, after a normalization procedure, results in the final immune infiltration score. Full details on the calculation and validation of the immune infiltration scores using BASE have been described previously.<sup>6</sup>

### Creation of consensus immune cell signatures

To increase the robustness and interpretability of our infiltration scores, we used the ImmGen reference profiles to create consensus immune cell signatures representing six distinct immune cell types: naive B cells, memory B cells, CD8+ T cells, CD4+ T cells, natural killer (NK) cells, and myeloid cells (Supplementary Table S4). In order to select the ImmGen signatures that were best suited for each cell type, we subjected each profile to a series of quality-filtering steps that measured the resulting infiltration scores' associations with tumor purity and flow cytometry fraction in PBMC and lung tumor settings. To begin, we calculated infiltration scores for each ImmGen signature in The Cancer Genome Atlas (TCGA)'s lung adenocarcinoma dataset and then removed all ImmGen signatures whose scores were positively correlated ( $R > -0.05$ ) with previously calculated consensus tumor purity estimates.<sup>43</sup> We then used each of the remaining cell types to generate infiltration scores for a series of gene expression profiles from PBMC mixtures whose composition was defined using flow cytometry measuring the percentages of naive B cells, memory B cells, CD8+ T cells, CD4+ T cells, NK cells, and monocytes (GSE65133).<sup>17</sup> For non-T cells, ImmGen profiles whose infiltration scores were correlated with a given cell's flow cytometry fraction at an SCC  $> 0.6$  and who were defined by ImmGen to be from the same lineage as the representative cell were selected to potentially be used in the creation of that cell type's consensus immune cell signature. For CD8+ and CD4+ T cells, whose transcriptional phenotypes in the tumor microenvironment differ from those of the naive cells used in the PBMC mixture experiments, we used a reduced correlation coefficient cutoff of 0.3. After filtering using PBMC criteria, we performed additional cell-specific filtering steps to ensure that each profile was well-suited to specifically measure infiltration of their respective cell type in the tumor microenvironment. To ensure greater discrimination between naive and memory B



cells, we required that the infiltration scores from the ImmGen cell types selected for these profiles exhibit flow cytometry correlations of  $> 0.6$  for either naive or memory B cells, but not both. For the CD8<sup>+</sup> T cell consensus signature, we filtered out ImmGen CD8<sup>+</sup> T cell profiles that were in the early stages of activation, as these cell types exhibit a high degree of proliferation-associated gene expression that could confound their resulting infiltration scores.<sup>44</sup> For the CD4<sup>+</sup> T cell consensus signature, we removed all ImmGen cell types whose correlation coefficients were greater when measuring CD8<sup>+</sup> T cell fraction than CD4<sup>+</sup> T cell fraction. For the monocyte correlations, where over 50 ImmGen cells met the PBMC filtering criteria, we selected the ImmGen profiles with the top 10 correlation coefficients to ensure maximum fidelity in measuring myeloid cell infiltration. The remaining ImmGen profiles were then subjected to a final benchmarking analysis, where each profile was applied to a dataset consisting of a series of NSCLC gene expression profiles measured on the Nanostring platform and paired flow cytometry fractions measuring the abundance of different cell types in the tumor microenvironment (GSE84797).<sup>24</sup> For B and T cells, all profiles whose scores were correlated with their respective cell type's flow cytometry fraction at a coefficient  $> 0.5$  were kept for the final consensus signatures. In myeloid cells, which have a diverse collection of cell surface markers, this threshold was lowered to 0.2. In NK cells, this step was skipped as only 1% of the total live cell events on average came from NK cells, making it difficult to detect NK-cell specific signals in this dataset. Once the profiles were chosen, the final consensus signatures were created by taking the mean gene expression profile of the ImmGen cell types in each group. The resulting six profiles were then median normalized against the overall ImmGen dataset, converted to immune cell weights, and input into BASE to calculate immune cell infiltration scores. A full list of the ImmGen cells chosen for each consensus signature, as well as their infiltration scores' flow cytometry and tumor purity correlations, is available (Supplementary Table S5).

### Survival analysis

For continuous univariate and multivariate survival analyses, infiltration scores and the appropriate clinical covariates were fit to a Cox proportional hazards model using the "coxph" function from the R "survival" package. Survival distributions for different cell types were visualized using Kaplan-Meier curves created by the "survfit" function from the R "survival" package. The median infiltration score was used to stratify patients into "high" and "low" infiltration groups when performing univariate two-class comparisons, while an infiltration score of 0 was used to stratify the two groups when performing multivariate analyses. Differences between the survival distributions in each Kaplan-Meier plot were calculated using a log-rank test through the "survdiff" function from the R "survival" package.

### Disclosure of potential conflicts of interests

The authors declare no potential conflicts of interest.

### Financial support

This study was supported by the American Cancer Society (IRG-82-003-30) and the National Center for Advancing Translational Sciences of the National Institutes of Health (KL2TR001088) (C. Cheng). F.S. Varn was additionally supported in part by the National Institute of General Medical Sciences of the National Institutes of Health (T32GM008704).

### Funding

HHS | NIH | National Center for Advancing Translational Sciences (NCATS), KL2TR001088, HHS | NIH | National Institute of General Medical Sciences (NIGMS), T32GM008704, American Cancer Society (ACS), IRG-82-003-30.

### References

- Pages F, Galon J, Dieu-Nosjean MC, Tartour E, Sautes-Fridman C, Fridman WH. Immune infiltration in human tumors: a prognostic factor that should not be ignored. *Oncogene*. 2010;29:1093–102. doi:10.1038/onc.2009.416.
- Kitamura T, Qian BZ, Pollard JW. Immune cell promotion of metastasis. *Nat Rev Immunol*. 2015;15:73–86. doi:10.1038/nri3789.
- Postow MA, Callahan MK, Wolchok JD. Immune Checkpoint Blockade in Cancer Therapy. *J Clin Oncol*. 2015;33:1974–82. doi:10.1200/JCO.2014.59.4358.
- About Non-Small Cell Lung Cancer. Atlanta, GA: American Cancer Society; 2017.
- Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. 2016;17:174. doi:10.1186/s13059-016-1028-7.
- Varn FS, Wang Y, Mullins DW, Fiering S, Cheng C. Systematic Pan-Cancer Analysis Reveals Immune Cell Interactions in the Tumor Microenvironment. *Cancer Res*. 2017;77:1271–82. doi:10.1158/0008-5472.CAN-16-2490.
- Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, Powderly JD, Carvajal RD, Sosman JA, Atkins MB, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med*. 2012;366:2443–54. doi:10.1056/NEJMoa1200690.
- Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015;348:124–8. doi:10.1126/science.aaa1348.
- Brambilla E, Le Teuff G, Marguet S, Lantuejoul S, Dunant A, Graziano S, Pirker R, Douillard JY, Le Chevalier T, Filipits M, et al. Prognostic Effect of Tumor Lymphocytic Infiltration in Resectable Non-Small-Cell Lung Cancer. *J Clin Oncol*. 2016;34:1223–30. doi:10.1200/JCO.2015.63.0970.
- Al-Shibli KI, Donnem T, Al-Saad S, Persson M, Bremnes RM, Busund LT. Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. *Clin Cancer Res*. 2008;14:5220–7. doi:10.1158/1078-0432.CCR-08-0133.
- Suzuki K, Kadota K, Sima CS, Nitadori J, Rusch VW, Travis WD, Sadelain M, Adusumilli PS. Clinical impact of immune microenvironment in stage I lung adenocarcinoma: tumor interleukin-12 receptor beta2 (IL-12Rbeta2), IL-7R, and stromal FoxP3/CD3 ratio are independent predictors of recurrence. *J Clin Oncol*. 2013;31:490–8. doi:10.1200/JCO.2012.45.2052.
- Kurebayashi Y, Emoto K, Hayashi Y, Kamiyama I, Ohtsuka T, Asamura H, Sakamoto M. Comprehensive Immune Profiling of Lung Adenocarcinomas Reveals Four Immunotypes with Plasma Cell Subtype a Negative Indicator. *Cancer Immunol Res*. 2016;4:234–47. doi:10.1158/2326-6066.CIR-15-0214.
- Welsh TJ, Green RH, Richardson D, Waller DA, O'Byrne KJ, Brading P. Macrophage and mast-cell invasion of tumor cell islets confers a marked survival advantage in non-small-cell lung cancer. *J Clin Oncol*. 2005;23:8959–67. doi:10.1200/JCO.2005.01.4910.

14. Dieu-Nosjean MC, Antoine M, Danel C, Heudes D, Wislez M, Poulot V, Rabbe N, Laurans L, Tartour E, de Chaisemartin L, et al. Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures. *J Clin Oncol.* 2008;26:4410–7. doi:10.1200/JCO.2007.15.0284.
15. Gaujoux R, Seighe C. CellMix: a comprehensive toolbox for gene expression deconvolution. *Bioinformatics.* 2013;29:2211–2. doi:10.1093/bioinformatics/btt351.
16. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612. doi:10.1038/ncomms3612.
17. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12:453–7. doi:10.1038/nmeth.3337.
18. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17:218. doi:10.1186/s13059-016-1070-5.
19. Cesano A. nCounter(R) PanCancer Immune Profiling Panel (NanoString Technologies, Inc., Seattle, WA). *J Immunother Cancer.* 2015;3:42. doi:10.1186/s40425-015-0088-7.
20. Jojic V, Shay T, Sylvia K, Zuk O, Sun X, Kang J, Best AJ, Knell J, Goldrath A, Jojic V, et al. Identification of transcriptional regulators in the mouse immune system. *Nat Immunol.* 2013;14:633–43. doi:10.1038/ni.2587.
21. Shay T, Jojic V, Zuk O, Rothamel K, Puyraimond-Zemmour D, Feng T, Wakamatsu E, Benoist C, Koller D, Regev A, et al. Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc Natl Acad Sci U S A.* 2013;110:2946–51. doi:10.1073/pnas.1222738110.
22. Varn FS, Andrews EH, Mullins DW, Cheng C. Integrative analysis of breast cancer reveals prognostic haematopoietic activity and patient-specific immune response profiles. *Nat Commun.* 2016;7:10248. doi:10.1038/ncomms10248.
23. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, Furuta K, Tsuta K, Shibata T, Yamamoto S, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* 2012;72:100–11. doi:10.1158/0008-5472.CAN-11-1403.
24. Lizotte PH, Ivanova EV, Awad MM, Jones RE, Keogh L, Liu H, Dries R, Almonte C, Herter-Sprie GS, Santos A, et al. Multiparametric profiling of non-small-cell lung cancers reveals distinct immunophenotypes. *JCI Insight.* 2016;1:e89014. doi:10.1172/jci.insight.89014.
25. Tomida S, Takeuchi T, Shimada Y, Arima C, Matsuo K, Mitsudomi T, Yatabe Y, Takahashi T. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. *J Clin Oncol.* 2009;27:2793–9. doi:10.1200/JCO.2008.19.7053.
26. Director's Challenge Consortium for the Molecular Classification of Lung A, Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med.* 2008;14:822–7. doi:10.1038/nm.1790.
27. Lee ES, Son DS, Kim SH, Lee J, Jo J, Han J, Kim H, Lee HJ, Choi HY, Jung Y, et al. Prediction of recurrence-free survival in postoperative non-small cell lung cancer patients by using an integrated model of clinical information and gene expression. *Clin Cancer Res.* 2008;14:7397–404. doi:10.1158/1078-0432.CCR-07-4937.
28. Kratz JR, Jablons DM. Genomic prognostic models in early-stage lung cancer. *Clin Lung Cancer.* 2009;10:151–7. doi:10.3816/CLC.2009.n.021.
29. Lou F, Huang J, Sima CS, Dycoco J, Rusch V, Bach PB. Patterns of recurrence and second primary lung cancer in early-stage lung cancer survivors followed with routine computed tomography surveillance. *J Thorac Cardiovasc Surg.* 2013;145:75–81;discussion -2. doi:10.1016/j.jtcvs.2012.09.030.
30. Nordquist LT, Simon GR, Cantor A, Alberts WM, Bepler G. Improved survival in never-smokers vs current smokers with primary adenocarcinoma of the lung. *Chest.* 2004;126:347–51. doi:10.1378/chest.126.2.347.
31. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer.* 2012;12:252–64. doi:10.1038/nrc3239.
32. Inoshima N, Nakanishi Y, Minami T, Izumi M, Takayama K, Yoshino I, Hara N. The influence of dendritic cell infiltration and vascular endothelial growth factor expression on the prognosis of non-small cell lung cancer. *Clin Cancer Res.* 2002;8:3480–6.
33. Petersen RP, Campa MJ, Sperlazza J, Conlon D, Joshi MB, Harpole DH, Jr., Patz EF Jr. Tumor infiltrating Foxp3+ regulatory T-cells are associated with recurrence in pathologic stage I NSCLC patients. *Cancer.* 2006;107:2866–72. doi:10.1002/cncr.22822.
34. Lavin Y, Kobayashi S, Leader A, Amir ED, Elefant N, Bigenwald C, Remark R, Sweeney R, Becker CD, Levine JH, et al. Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. *Cell.* 2017;169:750–65 e17. doi:10.1016/j.cell.2017.04.014.
35. Kinoshita T, Muramatsu R, Fujita T, Nagumo H, Sakurai T, Noji S, Takahata E, Yaguchi T, Tsukamoto N, Kudo-Saito C, et al. Prognostic value of tumor-infiltrating lymphocytes differs depending on histological type and smoking habit in completely resected non-small-cell lung cancer. *Ann Oncol.* 2016;27:2117–23. doi:10.1093/annonc/mdw319.
36. Lynch TJ, Bondarenko I, Luft A, Serwatowski P, Barlesi F, Chacko R, Sebastian M, Neal J, Lu H, Cuillerot JM, Reck M. Ipilimumab in combination with paclitaxel and carboplatin as first-line treatment in stage IIIB/IV non-small-cell lung cancer: results from a randomized, double-blind, multicenter phase II study. *J Clin Oncol.* 2012;30:2046–54. doi:10.1200/JCO.2011.38.4032.
37. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Foppen MHG, Goldinger SM, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science.* 2015;350:207–11. doi:10.1126/science.aad0095.
38. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell.* 2016;165:35–44. doi:10.1016/j.cell.2016.02.065.
39. Garon EB, Rizvi NA, Hui R, Leigh N, Balmanoukian AS, Eder JP, Patnaik A, Aggarwal C, Gubens M, Horn L, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med.* 2015;372:2018–28. doi:10.1056/NEJMoa1501824.
40. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, Sosman JA, McDermott DF, Powderly JD, Gettinger SN, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature.* 2014;515:563–7. doi:10.1038/nature14011.
41. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics.* 2004;20:307–15. doi:10.1093/bioinformatics/btg405.
42. Cheng C, Yan X, Sun F, Li LM. Inferring activity changes of transcription factors by binding association with sorted expression profiles. *BMC Bioinformatics.* 2007;8:452. doi:10.1186/1471-2105-8-452.
43. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun.* 2015;6:8971. doi:10.1038/ncomms9971.
44. Best JA, Blair DA, Knell J, Yang E, Mayya V, Doedens A, Dustin ML, Goldrath AW, Immunological Genome Project Consortium. Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. *Nat Immunol.* 2013;14:404–12. doi:10.1038/ni.2536.