ORIGINAL ARTICLE

B-Cell Infiltrate in the Tumor Microenvironment Is Associated With Improved Survival in Resected Lung Adenocarcinoma

Check for updates

Ibiayi Dagogo-Jack, MD,[a,b,*] Ivan Valiev, MD,[c] Nikita Kotlov, MS,[c]
Anna Belozerova, MD,[c] Aleksandra Lopareva, MS, MD,[c] Anna Butusova, MS,[c]
Naira Samarina, PhD,[c] Alexandra Boyko, MS,[c] Zhongmin Xiang, PhD,[c]
Monique Johnson, BS,[c] Sandrine Degryse, PhD,[c] Florence K. Keane, MD, [b,d]
Lecia V. Sequist, MD,[b,e] Michael Lanuti, MD,[b,f] Nathan Fowler, MD,[c]
Mari Mino-Kenudson, MD,[b,g] Alexander Bagaev, PhD[c]

[a]Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts
[b]Harvard Medical School, Boston, Massachusetts
[c]BostonGene Corporation, Waltham, Massachusetts
[d]Department of Radiation Oncology, Massachusetts General Hospital, Boston, Massachusetts
[e]Massachusetts General Hospital, Boston, Massachusetts
[f]Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts
[g]Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts

ABSTRACT

Introduction: Relapse is common after resection of lung adenocarcinoma (LUAD). Features of the tumor microenvironment (TME) which influence postsurgical survival outcomes are poorly characterized. Here, we analyzed the TME of more than 1500 LUAD specimens to identify the relationship between B-cell infiltration and prognosis.

Methods: Whole exome sequencing and bulk RNA sequencing were performed on LUADs and adjacent normal lung tissue. Relapse-free survival and overall survival (OS) were retrospectively correlated with characteristics of the tumor and TME in three data sets.

Results: High B-cell content (defined as >10% B cells) was associated with improved OS in both a The Cancer Genome Atlas–resected LUAD data set ($p = 0.01$) and a separate institutional stage II LUAD data set ($p = 0.04$, median not reached versus 89.5 mo). A validation cohort consisting of pooled microarray data representing more than 1400 resected stage I to III LUADs confirmed the association between greater B-cell abundance, specifically higher B-cell expression, and longer postsurgical survival (median OS 90 versus 71 mo, $p < 0.01$). Relapse-free survival was longer for patients with adenocarcinomas with high B-cell content across data sets, but it did not reach statistical significance. Subcategorization of B-cell subsets indicated that high naive B-cell content was most predictive of survival. There was no correlation between

programmed death-ligand 1 expression, lymphoid aggregates, or overall immune infiltrate density and survival outcomes across the cohorts.

Conclusions: The growing adjuvant immunotherapy repertoire has increased the urgency for identifying prognostic and predictive biomarkers. Comprehensive profiling of more than 1500 LUADs suggests that high tumor-infiltrating B-cell content is a favorable prognostic marker.

Introduction

Approximately 25% to 30% of patients with NSCLC will present with resectable disease at initial diagnosis.[1] Despite undergoing surgery with curative intent, up to 50% of patients experience disease recurrence within 5 years.[1] Among patients who have a relapse, distant metastasis is the most common manifestation,[2] underscoring the importance of addressing both the primary tumor and occult micrometastatic disease. Compared with surveillance, adjuvant chemotherapy improves survival by approximately 5%.[3] Yet, the adoption of adjuvant chemotherapy is poor, with studies suggesting that it is not pursued for half of eligible patients.[4] In recent months, osimertinib and immunotherapy (atezolizumab, pembrolizumab) have gained regulatory approval in the United States as a treatment-intensification strategy after surgery and chemotherapy for select patients on the basis of reduction in risk of recurrence by 80% and 35%, respectively.[5–7] As outcomes after surgery are variable with some patients achieving cure without further therapy, biomarkers that foreshadow disease biology can inform complex decisions regarding whether to pursue adjuvant systemic therapy.

Historically, most NSCLC prognostic models have been predicated on characteristics of the resected tumor, such as size, grade, nodal involvement, lymphovascular invasion, and extension of tumor cells into visceral pleura and air spaces.[8,9] Emerging approaches to recurrence risk stratification are also focused on the tumor, including recent studies that have associated undetectable tumor DNA in the circulation with favorable prognosis after resection of NSCLC.[10–12] Tumor-focused analyses, however, may not fully account for complex interactions between tumor cells and the microenvironment which may influence propensity toward relapse. In support of a potential role of the tumor microenvironment in modulating recurrence risk, smaller retrospective studies have reported an association between increased density of tumor-infiltrating lymphocytes (TILs) and improved survival among patients with resected NSCLC.[13–15] These studies, however, largely focused on T cells and were less equipped to elucidate whether other immune cell populations, such as B cells, similarly affect relapse risk.

Here, we performed bulk RNA sequencing and whole exome sequencing (WES) to dissect the microenvironment of resected lung adenocarcinoma from patients belonging to several cohorts (Supplementary Fig. 1) to evaluate whether the density of the B-cell infiltrate is associated with prognostic outcomes.

Methods

Patient Data Collection

Patient Cohorts From The Cancer Genome Atlas and External Data Sets. See the Molecular Analysis section for clinical data collection methods for The Cancer Genome Atlas (TCGA) (cohort 1) and other external data sets (cohort 2).

Massachusetts General Hospital Stage II Cohort (Cohort 3). Primary lung tumor and adjacent normal lung tissues were obtained from 32 patients who underwent resection of stage II (American Joint Committee on Cancer version 7.0 [AJCC v7.0]) lung adenocarcinoma at Massachusetts General Hospital (MGH) between April 2009 and October 2016. Medical records of the patients were retrospectively reviewed to extract data on demographic characteristics, treatment histories, and survival outcomes. Data were updated as of November 2021. This study was approved by the institutional review board (2020A015487). All patients provided written informed consent.

Molecular Analysis

WES of MGH Stage II Cohort (Cohort 3). Tissue embedded in optimal cutting temperature compound was used for DNA and RNA co-extraction using the AllPrep DNA, RNA, Protein Mini Kit (Qiagen, Hilden, Germany). Exome libraries were captured by IDT KAPA Hyper libraries with xGen Exome Research Panel v1.0. The libraries were sequenced on a NovaSeq S4 with at least $100\times$ coverage for tumor and $30\times$ for normal.

RNA Extraction, Library Preparation, and Sequencing of MGH Stage II Cohort (Cohort 3). Tissue was prepared using the AllPrep DNA/RNA Mini kit (Qiagen). Total RNA amounts and degradation values were obtained using RNA 6000 Nano Chip run on a 2100

BioAnalyzer (Agilent, Santa Clara, CA). RNA sequencing (RNA-seq) libraries were prepared using the TruSeq RNA Library Prep Kit v2 (Illumina, Inc., San Diego, CA). Libraries were purified through Ampure XP (Beckman Coulter Genomics, Indianapolis, IN) bead purification and enriched with standard polymerase chain reaction to create a complementary DNA library. Final library quality control was carried out by evaluating the fragment size on a DNA1000 chip run on a 2100 Bio-Analyzer. Library concentrations were determined by quantitative polymerase chain reaction using the KAPA Library Quantification Kit for Next Generation Sequencing (KAPA Biosystems, Woburn, MA). Before sequencing, the libraries were normalized to 2 nmol/liter in 10 mM Tris–Cl, pH 8.5 with 0.1% Tween 20, then divided evenly to yield approximately 100 million paired-end reads per sample. The libraries were denatured with 0.05 N sodium hydroxide and diluted to 20 pmol/liter. Cluster generation of the denatured libraries was performed according to the manufacturer's specifications (Illumina, Inc.) using the HiSeq PE Cluster Kit v2 and or v4 chemistry and flow cells. Libraries were clustered appropriately with a 1% PhiX spike-in. Sequencing-by-synthesis was performed on a HiSeq2500 using appropriate chemistry with paired-end 101 bp reads. Sequence-read data were processed and converted to FASTQ format for downstream analysis by Illumina BaseSpace analysis software, FASTQ Generation v1.0.0.

Transcriptomic Analysis of MGH Stage II Cohort (Cohort 3). Next-generation sequencing quality control analysis was performed using FastQC (v0.11.5 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), FastQ Screen3 (v0.11.1), RSeQC4 (v3.0.0), and MultiQC5 (v1.6). Reads were aligned using Kallisto to estimate expression levels (v0.46.0) to GENCODE v23.[16] Noncoding transcripts, histone coding transcripts, and mitochondrial transcripts were removed, resulting in 20,062 transcripts for downstream analysis. Gene expression was quantified as transcripts per million with log2 transformation. Bulk RNA-seq data were deconvolved using the decision tree machine learning algorithm Kassandra.[17]

RNA Expression Analysis Using External Data Sets. *TCGA Lung Adenocarcinoma (Cohort 1).* Transcriptomic data were downloaded from the USCS XENA portal (https://xena.ucsc.edu/) as transcripts per million units. Sample identification numbers were unified to patient identification numbers using the first 12 characters. Patients with more than one tumor RNA-seq sample or missing clinical annotation were removed. Clinical data were downloaded from the GDC TCGA data portal (MC3 data set).[18]

Other Lung Adenocarcinoma Data Sets (Cohort 2). For array expression analysis, GSE72094, GSE37745, GSE31210, GSE119267, GSE41271, GSE68465, GSE19188, GSE40791, GSE50081, GSE94601, and GSE43580 were downloaded in raw cell format, median normalized, and harmonized to be used as a pooled cohort (cohort 2). Only data derived from lung adenocarcinoma samples were used.

Identification of B-Cell Subsets (Cohorts 1 and 3). To quantify the abundance of select B-cell subsets, bulk RNA-seq data were deconvolved using a machine learning algorithm (Kassandra) applied to a resolution that permitted differentiation of the following B-cell subsets on the basis of gene expression: naive, class-switched, nonswitched memory, and plasmablasts.[17,19]

Histopathologic Analysis

Stage II Lung Adenocarcinomas in MGH Cohort (Cohort 3). Formalin-fixed, paraffin-embedded hematoxylin and eosin slides were reviewed to assess lymphocyte infiltration, determine histologic grade, and identify lymphoid aggregates.

Statistical Analysis

Python (version 3.8) was used for statistical analysis and graphing. Scikit-learn version 1.0.2, lifelines version 0.26.4, matplotlib version 3.5.1, and seaborn version 0.11.2 were used for data analysis and visualization.

Results

Prognostic Relevance of B-Cell Abundance in Resected Adenocarcinoma (Cohort 1)

To identify microenvironment characteristics, including B-cell infiltration, that were associated with recurrence risk, we first analyzed a TCGA resected lung adenocarcinoma RNA-seq data set (cohort 1) comprising stages I to IV NSCLC. Given the aim of identifying relevant biomarkers in all-comers with resected NSCLC, all stages were pooled in the analysis. To characterize the microenvironment composition, we used the previously described RNA-seq deconvolution algorithm, Kassandra, as previously described.[19]

We divided the TCGA lung adenocarcinoma data set into tertiles on the basis of the degree of B cell infiltration. There was no statistically significant difference in survival across tertiles ($p = 0.08$; Supplementary Fig. 2). On inspection, however, there was considerable overlap between B cells grouped in the 0 to 33rd percentile and those grouped in the 33rd to 66th percentile. On the basis of this finding, the data set was subsequently divided into two groups stratified by the 66th percentile resulting in separation in the survival curves
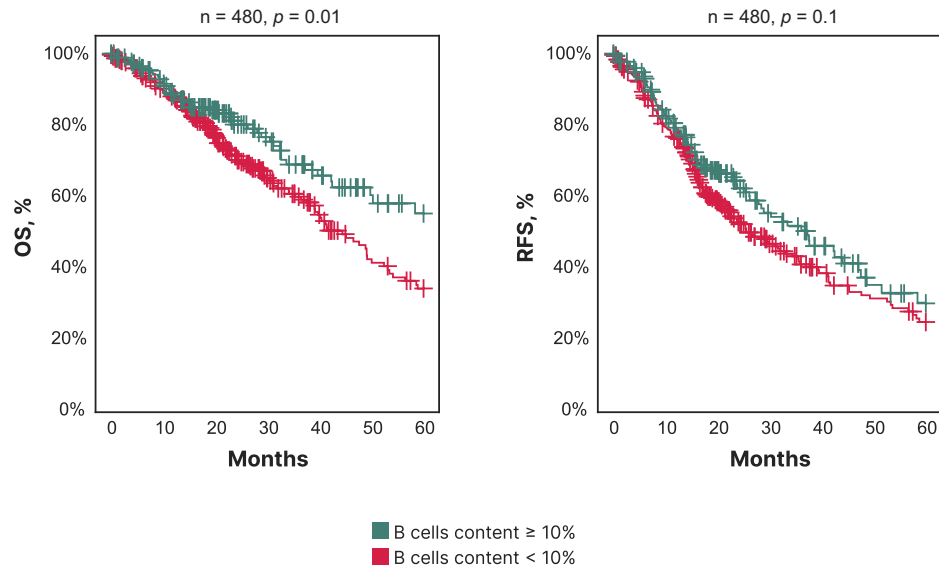
Figure 1. Prognostic relevance of B-cell abundance in resected lung adenocarcinoma in TCGA data set (cohort 1). Curves depict OS and RFS for patients with resected lung adenocarcinoma in TCGA data set according to high versus low B-cell content. The pooled cohort of 480 specimens includes stages I to IV resected lung cancer. OS, overall survival; RFS, relapse-free survival; TCGA, The Cancer Genome Atlas.

(Supplementary Fig. 2). As the 66th percentile corresponded to 10% absolute B cells, samples with more than or equal to 10% were thereafter categorized as having "high" B cell content whereas those with B cell content below this threshold were considered to have "low" B cell content.

Using this cutoff, we analyzed whether stratification by B cell abundance could identify patients at higher versus lower risk of recurrence and death after resection of lung adenocarcinoma. Overall survival was significantly longer in the high B cell content group ($p = 0.01$; Fig. 1). In contrast, B cell content was not significantly associated with relapse-free survival ($p = 0.10$, median 37.3 mo versus 26.5 mo), though relapse-free survival was numerically higher for patients with tumors with high B cell content (Fig. 1). We also assessed whether other tumor and microenvironment features, specifically histologic grade, tumor mutational burden (TMB), and programmed death-ligand 1 (*PD-L1*) expression by RNA-seq, correlated with survival in the TCGA data set. This analysis did not reveal a significant association between overall survival and either TMB or *PD-L1* expression. Nevertheless, overall survival was shorter for patients with grade 3 adenocarcinoma compared with those with lower grade adenocarcinoma (Fig. 2A–C).
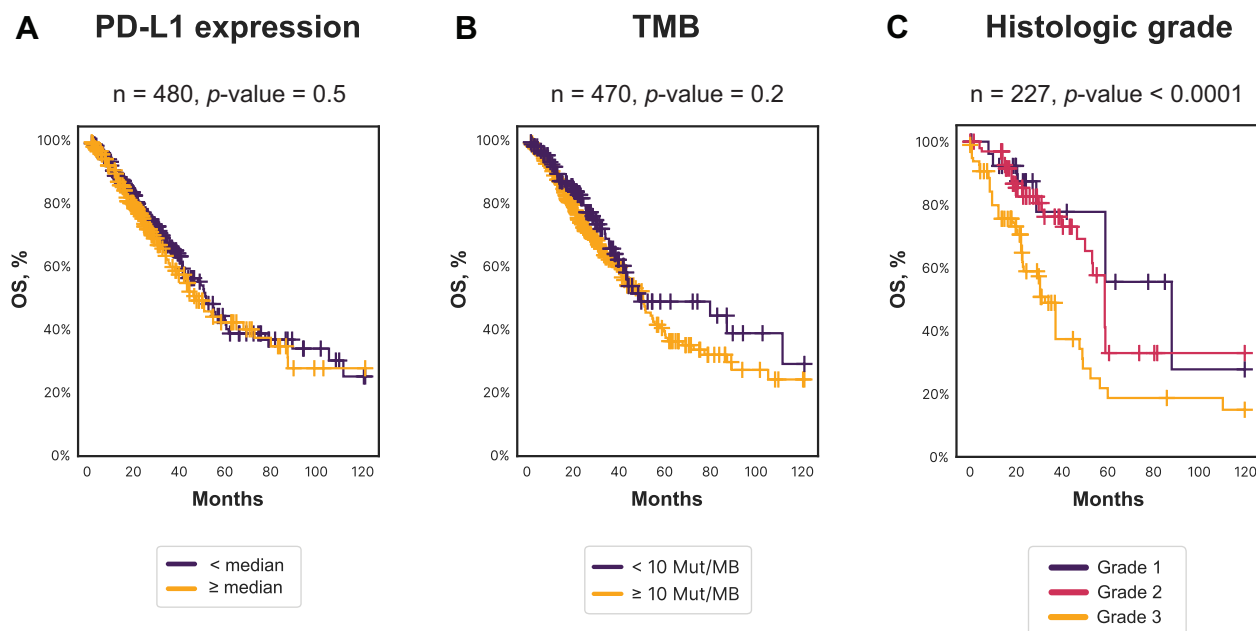
Validation in Microarray Data Sets (Cohort 2)

To validate our findings, we explored the prognostic relevance of B cell abundance in microarray data sets encompassing resected stages I to III lung adenocarcinoma. For this analysis, we generated a meta-cohort

(cohort 2) comprising 1422 resected lung adenocarcinomas captured in publicly available microarray data sets. Most (54%) of the tumors in the meta-cohort were categorized as stage I per AJCC v7.0. As RNA-seq data were not available for these tumors, we used a previously described custom gene expression signature to estimate B-cell abundance.[17] To confirm concordance between B-cell abundance as gauged by the custom gene expression signature versus bulk RNA sequencing using the deconvolution algorithm used in the analyses previously discussed, we applied both techniques to the TCGA data set and found overall agreement (Fig. 3A). Of note, owing to the nature of the data available for the microarray data set, B-cell expression that exceeded the median was considered representative of high B-cell abundance instead of the previously used cutoff of 10%.

When applied to the 1422 lung adenocarcinomas (cohort 2), high B-cell abundance (defined as B-cell expression above the median) was significantly associated with longer median overall survival ($p < 0.01$, 90 mo versus 71 mo; Fig. 3B), consistent with findings from bulk RNA sequencing in the TCGA cohort. The prognostic value of B-cell abundance was apparent across all stages (Fig. 3B), though statistical significance was only reached in the larger stage I subgroup.

B-Cell Content Correlates With Survival in MGH Stage II Cohort (Cohort 3)

We next sought to confirm our findings in an institutional cohort (cohort 3) representing resected stage II

A   PD-L1 expression

n = 480, *p*-value = 0.5

B   TMB

n = 470, *p*-value = 0.2

C   Histologic grade

n = 227, *p*-value < 0.0001

Figure 2. Survival outcomes in TCGA lung adenocarcinoma data set (cohort 1) according to *PD-L1* expression, TMB, and histologic grade. Curves depict OS for patients with resected stages I to IV lung adenocarcinoma according to (*A*) *PD-L1* expression, (*B*) TMB, and (*C*) histologic grade. Mut/MB, mutations per megabase; OS, overall survival; PD-L1, programmed death-ligand 1; TCGA, The Cancer Genome Atlas; TMB, tumor mutational burden.

NSCLC (AJCC v7.0). Stage II tumors were selected for this validation cohort as compared with stage I tumors which predominated in both the TCGA and microarray data sets and are associated with lower recurrence rates; stage II NSCLC is associated with nearly equivalent likelihood of cure versus relapse after surgical resection.[1] Furthermore, given limited available tissue specimens and the more aggressive biology ascribed to stage III lung adenocarcinoma, we intentionally focused on stage II tumors.

For this analysis, we performed WES and bulk RNA-seq on 32 resected stage II lung adenocarcinomas. The quality of the 32 specimens was sufficient for WES; however, only 29 specimens were of sufficient quality to support RNA-seq (Supplementary Fig. 3*A* and *B*). Most (62%) tumors were categorized as stage IIB (AJCC v7.0). The median age of the group was 69 (range: 44–87) years old, and most patients were female (62%; Supplementary Table 1). Seven patients (24%) were never smokers. Consistent with national trends, 16 patients (55%) had received adjuvant chemotherapy.[4] As presented in Supplementary Figure 3*C*, the most common driver alterations identified in the cohort of 29 lung adenocarcinomas included *KRAS* mutations (n = 12) and *EGFR* mutations (n = 8).

With median follow-up of 4.9 (range: 1.5–12.4) years after surgery, 13 patients (45%) developed recurrence of NSCLC and 10 patients (34%) died, including two patients who died without evidence of relapse. Eight of 13 patients (62%) with lung cancer recurrence previously received adjuvant chemotherapy. As with the TCGA data set (cohort 1), the validation cohort neatly clustered into two groups on the basis of the degree of B-cell infiltration per the 10% threshold (Fig. 4*A*). Consistent with findings from the larger TCGA data set (cohort 1), high B-cell content was significantly associated with improved overall survival (log-rank test, *p* = 0.04, median not reached versus 89.5 mo) and had a trend toward improved relapse-free survival (log-rank test *p* = 0.2, median not reached versus 48.8 mo; Fig. 4*B*).

We reviewed hematoxylin and eosin slides from specimens in cohort 3 to assess the relationship between the histomorphology of the tumor and other immune microenvironment features and survival outcomes (Supplementary Fig. 4*A*–*C*). No statistically significant association was identified between histologic grade, presence of lymphoid aggregates, or density of the immune infiltrate (including both B cells and T cells) and either overall or relapse-free survival. Notably, high B-cell content was identified in adenocarcinomas with and without lymphoid aggregates. We also evaluated whether tumor *PD-L1* expression (as assessed by RNA-seq) and TMB derived from WES distinguished patients with poorer versus more favorable postsurgical outcomes. *PD-L1* expression did not predict relapse risk or affect survival (Supplementary Fig. 4*D*). In contrast, lower TMB was associated with longer overall survival (Supplementary Fig. 4*E*).
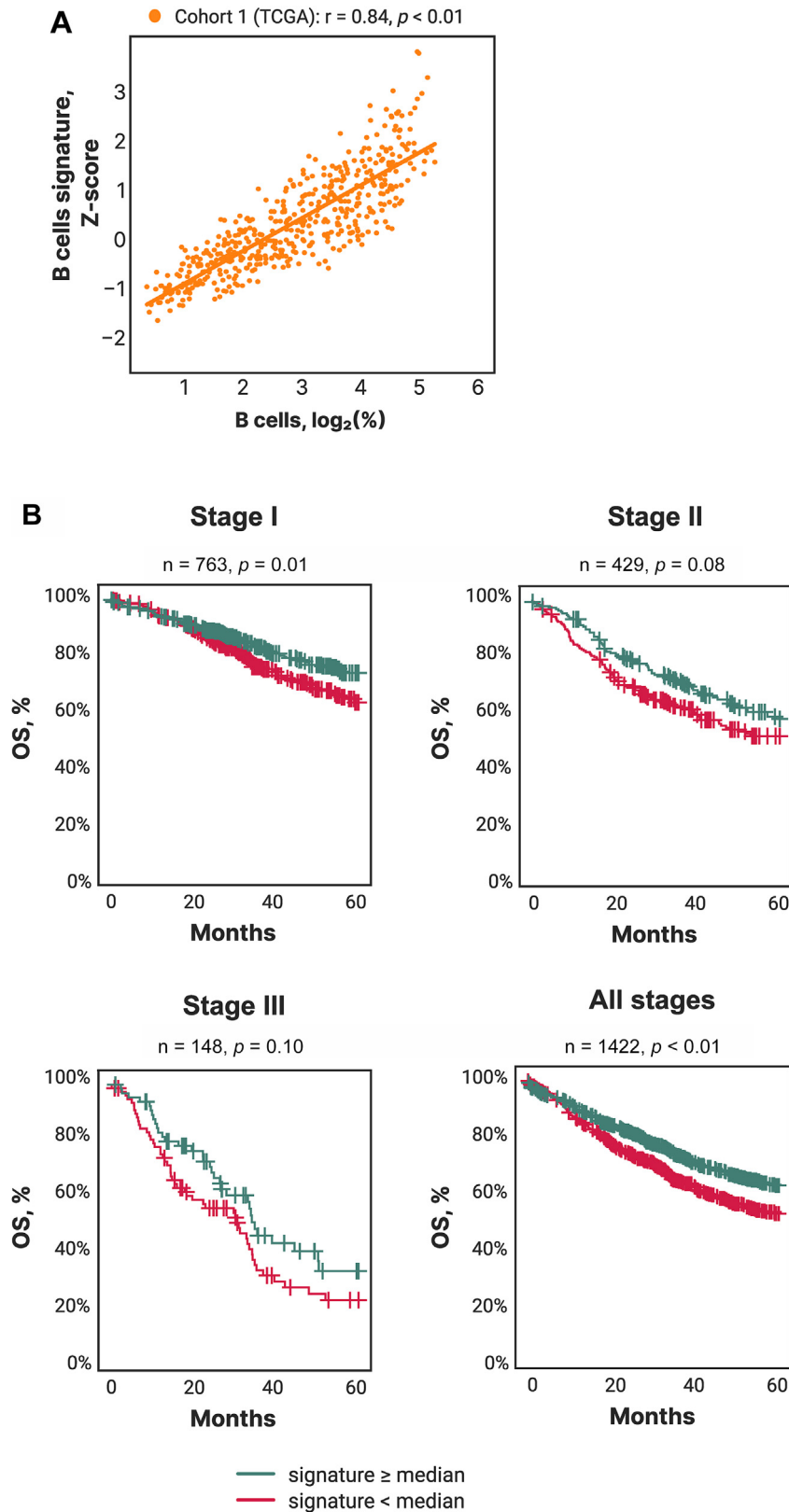
Figure 3. Relationship between B-cell content and survival outcomes in external microarray data sets (cohort 2). (*A*) Agreement between B-cell content determined by the Kassandra RNA-seq deconvolution algorithm versus the B-cell expression signature. (*B*) Curves illustrate OS for patients with stage I, II, and III resected lung adenocarcinomas and a pooled cohort of 1422 specimens (cohort 2) representing stage I to III lung adenocarcinomas according to B-cell expression as determined by a custom gene expression signature. OS, overall survival; RNA-seq, RNA-sequencing; TCGA, The Cancer Genome Atlas.
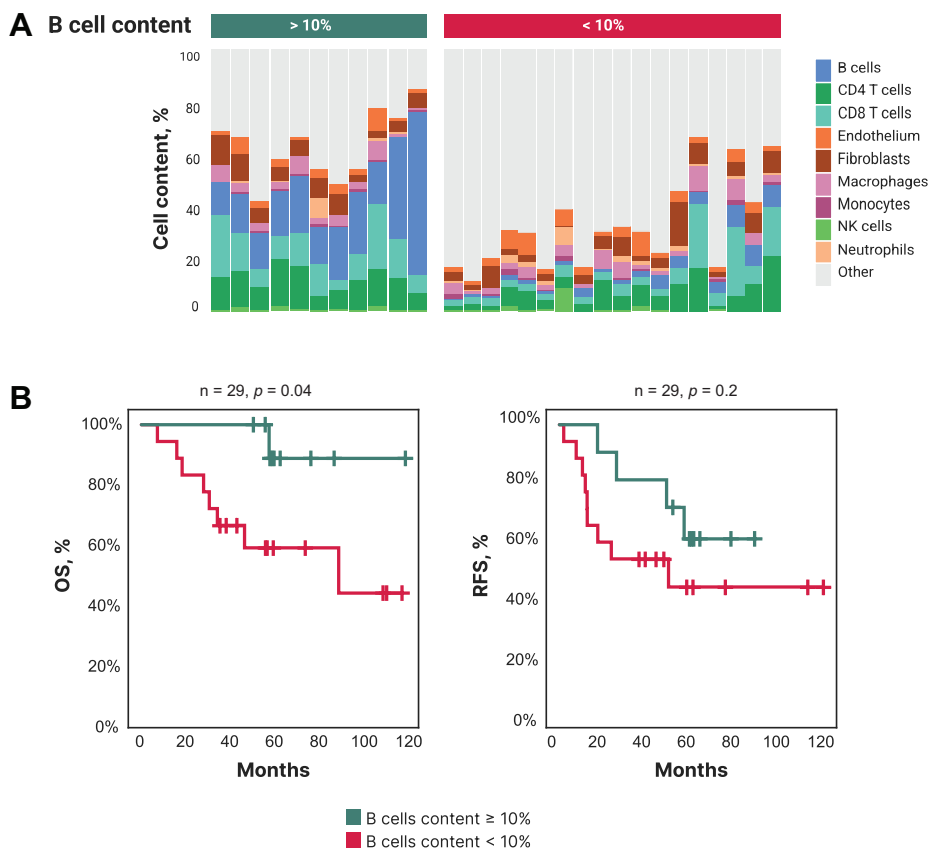
Figure 4. Correlation between B-cell content and survival in MGH stage II lung adenocarcinoma data set (cohort 3). (*A*) Tumor microenvironment deconvolution by Kassandra (Zaitsev et al. Cancer Cell 2022) of RNA-seq data derived from 29 resected lung adenocarcinoma specimens. Specimens are groups on the basis of B-cell content. (*B*) Curves depict OS and RFS for patients with stage II lung adenocarcinoma with tumors with high versus low B-cell content using a 10% cutoff. *p* values were calculated using log-rank tests. NK, natural killer; OS, overall survival; RFS, relapse-free survival; RNA-seq, RNA-sequencing.

Impact of Distinct B-Cell Subsets on Survival Outcomes. Overall, our analyses of three distinct cohorts revealed that B-cell content is a prognostic factor in resected lung adenocarcinoma. To determine whether survival outcomes were driven by particular subsets of B cells, we mined RNA-seq data from patients in cohorts 1 and 3 and applied the Kassandra algorithm referenced previously to subcategorize the B cells into the following subsets: naive, class-switch memory, nonswitched memory, and secreting (plasmablasts). Cohort 2 was not considered for this analysis as RNA-seq data were not available. We then reanalyzed outcomes for the MGH stage II cohort (cohort 3) and TCGA cohort (cohort 1) using the previously defined 10% threshold as the cutoff for "high" naive B cells. High naive B-cell content was associated with numerically improved overall (median not reached versus 90 mo, $p = 0.08$) and relapse-free survival (median not reached versus 48 mo, $p = 0.01$) in the MGH cohort (cohort 3), but only the latter reached statistical significance ([Fig. 5](#)*A*). Similarly, we observed longer relapse-free survival ($p = 0.09$, median 36.1 mo versus 29.7 mo) and overall survival ($p = 0.04$, median

60.1 mo versus 47.8 mo) in tumors with high naive B-cell content in the TCGA data set (cohort 1), though in contrast to the MGH stage II data set (cohort 3), only overall survival met statistical significance. ([Fig. 5](#)*B*). When the analyses were repeated on the TCGA data set (cohort 1) using other B-cell subsets, relationships between abundance of these distinct B-cell subsets and survival were not observed ([Supplementary Fig. 5](#)*A–C*).

B-Cell Infiltration Is Enriched in Tumor Compared With Adjacent Lung Tissue. Finally, to characterize the B-cell infiltrate in the tumor versus juxtaposed lung tissue, we reviewed RNA-seq data from the TCGA-resected lung adenocarcinoma cohort (cohort 1). The B-cell infiltrate was sparse in adjacent lung tissue in the TCGA data set (cohort 1) ([Fig. 6](#)*A* and *C*). To confirm these findings, we performed RNA-seq on adjacent normal lung tissue and compared findings to nearby tumor tissue in the MGH stage II cohort (cohort 3). In the MGH stage II adenocarcinoma cohort (cohort 3), B cells were also scarce in adjacent normal tissue ([Fig. 6](#)*B* and *D*), suggesting that the B-cell infiltrate was uniquely
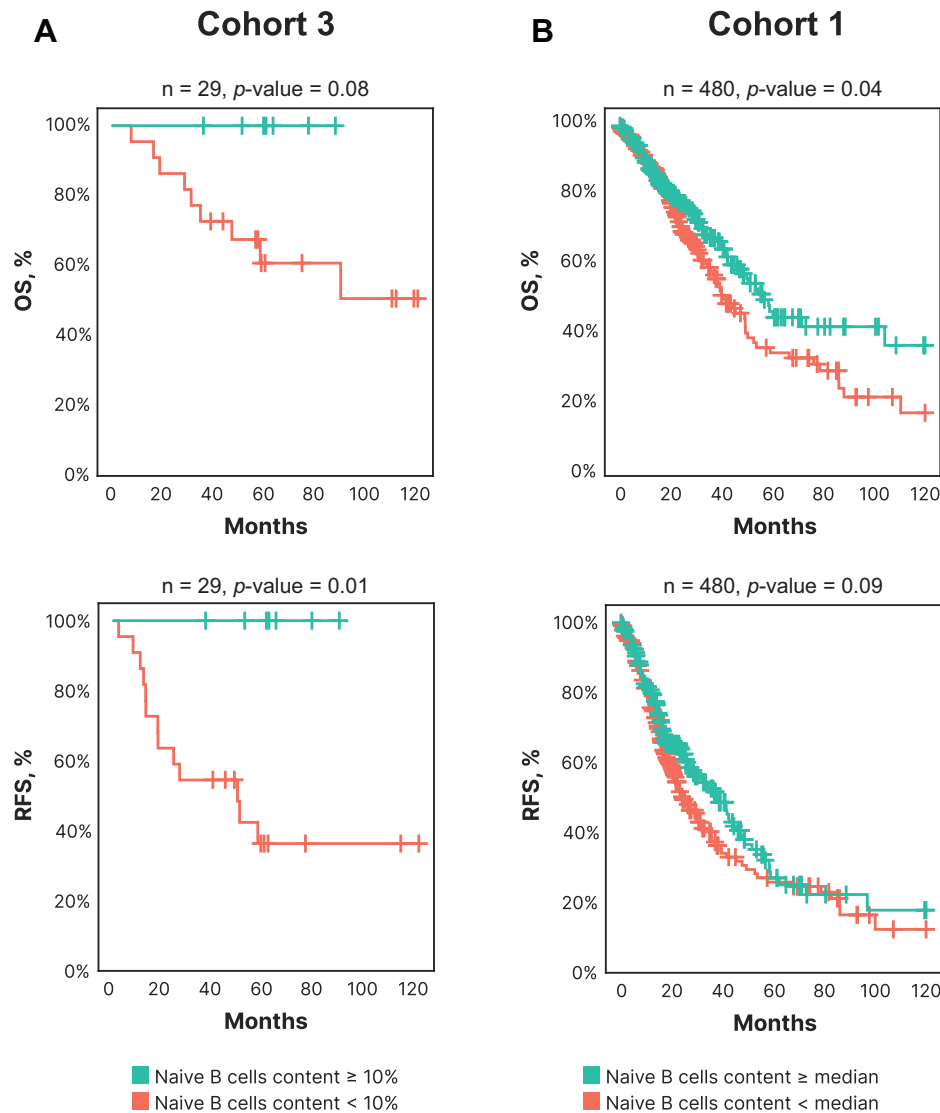
Figure 5. Association between abundance of naive B cells and survival in cohorts 1 and 3. Curves depict OS and RFS for patients with resected lung adenocarcinoma according to high versus low naive B-cell content as defined by a threshold of 10% in (*A*) stage II MGH data set (cohort 3) and (*B*) TCGA data set (cohort 1) encompassing stage I to IV lung adenocarcinomas. MGH, Massachusetts General Hospital; OS, overall survival; RFS, relapse-free survival; TCGA, The Cancer Genome Atlas.

enriched in the tumor. Collectively, these findings raise the possibility that B cells may be deliberately recruited into lung adenocarcinomas.

Discussion

Among patients who are medically fit for surgery, the extent of tumor involvement is a critical determinant of resectability and factors heavily into survival estimates.[1] Although stage has tremendous prognostic value, primarily relying on disease stage for survival projections and adjuvant systemic therapy decisions does not fully account for the biological and molecular diversity encompassed within stage groups. Indeed, in metastatic NSCLC, the molecular and immunologic profile of tumors modulates survival outcomes.[20] Thus, a better understanding of the immunologic and molecular contexture of resected tumors may provide invaluable insights for unraveling drivers of heterogeneous outcomes among patients with resected tumors of similar stage. In this study, we performed a comprehensive analysis of more than 1500 resected lung adenocarcinomas belonging to three distinct cohorts to find molecular factors potentially underlying disparate survival outcomes. Through our analysis, we identified an association between an abundance of tumor-infiltrating B cells and improved overall survival after surgery.

In this study, we analyzed three data sets of lung adenocarcinoma resected in an era predating recent adjuvant approvals of immunotherapy and osimertinib. As such, our study was primarily focused on identifying
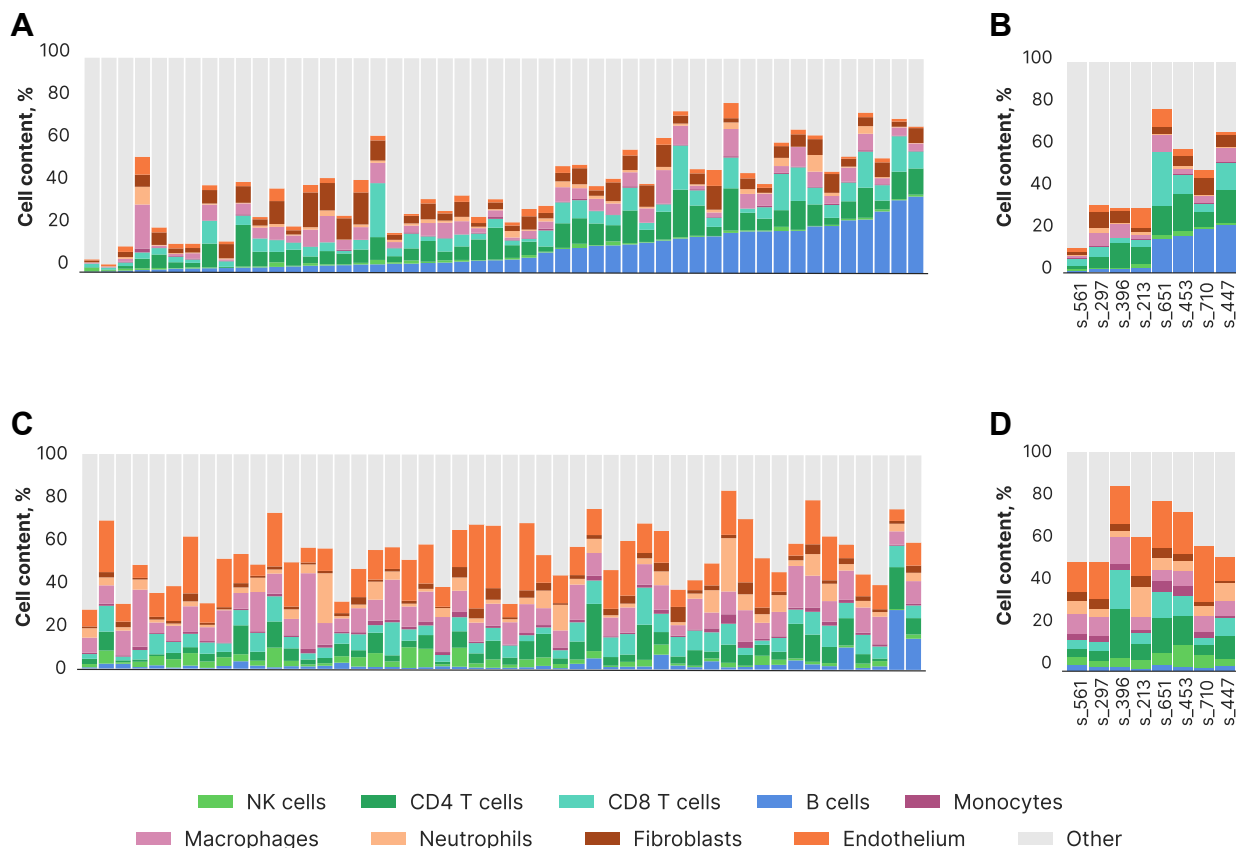
Figure 6. B-cell composition in tumor tissue and adjacent normal lung tissue in cohort 1 (TCGA) and cohort 3 (MGH stage II) supports enrichment of B cells in tumor but not adjacent normal tissue. The figures illustrate the content of the microenvironment (color-coded per legend) in the samples. Each bar represents a separate specimen. (*A, B*) Tissue deconvolution of tumor samples of TCGA (*A*) and MGH (*B*) cohorts. (*C, D*) Tissue deconvolution of adjacent normal samples matched to the tumor specimen found in panels *A* and *B*. MGH, Massachusetts General Hospital; NK, natural killer; TCGA, The Cancer Genome Atlas.

prognostic rather than predictive biomarkers. In our analysis, biomarkers currently used to stratify treatments in the metastatic setting, including TMB and *PD-L1* expression, typically associated with response to immunotherapy did not consistently predict risk of relapse or death after resection. Instead, B-cell infiltration, a less studied characteristic, emerged as a relevant prognostic marker that repeatedly identified a subgroup of patients with improved survival across unique data sets. The association between increased density of infiltrating lymphocytes and improved survival after lung cancer resection has been established.[13–15,21,22] Nevertheless, studies that support this association largely characterized the T cell infiltrate, with the relatively fewer studies that analyzed intratumoral B-cell composition reporting inconsistent results in small patient cohorts.[14,23] Besides T cell density, studies suggest that spatial distribution of T lymphocytes, including organization into TLS, also influences overall prognosis.[14,22,24,25] In our analysis, we did not assess TLS directly; however, the prognostic implications of a robust B-cell infiltrate were independent of the formation of lymphoid aggregates. As the spatial architecture of the B-cell infiltrate was only a minor component of our current analysis, future studies are warranted to fully resolve spatial relationships between infiltrating B cells and other stromal cells and assess how these interactions affect disease outcomes. Notably, recent studies using single-cell RNA-seq have made inroads toward addressing these unresolved questions.[23]

As mentioned previously, the three data sets used to support our analysis were generated at a time where the only available adjuvant systemic therapy was chemotherapy. Our analysis did not explore the possibility that the improved survival observed among patients with tumors with high B-cell content may have resulted from improved efficacy of adjuvant chemotherapy. Nevertheless, the prognostic value of high B-cell content among stage I tumors in the microarray data set (cohort 2) where chemotherapy is sparingly used suggests that our observations cannot be explained by chemotherapy effect alone. Moreover, correlative analyses from four practice-defining randomized adjuvant chemotherapy trials failed to identify a relationship between lymphocyte

infiltration and benefit from adjuvant chemotherapy,[15] lending further support to the notion that the differences in survival noted in our study were not due to imbalances in exposure to or differential benefit from chemotherapy. With recent adoption of immunotherapy and osimertinib for adjuvant treatment of resected NSCLC and the potential for these therapies to ultimately improve overall survival,[5,6] these analyses should be replicated using modern era cohorts to confirm the ongoing relevance of B-cell infiltration as a prognostic biomarker and to assess whether B-cell infiltration affects response to these therapies.

Our study has several limitations. First, the bulk of the analyses was conducted using pooled data sets containing a variety of stages. As stage is a proven prognostic factor,[1] we independently analyzed outcomes in distinct stage groups in the microarray data set (cohort 2), given the small sample size of stage II to III adenocarcinomas in TCGA (cohort 1), and included a separate institutional stage II cohort (cohort 3). These analyses suggested that our findings are relevant across disease stages, though the magnitude of prognostic effect of B-cell infiltration may vary by stage. In addition, our study exclusively considered patients with adenocarcinoma and; therefore, findings may not be applicable to other histologic subsets, including squamous NSCLC. Indeed, studies suggest that both the makeup and prognostic value of infiltrating lymphocytes in resected squamous and adenocarcinomas may differ.[26] Third, cohort 3 was inadvertently enriched for tumors with *KRAS* and *EGFR* mutations, potentially affecting generalizability of findings from this particular cohort. Fourth, owing to the nature of available data, validating findings in the microarray data set (cohort 2) required use of a different threshold to delineate high B-cell-infiltrate. Fifth, we did not functionally characterize the B-cell infiltrate as part of this analysis and our subcategorization of B-cell subsets was limited. Although these analyses are outside of the scope of our study owing to the limitations of bulk RNA sequencing, recent studies using single-cell RNA-seq have provided valuable insights into the immunogenomic and spatial landscape of B-cell in resected lung adenocarcinoma.[23] Finally, we observed an impact of B-cell abundance on overall survival, but not relapse-free survival. As co-morbidities affect life expectancy and prognosis, it is possible that absent or low B-cell infiltration in the tumor may be associated with or may be a surrogate for another unidentified adverse prognostic marker that is unrelated to lung cancer-specific mortality.

In summary, robust analyses of multiple data sets representing more than 1500 unique lung adenocarcinomas nominate B-cell content as a potentially relevant prognostic biomarker for resected lung adenocarcinoma.

These findings provide rationale for expanding the scope of future studies of prognostic immunologic biomarkers in NSCLC beyond the traditional focus on T cells.

CRediT Authorship Contribution Statement

Ibiayi Dagogo-Jack: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review and editing.

Ivan Valiev: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review and editing.

Nikita Kotlov: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review and editing.

Anna Belozerova: Data curation, Formal analysis, Investigation, Methodology, Writing—review and editing.

Aleksandra Lopareva: Data curation, Formal analysis, Investigation, Methodology, Writing—review and editing.

Anna Butusova: Data curation, Formal analysis, Investigation, Methodology, Writing—review and editing.

Naira Samarina: Data curation, Formal analysis, Investigation, Methodology, Writing—review and editing.

Alexandra Boyko: Data curation, Formal analysis, Investigation, Methodology, Writing—review and editing.

Zhongmin Xiang: Formal analysis, Investigation, Writing—review and editing.

Monique Johnson: Formal analysis, Investigation, Writing—review and editing.

Sandrine Degryse: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review and editing.

Jaimie Barth: Data curation, Writing—review and editing.

Florence K. Keane: Data curation, Writing—review and editing.

Lecia Sequist: Data curation, Methodology, Writing—review and editing.

Michael Lanuti: Conceptualization, Data curation, Investigation, Methodology, Writing—review and editing.

Nathan Fowler: Methodology, Writing—review and editing.

Mari Mino-Kenudson: Data curation, Formal analysis, Investigation, Methodology, Writing—review and editing.

Alexander Bagaev: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review and editing.

Acknowledgments

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *JTO Clinical and Research Reports* at www.jtocrr.org and at https://doi.org/10.1016/j.jtocrr.2023.100527.

References

1. Chansky K, Detterbeck FC, Nicholson AG, et al. The IASLC lung cancer staging project: external validation of the revision of the TNM stage groupings in the eighth edition of the TNM classification of lung cancer. *J Thorac Oncol*. 2017;12:1109-1121.
2. Mouillet G, Monnet E, Milleron B, et al. Pathologic complete response to preoperative chemotherapy predicts cure in early-stage non-small-cell lung cancer: combined analysis of two IFCT randomized trials. *J Thorac Oncol*. 2012;7:841-849.
3. Pignon JP, Tribodet H, Scagliotti GV, et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. *J Clin Oncol*. 2008;26:3552-3559.
4. Kehl KL, Zahrieh D, Yang P, et al. Rates of guideline-concordant surgery and adjuvant chemotherapy among patients with early-stage lung cancer in the US ALCHEMIST study (alliance A151216). *JAMA Oncol*. 05 01 2022;8:717-728.
5. Felip E, Altorki N, Zhou C, et al. Adjuvant atezolizumab after adjuvant chemotherapy in resected stage IB-IIIA non-small-cell lung cancer (IMpower010): a randomised, multicentre, open-label, phase 3 trial. *Lancet*. 2021;398:1344-1357.
6. Wu YL, Tsuboi M, He J, et al. Osimertinib in resected. *N Engl J Med*. 10 29 2020;383:1711-1723.
7. O'Brien M, Paz-Ares L, Marreaud S, et al. Pembrolizumab versus placebo as adjuvant therapy for completely resected stage IB-IIIA non-small-cell lung cancer (PEARLS/KEYNOTE-091): an interim analysis of a randomised, triple-blind, phase 3 trial. *Lancet Oncol*. 2022;23:1274-1286.
8. Kris MG, Gaspar LE, Chaft JE, et al. Adjuvant systemic therapy and adjuvant radiation therapy for stage I to IIIA completely resected non-small-cell lung cancers: American Society of Clinical Oncology/Cancer Care Ontario Clinical Practice Guideline Update. *J Clin Oncol*. 2017;35:2960-2974.
9. Mino-Kenudson M. Significance of tumor spread through air spaces (STAS) in lung cancer from the pathologist perspective. *Transl Lung Cancer Res*. 2020;9:847-859.
10. Zhang EW, Dagogo-Jack I, Kuo A, Rooney MM, Shaw AT, Digumarthy SR. Association between circulating tumor DNA burden and disease burden in patients with ALK-positive lung cancer. *Cancer*. 2020;126:4473-4484.
11. Provencio M, Serna-Blasco R, Nadal E, et al. Overall survival and biomarker analysis of neoadjuvant nivolumab plus chemotherapy in operable stage IIIA non-small-cell lung cancer (NADIM phase II trial). *J Clin Oncol*. 2022;40:2924-2933.
12. Forde PM, Spicer J, Lu S, et al. Neoadjuvant nivolumab plus chemotherapy in resectable lung cancer. *N Engl J Med*. 2022;386:1973-1985.
13. Donnem T, Hald SM, Paulsen EE, et al. Stromal CD8+ T-cell density—a promising supplement to TNM staging in non-small cell lung cancer. *Clin Cancer Res*. 2015;21:2635-2643.
14. Schalper KA, Brown J, Carvajal-Hausdorf D, et al. Objective measurement and clinical significance of TILs in non-small cell lung cancer. *J Natl Cancer Inst*. 2015;107:dju435.
15. Brambilla E, Le Teuff G, Marguet S, et al. Prognostic effect of tumor lymphocytic infiltration in resectable non-small-cell lung cancer. *J Clin Oncol*. 2016;34:1223-1230.
16. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34:525-527.
17. Bagaev A, Kotlov N, Nomie K, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell*. 2021;39:845-865.e7.
18. Ellrott K, Bailey MH, Saksena G, et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst*. 2018;6:271-281.e7.
19. Zaitsev A, Chelushkin M, Dyikanov D, et al. Precise reconstruction of the TME using bulk RNA-seq and a machine learning algorithm trained on artificial transcriptomes. *Cancer Cell*. 2022;40:879-894.e16.
20. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. *Lancet*. 2021;398:535-554.
21. Barua S, Fang P, Sharma A, et al. Spatial interaction of tumor cells and regulatory T cells correlates with survival in non-small cell lung cancer. *Lung Cancer*. 2018;117:73-79.
22. AbdulJabbar K, Raza SEA, Rosenthal R, et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat Med*. 2020;26:1054-1062.
23. Hao D, Han G, Sinjab A, et al. The single-cell immuno-genomic landscape of B and plasma cells in early-stage lung adenocarcinoma. *Cancer Discov*. 2022;12:2626-2645.
24. Rakaee M, Kilvaer TK, Jamaly S, et al. Tertiary lymphoid structure score: a promising approach to refine the TNM staging in resected non-small cell lung cancer. *Br J Cancer*. 2021;124:1680-1689.
25. Dieu-Nosjean MC, Antoine M, Danel C, et al. Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures. *J Clin Oncol*. 2008;26:4410-4417.
26. Ding R, Prasanna P, Corredor G, et al. Image analysis reveals molecularly distinct patterns of TILs in NSCLC associated with treatment outcome. *NPJ Precis Oncol*. 2022;6:33.