RESEARCH   **Open Access**

A lepidic gene signature predicts patient prognosis and sensitivity to immunotherapy in lung adenocarcinoma

Thinh T. Nguyen[1], Hyun-Sung Lee[2], Bryan M. Burt[2], Jia Wu[3,4], Jianjun Zhang[3,5], Christopher I. Amos[1,6,7*] and Chao Cheng[1,6,7*] [ID]

Abstract

Background: Lung adenocarcinoma, the most common type of lung cancer, has a high level of morphologic heterogeneity and is composed of tumor cells of multiple histological subtypes. It has been reported that immune cell infiltration significantly impacts clinical outcomes of patients with lung adenocarcinoma. However, it is unclear whether histologic subtyping can reflect the tumor immune microenvironment, and whether histologic subtyping can be applied for therapeutic stratification of the current standard of care.

Methods: We inferred immune cell infiltration levels using a histological subtype-specific gene expression dataset. From differential gene expression analysis between different histological subtypes, we developed two gene signatures to computationally determine the relative abundance of lepidic and solid components (denoted as the L-score and S-score, respectively) in lung adenocarcinoma samples. These signatures enabled us to investigate the relationship between histological composition and clinical outcomes in lung adenocarcinoma using previously published datasets.

Results: We found dramatic immunological differences among histological subtypes. Differential gene expression analysis showed that the lepidic and solid subtypes could be differentiated based on their gene expression patterns while the other subtypes shared similar gene expression patterns. Our results indicated that higher L-scores were associated with prolonged survival, and higher S-scores were associated with shortened survival. L-scores and S-scores were also correlated with global genomic features such as tumor mutation burdens and driver genomic events. Interestingly, we observed significantly decreased L-scores and increased S-scores in lung adenocarcinoma samples with *EGFR* gene amplification but not in samples with *EGFR* gene mutations. In lung cancer cell lines, we observed significant correlations between L-scores and cell sensitivity to a number of targeted drugs including EGFR inhibitors. Moreover, lung cancer patients with higher L-scores were more likely to benefit from immune checkpoint blockade therapy.

* Correspondence: chris.amos@bcm.edu; chao.cheng@bcm.edu
[1]Institute for Clinical and Translational Research, Baylor College of Medicine,
Houston, TX 77030, USA
Full list of author information is available at the end of the article

Conclusions: Our findings provided further insights into evaluating histology composition in lung adenocarcinoma. The established signatures reflected that lepidic and solid subtypes in lung adenocarcinoma would be associated with prognosis, genomic features, and responses to targeted therapy and immunotherapy. The signatures therefore suggested potential clinical translation in predicting patient survival and treatment responses. In addition, our framework can be applied to other types of cancer with heterogeneous histological subtypes.

Keywords: Lung adenocarcinoma, Histology, Tumor immune microenvironment, Gene signature, Prognosis, Genomic alterations, Drug sensitivity, Immunotherapy

Background

Lung cancer is the leading cause of cancer-related deaths in the USA, causing more than 20% of all cancer deaths in 2020 [1]. Depending on the stage at the time of diagnosis, the prognosis of lung cancer differs substantially among patients. The 5-year survival rate is approximately 55% for localized lung cancer, in contrast to only 5% for distant spread [1]. Of all lung cancer cases, about 85% are non-small cell lung cancer (NSCLC) and 15% are small cell lung cancer [2]. Lung adenocarcinoma, squamous lung carcinoma, and large-cell lung carcinoma are the three major histological types of NSCLC, accounting for 47%, 35%, and 12% of all NSCLC cases, respectively. The application of immune checkpoint blockade therapy (ICBT) has improved the overall prognosis of NSCLC. However, only about 15% of patients have durable benefit from it [3]. The limited response rate to immunotherapy has become a major impediment preventing further prognostic improvement in late-stage NSCLC.

Histology is an important feature of tumors and is associated with major clinical outcomes. Early-stage lung adenocarcinoma can be further divided into five histological subtypes: lepidic, solid, acinar, papillary, and micropapillary [4]. Typically, lung adenocarcinoma is composed of cells with varying histological subtypes, which is known as intratumor heterogeneity. For example, a lung adenocarcinoma tissue may show 50% of papillary, 30% of acinar, and 20% of lepidic patterns at different tumor sub-regions. As such, it has been highly recommended by pathologists to include percentages of histological subtypes during histological review [4]. In general, the lepidic subtype is associated with low-grade tumor; acinar and papillary subtypes are associated with intermediate-grade tumor; solid and micropapillary subtypes are associated with high-grade tumor [5, 6]. The lepidic subtype is associated with good prognosis [5–9], while the solid and micropapillary subtypes are associated with poor prognosis [7, 10–14]. Interestingly, the expression level of PD-L1 varies among different histological subtypes: solid and micropapillary tumors have higher levels than the other subtypes [15–17]. However, it remains unclear how these different histological subtypes vary in their immune microenvironment and whether they show different sensitivity to targeted treatments and immunotherapies.

Several studies have shown that the tumor immune microenvironment (TIME) plays critical roles in the development, progression, and metastasis of cancer [18–20]. In many different cancer types including lung adenocarcinoma, the infiltration levels of immune cells in TIME are associated with patient prognosis [21–25]. In lung adenocarcinoma, T cells seem to be the main type of infiltrating immune cell [26] and the extent of T cell infiltration is positively correlated with patient prognosis [25]. Varn et al. [22] reported that the infiltration of CD8+ T cells and naïve B cells were associated with good prognosis while the infiltration of myeloid cells was associated with poor prognosis in early-stage lung adenocarcinoma. Generally, the presence of immunosuppressive features, including the absence of T cell infiltration and the presence of suppressive macrophages, is associated with poor prognosis [25]. It has also been reported that NSCLC patients with higher PD-L1 protein level tend to have longer survival times when treated by ICBT [27–29]. Nevertheless, the effectiveness of using PD-L1 expression for predicting ICBT response is still under debate [27, 30, 31].

In this study, we analyzed the gene expression data for micro-dissected lung adenocarcinoma tumors with homogenous histological subtypes and found that the five histological subtypes varied dramatically in their immunological features. Additionally, we found the lepidic and solid subtypes had distinguishable gene expression patterns while papillary, micropapillary, and acinar subtypes shared similar expression patterns. According to these observations, we defined two gene expression signatures to characterize lepidic and solid lung adenocarcinoma subtypes, respectively. Given a gene expression profile for a lung adenocarcinoma sample, these two signatures can be used to calculate a lepidic-score (L-score) and a solid-score (S-score) that quantify the relative abundance of lepidic and solid components in the sample. We found that higher L-scores were associated with better prognosis and response to immunotherapy. Moreover, the L-scores of lung cancer cell lines were correlated with their sensitivity to targeted drugs such as EGFR inhibitors. Our analyses revealed critical links

between histological composition and clinical outcomes in lung adenocarcinoma.

Methods

Dataset collection and processing

The level 3 RNA-seq data of lung adenocarcinoma (LUAD) generated by The Cancer Genome Atlas (TCGA) were retrieved from FireBrowse (http://firebrowse.org) [32]. Processed RNA-seq data of LUAD samples from The National Cancer Institute's (NCI's) Clinical Proteomic Tumor Analysis Consortium (CPTAC) were retrieved from NCI's Genomic Data Commons (GDC) [33, 34]. The large-scale drug sensitivity data generated by the Genomics of Drug Sensitivity in Cancer (GDSC) was retrieved from the GDSC database [35]. Ten microarray datasets used in this study were retrieved from the Gene Expression Omnibus (GEO) [36] with accession numbers GSE58772 (n = 48) [37], GSE14814 (n = 133) [38], GSE8894 (n = 138) [39], GSE13213 (n = 117) [40], GSE31210 (n = 246) [41], GSE32989 (n = 69) [42], GSE93157 (n = 65) [43], GSE11969 (n = 149) [44], GSE26939 (n = 116) [45], and GSE72094 (n = 442) [46]. Microarray data from the GEO database provide expression at the probeset level and were transformed into gene-level expression by selecting the probeset with the maximum intensities to represent a gene (for data from one-channel microarray: GSE58772 [37], GSE14814 [38], GSE8894 [39], GSE31210 [41], GSE32989 [42], GSE93157 [43], GSE72094 [46]) or by calculating the average levels of probesets for the same gene (for data from two-channel microarray: GSE13213 [40], GSE11969 [44], GSE26939 [45]). Processed RNA-seq gene expression data represented as Transcript Per Million and clinical data from Bachereau et al. [47] were retrieved from the European Genome-phenome Archive with accession number EGAS00001004343.

TCGA somatic mutation and copy number variation data were also downloaded from FireBrowse (http://firebrowse.org). Only non-silent mutations from somatic mutation data were indicated in the analyses. Some genomic features such as non-synonymous mutation rate and homologous recombination pathway defect score of TCGA LUAD samples were retrieved from a published article by Thorsson et al. [48].

The Zabeck dataset GSE58772 [37] was used to perform immune infiltration analysis and to define gene signatures for lepidic and solid histological subtypes. The data contains expression profiles for 10 acinar, 10 solid, 10 lepidic, 9 micropapillary, and 9 papillary tumor samples purified by microdissection. Zhu dataset GSE14814 [38], TCGA LUAD [32], and CPTAC LUAD [33] were used to validate the signatures with provided histological subtype information. The datasets TCGA LUAD [32],

GSE8894 [39], GSE13213 [40], and GSE31210 [41], with considerable number of patients (n > 60) with high-quality patient survival times and statuses, were used for prognostic analysis. These data have survival information in the form of either overall survival (TCGA LUAD and GSE13213) or recurrence-free survival (GSE8894 and GSE31210). With a high number of patients and available somatic mutation and copy number variation data, TCGA LUAD [32] was used to investigate the association of signatures with genomic features. Additionally, datasets GSE11969 [44], GSE26939 [45], and GSE72094 [46] were used to validate the findings, with clinical information regarding *TP53/EGFR* mutation status for each patient. Regarding the association of signatures with drug sensitivity of cancer cell lines, datasets GSE32989 [42] and GDSC [35] with corresponding information were utilized. Lastly, to investigate the association of signatures with patient response to immunotherapy, Prat dataset GSE93157 [43] and Banchereau dataset [47] were implemented. The former contains 22 adenocarcinoma and 13 squamous cell carcinoma patients treated with pembrolizumab and nivolumab, and the latter contains 81 non-small cell lung cancer patients treated with atezolizumab, providing treatment outcome information for each patient.

In the analyses, only lung adenocarcinoma samples/cell lines included in these datasets were used unless specified. Details about all datasets, including the number of samples/cell lines, descriptions, and relevant publications, are included in Additional file 1: Table S1.

Computational inference of immune cell infiltration in tumor samples

We applied the previously published BASE computational algorithm [21, 22, 49, 50] to infer the infiltration level of immune cells in tumor samples based on their gene expression profiles. This algorithm estimates the infiltration levels of immune cells by examining the expression levels of immune-cell-specific genes. Full details of the algorithm have been described in a prior publication [22]. In this study, we calculated the infiltration scores of six different immune cell types including naive B cells, memory B cells, CD8+ T cells, CD4+ T cells, NK cells, and myeloid cells.

Differential gene expression analysis between different histological subtypes

We identified genes that are specifically expressed in each of the five histological subtypes using the Zabeck dataset [37]. For each subtype, we divided samples into two groups with one containing samples of this subtype and the other containing the remaining samples in the dataset. The expression of all genes was compared between the two groups by using the Student's *t*-test. The

resulting p-values were adjusted for multiple tests using the Benjamini-Hochberg method. Principal component analysis (PCA) was based on the expression of 1000 genes, which were obtained by merging the top 200 most specific genes (it may not be significant) for each subtype.

Defining gene signatures to characterize lepidic and solid histological subtypes

Here we use the lepidic subtype as the example. From the above-described differential expression analysis, we have determined the t-score and p-value for all genes by comparing lepidic samples with all other samples using Student's t-test, which resulted in two vectors {t} and {p}. Based on them, we defined a pair of weighted profiles, denoted as $w^+ = (w_1^+, w_2^+, ..., w_g^+)$ and $w^- = (w_1^-, w_2^-, ..., w_g^-)$. For gene k with a t-score of t_k and p-value of p_k, w_k^+ and w_k^- were obtained by (i) calculating $-I(t_k > 0) \log p_k$ and $-I(t_k < 0) \log p_k$, (ii) trimming at 10 to avoid extreme values, and (iii) rescaling the values to [0, 1]. The w^+ and w^- form the gene signature for the lepidic subtype with genes that are more up-regulated and down-regulated being assigned larger values in w^+ and w^-, respectively. The gene signature for solid subtype can be defined similarly.

Calculation of lepidic-scores and solid-scores in tumor samples

The defined gene signatures were then used to calculate sample-specific L/S-scores to measure the lepidic/solid components in lung adenocarcinoma samples. To this end, we applied a rank-based method similar to the gene set enrichment analysis (GSEA) [51, 52]. First, given the expression profile for a tumor sample, we sorted all genes in decreasing order of expression level to obtain a ranked expression vector $e = (e_1, e_2, ..., e_n)$. Second, we examined whether genes with high weight in w (w^+ or w^-) had a skewed distribution in {e}, and quantified skewness by comparing a foreground function $f(i)$ with a background function $b(i)$.

$$f(i) = \frac{\sum_{k=1}^{i}|e_k w_k|}{\sum_{k=1}^{n}|e_k w_k|}, 1 \le i \le n$$

$$b(i) = \frac{\sum_{k=1}^{i}|e_k(1-w_k)|}{\sum_{k=1}^{n}|e_k(1-w_k)|}, 1 \le i \le n$$

While $b(i)$ represents a random background distribution, $f(i)$ captures the skewed distribution of highly informative genes in a gene expression profile. For example, if genes with high weights tend to have higher expression, $f(i)$ will increase more rapidly than $b(i)$. Third, we calculated the maximum deviation between $f(i)$ and $b(i)$ and normalized it against a null distribution

estimated from permuted data, resulting in pairs of scores, S^+ and S^- from w^+ and w^-, respectively. Fourth, the final score for this sample was calculated as $S = S^+ - S^-$. A higher L- or S-score indicates a higher relative abundance of lepidic or solid components in a tumor.

Association of L-scores and S-scores with genomic features

We used the following steps to identify genes with mutation status correlated with L-scores and S-scores of samples. First, we selected all genes that were mutated in at least 20 tumor samples in TCGA LUAD [32]. Second, for each of the selected genes, we defined two sample groups based on the mutation status of this gene. Third, we conducted the Wilcoxon rank sum test to compare the L- or S-scores of the two groups. Genes with $p < 0.01$ were selected as significant genes. A similar procedure was used to identify genes with amplification/deletion status correlated with L- and S-scores. The somatic mutation, amplification, and deletion information of genes in TCGA LUAD [32] were downloaded from Fire-Browse (http://firebrowse.org).

Statistical analysis

The R package "*survival*" was implemented to perform survival analyses. Survival distributions between two patient groups were compared using the log-rank test through the "survdiff" function. Univariate and multivariate Cox regression models were performed using the "coxph" function. Wilcoxon rank sum test and Student's t-test were performed for comparing the values between two groups of samples through the R function "wilcox.test" and "t.test," respectively. Multiple test correction was performed by using the Benjamini-Hochberg method to obtain adjusted p-values in the form of false discovery rate (FDR). p-values less than 0.05 are considered statistically significant, if stated otherwise. All statistical analyses in this study were conducted using the R software.

Results

Different histological subtypes of lung adenocarcinoma vary substantially in the tumor immune microenvironment

To compare the TIME among different histological subtypes, we investigated a gene expression dataset for 48 lung adenocarcinoma samples (Zabeck dataset GSE58772 [37]). These samples were carefully prepared by microdissection to obtain pure histology and categorized based on the tumor growth pattern. We applied the BASE computational method [21, 22, 49, 50] to determine the infiltration levels of immune cells in these samples based on their gene expression profiles. Specifically, we calculated the infiltration of six major immune

cell types: naïve B cells, memory B cells, CD8+ T cells, CD4+ T cells, NK cells, and myeloid cells. We found that different histological subtypes varied substantially in their immune cell infiltration patterns. Interestingly, lepidic growth pattern tumors tended to have low infiltration in myeloid cells but high infiltration in other immune cell types. In contrast, solid growth pattern tumors demonstrated an opposite trend. As shown, compared with other histological subtypes, lepidic growth pattern tumors had significantly higher infiltration in naïve B cells, memory B cells, CD8+ T cells, and NK cells ($p < 0.05$); solid growth pattern tumors had significantly lower infiltration of these immune cell types but significantly higher infiltration of myeloid cells (Fig. 1; Additional file 2: Fig. S1). Generally, the other three histological subtypes (acinar, papillary, and micropapillary) showed an intermediate level of immune infiltration, with the exception that acinar growth pattern tumors had the lowest infiltration level of NK cells (Fig. 1; Additional file 2: Fig. S1).

Then, we investigated the expression levels of a set of immunostimulatory and immunoinhibitory genes and found that many of these genes were differentially expressed between different histological subtypes. Lepidic growth pattern tumors showed the lowest levels of immune checkpoint B7-H3 (*CD276*) expression (Additional file 2: Fig. S2), which was in line with high infiltration levels of various immune cells in lepidic growth pattern tumors (Fig. 1). Markers of myeloid cells were expressed with a significantly higher level in the solid growth pattern tumors (Additional file 2: Fig. S2), which was consistent with the highest myeloid infiltration level in this histological subtype (Fig. 1). On the contrary, markers associated with immune activation, *CD28* and *CD70*, showed significantly higher expression levels in the lepidic growth pattern tumors (Additional file 2: Fig. S2), which, again, was consistent with the immune cell infiltration results.

Defining gene signatures to characterize lepidic and solid tumor cells in lung adenocarcinoma samples

We investigated whether gene expression patterns could distinguish different histological subtypes of lung adenocarcinoma. Using the above-described gene expression data from the Zabeck dataset [37], we identified subtype-specific genes by comparing gene expression of
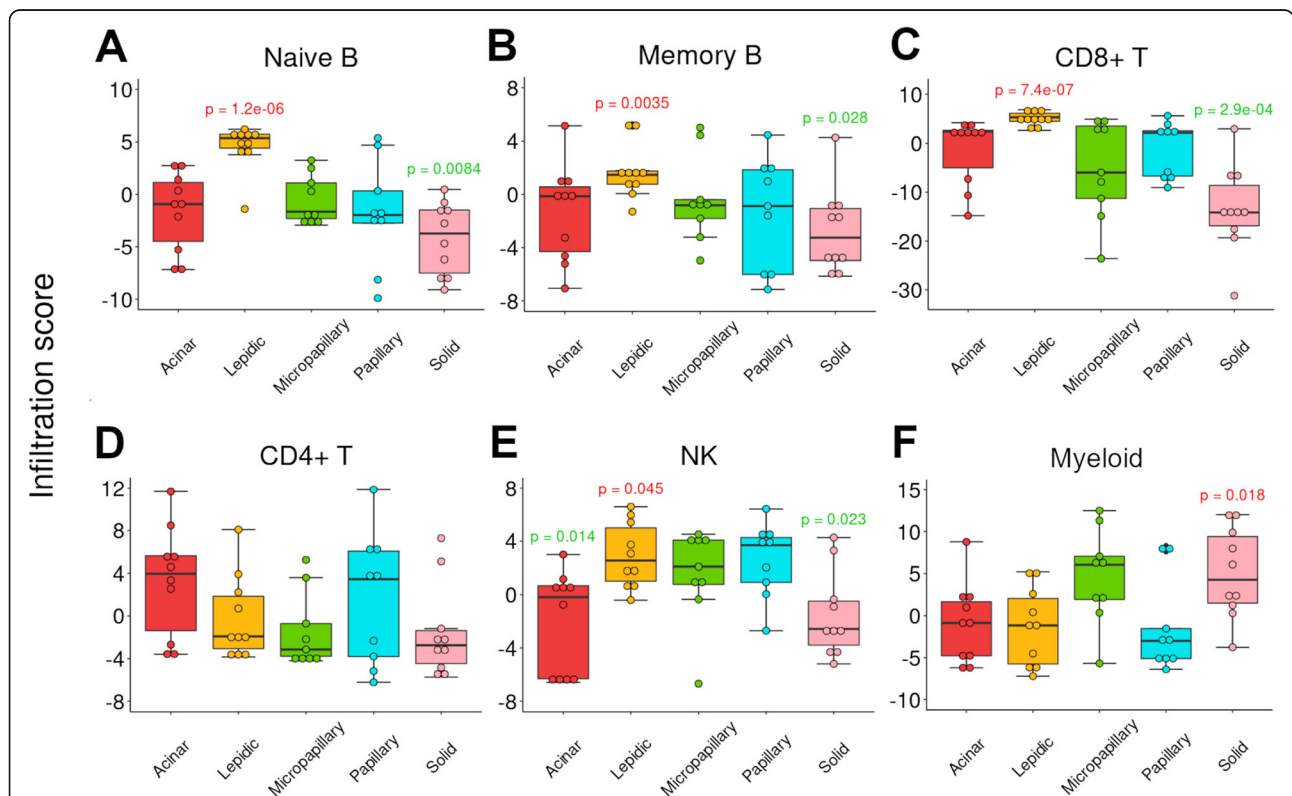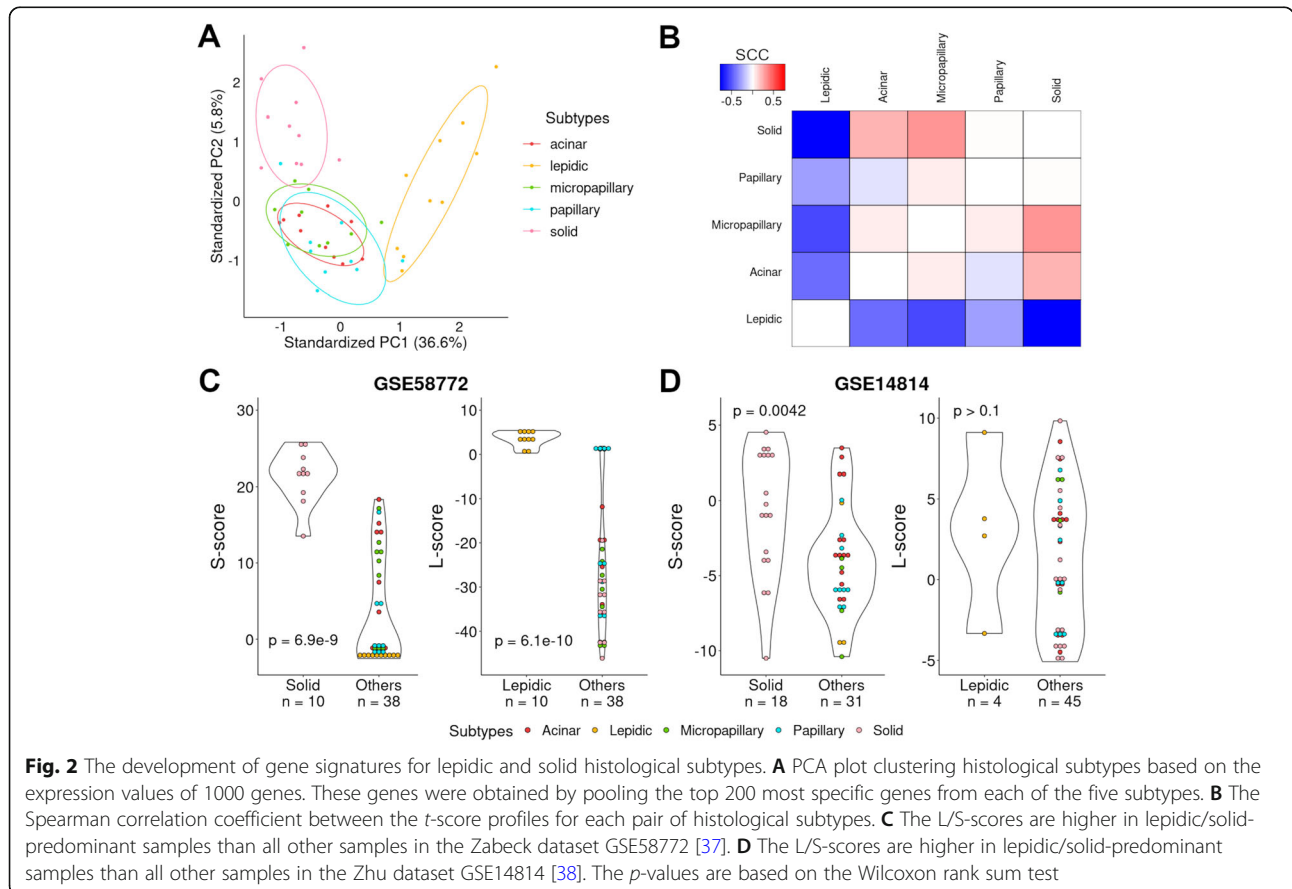

Fig. 1 The infiltration levels of immune cells vary significantly between different histological subtypes of lung adenocarcinoma. Boxplots showing the infiltration levels of **A** Naive B, **B** Memory B, **C** CD8+ T, **D** CD4+ T, **E** NK, and **F** myeloid in five different lung adenocarcinoma histological subtypes (acinar, lepidic, micropapillary, papillary, solid). *p*-values were calculated by comparing samples of the corresponding subtype with all other samples using the Wilcoxon rank sum test. *p*-values of significantly higher and lower immune infiltration levels are shown in red and green colors, respectively. The Zabeck dataset GSE58772 [37] was used in this analysis

each histological subtype with gene expression of all other subtypes using Student's *t*-test. The significant genes for each subtype (FDR < 0.01) are listed in Additional file 1: Table S2A-E. We identified 7325, 3579, 438, 532, and 244 significant genes (unadjusted *p*-value < 0.005) for the lepidic, solid, acinar, micropapillary, and papillary subtypes, respectively. We noted that no significant genes could be identified to differentiate the papillary subtype from other types after multiple test correction (FDR < 0.05). Thus, the expression patterns of lepidic and solid subtypes are more specific compared with the other subtypes. This result was confirmed by principal component analysis (PCA) based on 1000 selected genes (merging the top 200 most specific genes from each of the five histological subtypes). PCA results revealed distinct clustering of the lepidic and solid subtypes while the acinar, papillary, and micropapillary samples could not be distinguished from each other (Fig. 2A). For each subtype, we obtained a *t*-score profile that characterized its subtype-specific expression pattern. Pairwise correlation analysis between their *t*-score profiles indicated that the lepidic subtype tended to have a negative correlation with the others, especially, the solid subtype while the other four subtypes were positively correlated with each other (Fig. 2B). The same

conclusion can be made based on pairwise correlation analysis of gene expression profiles at the sample level (Additional file 2: Fig. S3).

Given these observations, we developed two gene signatures from the gene expression data of the Zabeck dataset [37] to characterize separately the lepidic and solid subtypes. These signatures can be used with gene expression data to determine the relative abundance of lepidic versus solid components (denoted as L-score and S-score, respectively) in lung adenocarcinoma samples. As shown when applied to the Zabeck dataset [37], the lepidic and solid samples showed scientifically higher L-scores and S-scores, respectively, than the other samples (Fig. 2C). This is expected, because the gene signatures for lepidic and solid subtypes are defined based on the same dataset. To further evaluate the two signatures, we applied them to the Zhu dataset GSE14814 [38], which provides histological annotation for 49 lung adenocarcinoma samples. Among these samples, 4 and 18 are annotated as lepidic-predominant and solid-predominant, respectively. In this data, the solid-predominant samples showed significantly higher S-scores than the other samples (*p* = 0.0042, Fig. 2D). Similarly, the lepidic-predominant samples tended to have higher L-scores than the other samples (Fig. 2D) but did not reach

Fig. 2 The development of gene signatures for lepidic and solid histological subtypes. **A** PCA plot clustering histological subtypes based on the expression values of 1000 genes. These genes were obtained by pooling the top 200 most specific genes from each of the five subtypes. **B** The Spearman correlation coefficient between the *t*-score profiles for each pair of histological subtypes. **C** The L/S-scores are higher in lepidic/solid-predominant samples than all other samples in the Zabeck dataset GSE58772 [37]. **D** The L/S-scores are higher in lepidic/solid-predominant samples than all other samples in the Zhu dataset GSE14814 [38]. The *p*-values are based on the Wilcoxon rank sum test

statistical significance due to the small sample size (only 4 samples are lepidic-predominant). In addition, we have also validated the signatures in TCGA and CPTAC lung adenocarcinoma datasets [32, 33], in which a small number of samples were annotated as lepidic/solid-predominant (Additional file 2: Fig. S4). These results indicated that the L- and S-scores provided reliable quantifications of the lepidic/solid components in lung adenocarcinoma. In the following analyses, we will focus on L- and S-scores separately, especially the L-score of samples, because the expression profile of the lepidic subtype deviates from and is negatively correlated with those of the other four subtypes.
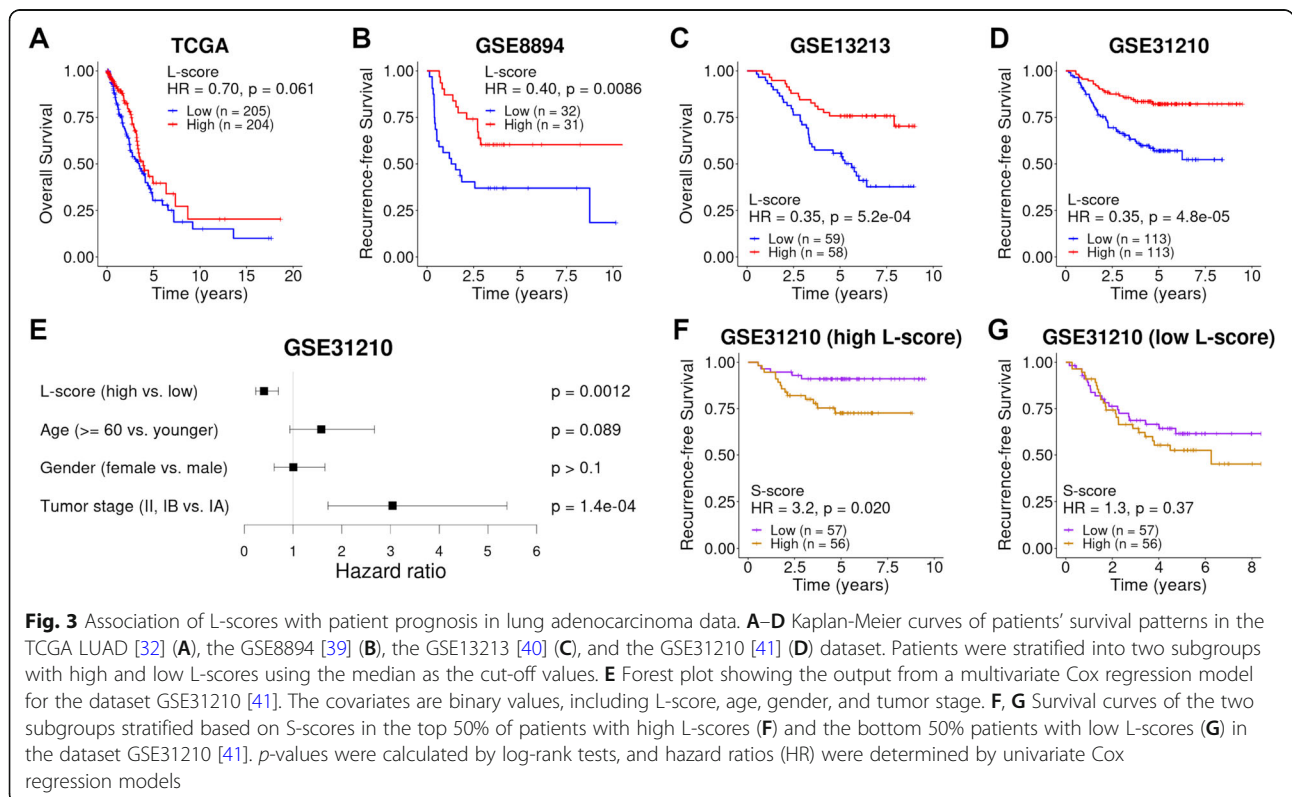
High lepidic score is associated with good prognosis

A large number of lung adenocarcinoma gene expression datasets have been generated in previous studies. For many of them, detailed clinical information was provided, but in most cases, histological composition was not available. The lepidic/solid gene signatures provide a useful tool to investigate the association between lung adenocarcinoma histology and clinical outcomes. To investigate the effect of histological composition on patient prognosis, we calculated patient-specific L/S-scores in four lung adenocarcinoma datasets, including TCGA LUAD [32], GSE8894 [39], GSE13213 [40], and GSE31210 [41]. Using the median value of the L/S-scores, we divided patients into two groups and

compared their survival. As shown, patients with high L-scores had significantly prolonged survival compared to patients with low L-scores in all datasets (Fig. 3B–D; Additional file 1: Table S3). Similar results were obtained when L-score = 0 was used to dichotomize patient samples (Additional file 1: Table S3). Note that either overall or recurrence-free survival was used depending on the availability in different datasets.

Moreover, multivariate Cox regression analysis indicated that the L-score was still significant after considering well-established clinical factors, including age, gender, and tumor stage (Fig. 3E). When the L-score was used as a continuous variable, a more significant association between L-score and prognosis was observed in both univariate and multivariate Cox regression analyses (Additional file 1: Table S3). Similarly, we performed the same analyses for the solid signature and found that the resulting S-score was also predictive of patient prognosis in lung adenocarcinoma, with higher S-scores associated with worse prognosis (Additional file 1: Table S3).

Furthermore, our results showed that combining the L-score with the S-score could further improve prognostic prediction. In Fig. 3F, we selected patients with high L-scores and further divided them into two subgroups with high and low S-scores, respectively. As shown, significantly differential survival was observed between these two subgroups. No similar pattern was found in

Fig. 3 Association of L-scores with patient prognosis in lung adenocarcinoma data. **A–D** Kaplan-Meier curves of patients' survival patterns in the TCGA LUAD [32] (**A**), the GSE8894 [39] (**B**), the GSE13213 [40] (**C**), and the GSE31210 [41] (**D**) dataset. Patients were stratified into two subgroups with high and low L-scores using the median as the cut-off values. **E** Forest plot showing the output from a multivariate Cox regression model for the dataset GSE31210 [41]. The covariates are binary values, including L-score, age, gender, and tumor stage. **F, G** Survival curves of the two subgroups stratified based on S-scores in the top 50% of patients with high L-scores (**F**) and the bottom 50% patients with low L-scores (**G**) in the dataset GSE31210 [41]. p-values were calculated by log-rank tests, and hazard ratios (HR) were determined by univariate Cox regression models

the patient subset with low L-scores (Fig. 3G). Nevertheless, these results indicated that the L-score and S-score provided related but complementary information for prognosis prediction.

Association of lepidic and solid scores with genomic features

To gain insight on what genomic features might have an impact on the histological composition in lung adenocarcinoma, we examined the association between L-score and several genomic features using TCGA LUAD [32]. We found that L-score was negatively correlated with tumor mutation burden (TMB), represented by non-silent mutation rate (SCC = −0.21, p = 2.8e−06, Fig. 4A). The L-score was also negatively correlated with tumor aneuploidy score, which was calculated as the sum of altered arms (SCC = −0.28, p = 1.5e−10, Fig. 4B). The aneuploidy score reflects the degree of chromosome instability, which is associated with the deficiency of the homologous

recombination pathway. Consistently, the L-score was also significantly correlated with the deficiency of this pathway (SCC = −0.31, p = 1.2e−12, Fig. 4C). In contrast, similar analyses with S-score and the genomic features showed positive correlations. In particular, S-score was positively correlated with TMB (SCC = 0.35, p = 1.1e−15, Additional file 2: Fig. S5A), with tumor aneuploidy score (SCC = 0.28, p = 2.1e−10, Additional file 2: Fig. S5B), and with homologous recombination pathway deficiency (SCC = 0.41, p < 2.2e−16, Additional file 2: Fig. S5C).

Following that, we identified genes whose mutation or gain/loss was correlated with the samples' L-scores. In total, the mutation statuses of 32 genes were significantly correlated with L-scores. All of them except for *MAGI2* had lower L-scores in mutated samples than the corresponding wild-type samples (Fig. 4D; Additional file 1: Table S4A). In other words, samples with mutations in these genes tend to have a lower lepidic composition in tumors. This result aligns with the association of L-
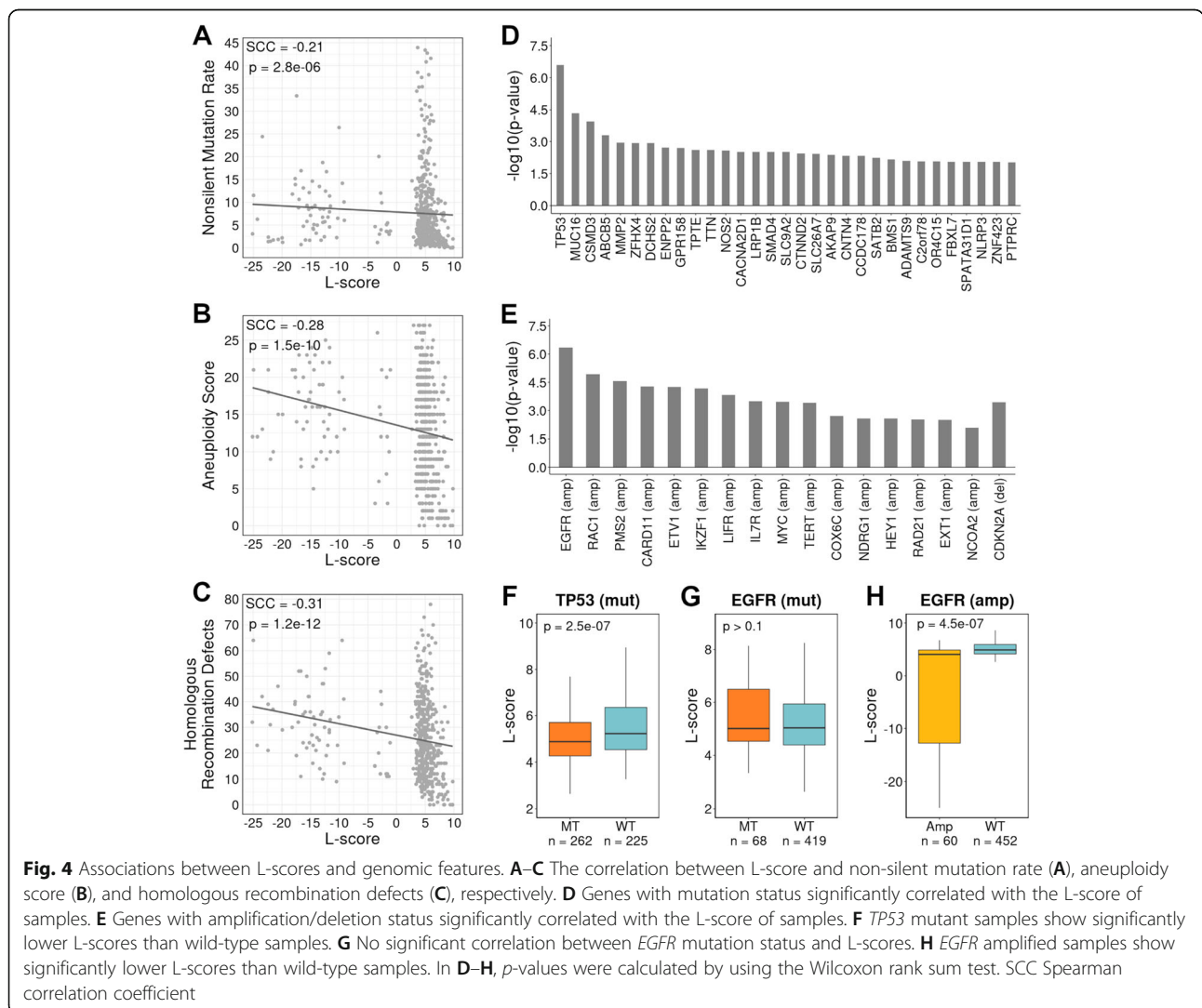

Fig. 4 Associations between L-scores and genomic features. **A–C** The correlation between L-score and non-silent mutation rate (**A**), aneuploidy score (**B**), and homologous recombination defects (**C**), respectively. **D** Genes with mutation status significantly correlated with the L-score of samples. **E** Genes with amplification/deletion status significantly correlated with the L-score of samples. **F** *TP53* mutant samples show significantly lower L-scores than wild-type samples. **G** No significant correlation between *EGFR* mutation status and L-scores. **H** *EGFR* amplified samples show significantly lower L-scores than wild-type samples. In **D–H**, *p*-values were calculated by using the Wilcoxon rank sum test. SCC Spearman correlation coefficient

scores with a good prognosis as described in the previous section. In addition, samples with amplification in 16 genes had higher L-scores than wild-type samples, and samples with *CDKN2A* deletion had higher L-scores than wild-type samples (Fig. 4E; Additional file 1: Table S4B). We performed the analyses with S-score and identified the mutation status of 279 genes, as well as the amplification status of 12 genes which were significantly correlated with S-scores (Additional file 1: Table S4C-D). There are no genes whose deletion status was significantly correlated with S-scores.

The mutation status of the *TP53* gene was significantly correlated with L-scores, with mutated samples being associated with lower L-scores compared to wild-type samples (p = 2.5e−07, Fig. 4D; Fig. 4F; Additional file 1: Table S4A). *TP53* plays critical roles in tumorigenesis [53] and its mutation correlates with poor prognosis in many cancer types [54–56]. *EGFR* is another important driver gene in lung cancer with multiple drugs targeting its protein products (EGFR inhibitors) being approved by the FDA [57–59]. The somatic mutation status of *EGFR* was not correlated with the L-score (p > 0.1, Fig. 4G). However, samples with *EGFR* amplification showed significantly lower L-scores than wild-type samples (p = 4.5e−07, Fig. 4H; Additional file 1: Table S4B). In fact, it is the most significant among the 16 genes with gain/loss status correlating with L-score in lung adenocarcinoma (Fig. 4E). To further verify this result, we combined *EGFR* mutation and amplification information to divide patients into four groups: with only mutation (mut + amp-), with only amplification (amp + mut-), with both (amp + mut+), and with neither (wt). Patients with only mutated *EGFR* (mut + amp-) had similar L-scores with the wild-type patients (wt) (p > 0.1, Additional file 2: Fig. S6). In contrast, patients with amplified *EGFR* had significantly lower L-scores than the wild-type patients with p = 5.1e−05 for amp + mut- and p = 0.0022 for amp + mut + with respect to the wt (Additional file 2: Fig. S6). These results confirmed that a reduced L-score is associated with *EGFR* amplification but not *EGFR* mutation.

In addition, we observed the opposite patterns for the association of S-scores with these genes. Samples with mutated *TP53* had higher S-scores than wild-type samples (p = 4.6e−24, Additional file 1: Table S4C; Additional file 2: Fig. S5D), and while *EGFR* mutation did not have a significant association with S-score (p > 0.1, Additional file 2: Fig. S5E), samples with amplified *EGFR* had higher S-scores than wild-type samples with the most significance (p = 2.3e−05, Additional file 1: Table S4D; Additional file 2: Fig. S5F). These results suggest that tumor histology may not just be simply affected by deregulated oncogenic pathways (e.g., overactivation of the EGFR pathway), rather, the molecular status of the driver gene product (e.g., abnormal structure vs. elevated abundance of the EGFR proteins) might also be important.

Association of lepidic and solid scores with lung cancer cell line sensitivity to targeted drugs
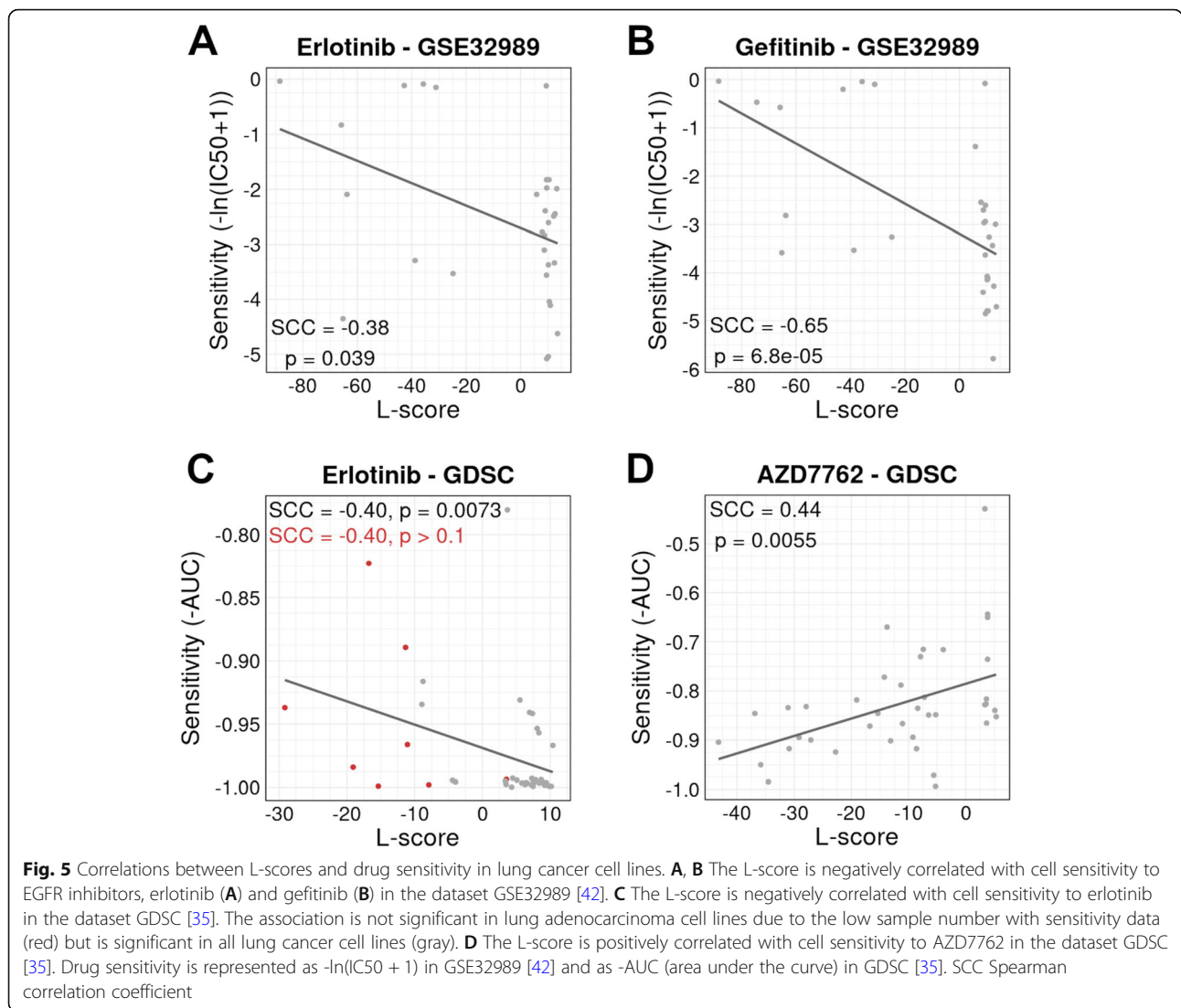
Knowing that L-score was negatively correlated with *EGFR* amplification status in lung adenocarcinoma, we then examined the association between L-score and lung adenocarcinoma cell line sensitivity to EGFR inhibitors. The GSE32989 data provides the baseline gene expression profiles of 31 lung adenocarcinoma cell lines as well as their sensitivity (IC50) to two EGFR inhibitors (erlotinib and gefitinib) [42]. In this data, we observed a significantly negative correlation between L-score and erlotinib sensitivity (SCC = −0.38, p = 0.039, Fig. 5A) and between L-score and gefitinib sensitivity (SCC = −0.65, p = 6.8e−05, Fig. 5B).

We then examined the Genomics of Drug Sensitivity in Cancer (GDSC) data, which provide cell line sensitivity information for a total of 138 different drugs (including erlotinib) in 714 cancer cell lines, including 124 lung cancer cell lines (42 are lung adenocarcinoma cell lines) [35]. In the data, we confirmed the correlation between L-score and erlotinib sensitivity in lung adenocarcinoma cell lines with SCC = −0.40 (Fig. 5C), although statistical significance was not reached due to lack of statistical power (the IC50 value was only available for only 8 cell lines). When all lung cancer cell lines were used, we observed a significant correlation with SCC = −0.40 and p = 0.0073.

In addition to EGFR inhibitors, our results indicated that the L-score was correlated with lung cancer cell line sensitivity to other drugs tested in the GDSC data. For example, AZD7762, a CHK1 inhibitor, presented a significant association with L-score (SCC = 0.44, p = 0.0055, Fig. 5D), suggesting that lung tumors with higher L-scores might be more responsive to this drug. CHK1 pathway regulating DNA damage and cell cycle responses is suggested as one of the causes of treatment resistance in lung cancer and is currently an emerging target for therapy [60–62]. Similar analyses were performed for the S-signature and identified drugs correlated with S-scores in lung cancer. For example, lung cancer cell line sensitivity to EHT 1864 (a RAC inhibitor) was correlated with S-score (SCC = −0.47, p = 0.0027, Additional file 2: Fig. S7). Inhibiting the RAC pathway has been recommended as an alternative option for treating EGFR inhibitor-resistant patients with lung cancer [63, 64].
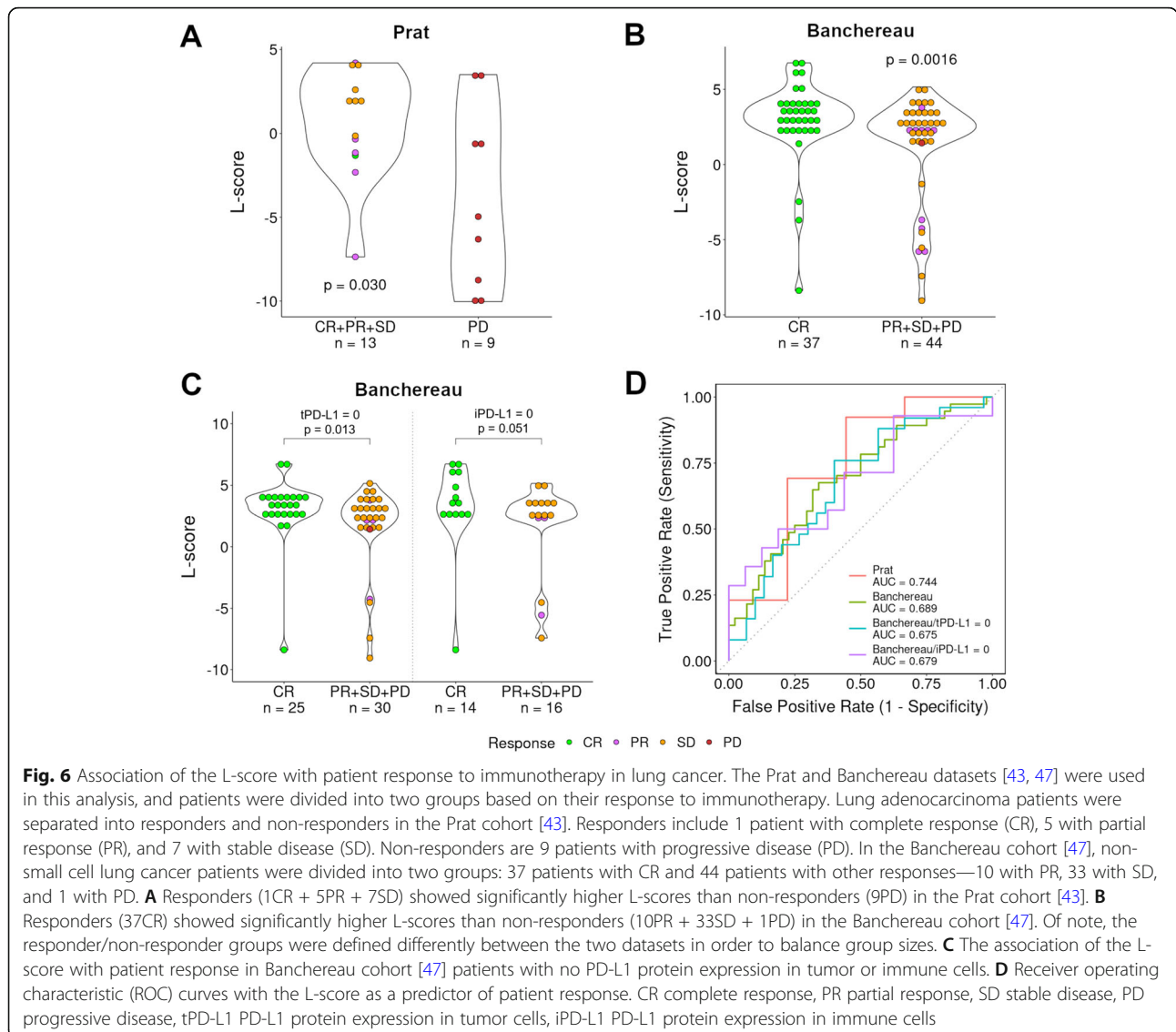
L-score is predictive of patient sensitivity to immunotherapy in lung cancer

As we have found that different histological subtypes varied dramatically in the immune microenvironment,

Fig. 5 Correlations between L-scores and drug sensitivity in lung cancer cell lines. **A**, **B** The L-score is negatively correlated with cell sensitivity to EGFR inhibitors, erlotinib (**A**) and gefitinib (**B**) in the dataset GSE32989 [42]. **C** The L-score is negatively correlated with cell sensitivity to erlotinib in the dataset GDSC [35]. The association is not significant in lung adenocarcinoma cell lines due to the low sample number with sensitivity data (red) but is significant in all lung cancer cell lines (gray). **D** The L-score is positively correlated with cell sensitivity to AZD7762 in the dataset GDSC [35]. Drug sensitivity is represented as -ln(IC50 + 1) in GSE32989 [42] and as -AUC (area under the curve) in GDSC [35]. SCC Spearman correlation coefficient

we investigated the relationship between L-score and patient response to immunotherapy. We retrieved two datasets that contained gene expression profiles and treatment outcome information for lung cancer patients. The Prat dataset (GSE93157) consists of 22 adenocarcinoma and 13 squamous cell carcinoma treated with pembrolizumab and nivolumab [43]. The Banchereau dataset consists of 81 non-small cell lung cancer treated with atezolizumab [47]. We calculated the L-scores of these patients based on their gene expression data. Based on the treatment outcomes, we divided the lung cancer patients into responders and non-responders. We observed a significantly higher L-score in the responder group compared to the non-responder group in both the Prat cohort [43] (p = 0.030, Wilcoxon rank sum test, Fig. 6A) and the Banchereau cohort [47] (p = 0.0016, Wilcoxon rank sum test, Fig. 6B). The PD-L1 protein expression in tumor (tPD-L1) or immune cells (iPD-L1) has been

measured in the Banchereau dataset [47]. We found that in patients with no PD-L1 expression the L-score was predictive of patient response with p = 0.013 for the tPD-L1 = 0 group and p = 0.051 for the iPD-L1 = 0 group (Fig. 6C). To quantify the ability of using an L-score to classify responders vs. non-responders, we determined their receiver operating characteristic (ROC) curves and the area under the curve (AUC). The L-score could classify the patient groups with AUC = 0.744 in the Prat cohort [43] (Fig. 6D), AUC = 0.689 for all (n = 81), AUC = 0.675 for tPD-L1 = 0 (n = 55), and AUC = 0.679 for 30 iPD-L1 = 0 NSCLC samples in the Banchereau cohort [47] (Fig. 6D). Furthermore, multivariate logistic regression analysis indicated the predictive value of L-score for patient response after considering established clinical factors, including age, sex, and *CD274* expression level (Additional file 1: Table S5). In contrast with L-score, we did not observe a significant association

Fig. 6 Association of the L-score with patient response to immunotherapy in lung cancer. The Prat and Banchereau datasets [43, 47] were used in this analysis, and patients were divided into two groups based on their response to immunotherapy. Lung adenocarcinoma patients were separated into responders and non-responders in the Prat cohort [43]. Responders include 1 patient with complete response (CR), 5 with partial response (PR), and 7 with stable disease (SD). Non-responders are 9 patients with progressive disease (PD). In the Banchereau cohort [47], non-small cell lung cancer patients were divided into two groups: 37 patients with CR and 44 patients with other responses—10 with PR, 33 with SD, and 1 with PD. **A** Responders (1CR + 5PR + 7SD) showed significantly higher L-scores than non-responders (9PD) in the Prat cohort [43]. **B** Responders (37CR) showed significantly higher L-scores than non-responders (10PR + 33SD + 1PD) in the Banchereau cohort [47]. Of note, the responder/non-responder groups were defined differently between the two datasets in order to balance group sizes. **C** The association of the L-score with patient response in Banchereau cohort [47] patients with no PD-L1 protein expression in tumor or immune cells. **D** Receiver operating characteristic (ROC) curves with the L-score as a predictor of patient response. CR complete response, PR partial response, SD stable disease, PD progressive disease, tPD-L1 PD-L1 protein expression in tumor cells, iPD-L1 PD-L1 protein expression in immune cells

between S-score and immunotherapy response in both datasets.

Discussion

In this study, we identified significantly differential immune cell infiltration between the five different histological subtypes of lung adenocarcinoma using a histology-specific gene expression data from microdissected tumor samples. We found that the lepidic and solid subtypes could be determined based on their gene expression patterns, while the other three subtypes (acinar, micropapillary, and papillary) had similar expression patterns. Motivated by these observations, we defined two gene signatures and used them to determine the relative abundance of lepidic and solid tumor components in lung adenocarcinoma samples. We found that L-scores and S-scores were significantly associated with

patient prognosis and with genomic features such as TMB in lung adenocarcinoma. In particular, the L-scores of lung cancer cell lines were negatively correlated with their sensitivity to EGFR inhibitors. Moreover, using an immunotherapy cohort data, we found that the L-score was significantly associated with patient response to ICBT in lung cancer.

Consistent with previous studies, our results indicated that samples with higher L-scores (indicating higher proportion of lepidic histology) tended to have a better prognosis, whereas higher S-scores were associated with a worse prognosis. These results might be explained at least partially by results from our immune infiltration analysis. The lepidic subtype had the highest levels of B cells, CD8+ T cells, and NK cells, which are known to play positive roles during immunosurveillance, whereas the immunosuppressive myeloid cells had the lowest

infiltration. In contrast, the prognostically unfavorable solid subtype displayed the opposite pattern in immune cell infiltration.

It is well known that distinct histological subtypes are associated with different mutation landscapes in tumor cells [65, 66]. However, it remains unclear how genomic alterations of driver genes determine the tumor histology, and vice versa. In this study, we determined the significant association of L- and S-scores with tumor mutation burden and *TP53* mutation status, as well as poor association with *EGFR* mutation status. We validated our results in the immunotherapy response dataset Banchereau [47] and three independent GEO datasets: GSE11969 [44], GSE26969 [45], and GSE72094 [46] (Additional file 1: Table S6). Interestingly, in this study, we found that *EGFR* amplification but not *EGFR* mutation was significantly correlated with the L-score and S-score of tumors. Both amplification and mutation of *EGFR* are frequently observed in lung cancers, which lead to the overactivation of the EGFR singling pathway. *EGFR* amplification results in enhanced protein abundance on the membrane of tumor cells, while *EGFR* mutation gives rise to abnormal protein products. It seems that, at least in this case for *EGFR*, the oncogenic pathways associated with more protein product has a more substantial impact on histology compared with effects from an abnormal protein.

Initially, we planned to develop a deconvolution-based method to calculate the proportion of the five histological subtypes in lung adenocarcinoma samples. In fact, we have attempted to do this by using the CIBER-SORT [67] and WISP [68] algorithms. However, none of them could generate effective reference signatures for all five histological subtypes. Specifically, the gene expression patterns of acinar, papillary, and micropapillary subtypes were similar. No subtype-specific genes can be identified using the Zabeck dataset [37] after correcting multiple tests. As such, in this study, we focused on the lepidic and solid subtypes, which had expression patterns distinguishable from the other subtypes. Instead of calculating the histology proportion, we defined two gene signatures to quantify the relative abundance of lepidic and solid subtypes of tumor cells, respectively.

We showed here that a higher L-score was associated with low sensitivity to EGFR inhibitors in lung cancer cell lines. However, the predictive value of the L-signature was not validated in lung adenocarcinoma patients treated with EGFR inhibitors due to the lack of appropriate datasets. Additionally, we showed that this signature is predictive of patient response to immunotherapy in the Prat and the Banchereau datasets [43, 47]. As another caveat, we used all NSCLC samples in the Banchereau cohort [47], since the histological types (adenocarcinoma or squamous) of the tumor samples were not provided. According to our results, we expect that both EGFR-targeted therapy and immunotherapy might change the histological composition of lung adenocarcinoma. Specifically, the tumor subclones of lepidic subtypes are likely to survive from EGFR-targeted treatment but may be killed by immunotherapy. We have not found data to perform this analysis, but it would be interesting to examine how the L-score changes before and after therapeutic treatment when data become available.

The current L/S-signature is a whole gene signature that includes all genes with weights assigned according to their subtype specificity. The signature can be further simplified by choosing the top most informative (i.e., highly weighted) genes to facilitate its clinical application. The L/S-scores indicate the relative abundance of L/S-subtype tumor cells, but do not automatically identify subtype predominance. To identify L/S-predominant tumor samples, the scores can be compared with reference scores that are produced for L/S-specific reference samples using the same gene expression platform.

Conclusions

In conclusion, we have defined gene signatures to quantify the relative abundance of lepidic and solid subtypes in lung adenocarcinoma. By using these signatures, we detected important associations between the L/S-score and clinical outcomes, which suggested potential clinical translation. The framework developed in this study can also be applied to other cancer types with heterogeneous subtypes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-021-01010-w.

Additional file 1: Table S1-S7.

Additional file 2: Figs. S1-S7.

Declarations

Ethics approval and consent to participate

All datasets used in this study have been previously published.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX 77030, USA. [2]Division of General Thoracic Surgery, Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX 77030, USA. [3]Department of Thoracic/Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [4]Department of Imaging Physics, Division of Diagnostic Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [5]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [6]Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA. [7]Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020; 70(1):7–30. https://doi.org/10.3322/caac.21590.
2. Gridelli C, Rossi A, Carbone DP, Guarize J, Karachaliou N, Mok T, et al. Non-small-cell lung cancer. Nat Rev Dis Prim. 2015;1(1):15009. https://doi.org/10.1038/nrdp.2015.9.
3. Garon EB, Hellmann MD, Rizvi NA, Carcereny E, Leighl NB, Ahn M-J, et al. Five-year overall survival for patients with advanced non–small-cell lung cancer treated with pembrolizumab: results from the phase I KEYNOTE-001 study. J Clin Oncol. 2019;37(28):2518–27. https://doi.org/10.1200/JCO.19.00934.
4. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization classification of lung tumors. J Thorac Oncol. 2015;10(9):1243–60. https://doi.org/10.1097/JTO.0000000000000630.
5. Yoshizawa A, Motoi N, Riely GJ, Sima CS, Gerald WL, Kris MG, et al. Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases. Mod Pathol. 2011;24(5):653–64. https://doi.org/10.1038/modpathol.2010.232.
6. Warth A, Muley T, Meister M, Stenzinger A, Thomas M, Schirmacher P, et al. The Novel Histologic International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification system of lung adenocarcinoma is a stage-independent predictor of survival. J Clin Oncol. 2012;30(13):1438–46. https://doi.org/10.1200/JCO.2011.37.2185.
7. Russell PA, Wainer Z, Wright GM, Daniels M, Conron M, Williams RA. Does lung adenocarcinoma subtype predict patient survival?: a clinicopathologic study based on the New International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Lung Adeno. J Thorac Oncol. 2011;6(9):1496–504. https://doi.org/10.1097/JTO.0b013e318221f701.
8. Yoshizawa A, Sumiyoshi S, Sonobe M, Kobayashi M, Fujimoto M, Kawakami F, et al. Validation of the IASLC/ATS/ERS lung adenocarcinoma classification for prognosis and association with EGFR and KRAS gene mutations: analysis of 440 Japanese patients. J Thorac Oncol. 2013;8(1):52–61. https://doi.org/10.1097/JTO.0b013e3182769aa8.
9. Kadota K, Villena-Vargas J, Yoshizawa A, Motoi N, Sima CS, Riely GJ, et al. Prognostic significance of adenocarcinoma in situ, minimally invasive adenocarcinoma, and nonmucinous lepidic predominant invasive adenocarcinoma of the lung in patients with stage I disease. Am J Surg Pathol. 2014;38(4):448–60. https://doi.org/10.1097/PAS.0000000000000134.
10. Song Z, Zhu H, Guo Z, Wu W, Sun W, Zhang Y. Prognostic value of the IASLC/ATS/ERS classification in stage I lung adenocarcinoma patients—based on a hospital study in China. Eur J Surg Oncol. 2013;39(11):1262–8. https://doi.org/10.1016/j.ejso.2013.08.026.
11. Tsuta K, Kawago M, Inoue E, Yoshida A, Takahashi F, Sakurai H, et al. The utility of the proposed IASLC/ATS/ERS lung adenocarcinoma subtypes for disease prognosis and correlation of driver gene alterations. Lung Cancer. 2013;81(3):371–6. https://doi.org/10.1016/j.lungcan.2013.06.012.
12. Yanagawa N, Shiono S, Abiko M, Ogata S, Sato T, Tamura G. New IASLC/ATS/ERS classification and invasive tumor size are predictive of disease recurrence in stage I lung adenocarcinoma. J Thorac Oncol. 2013;8(5):612–8. https://doi.org/10.1097/JTO.0b013e318287c3eb.
13. Yanagawa N, Shiono S, Abiko M, Ogata S, Sato T, Tamura G. The correlation of the International Association for the Study of Lung Cancer (IASLC)/American Thoracic Society (ATS)/European Respiratory Society (ERS) classification with prognosis and EGFR mutation in lung adenocarcinoma. Ann Thorac Surg. 2014;98(2):453–8. https://doi.org/10.1016/j.athoracsur.2014.04.108.
14. Woo T, Okudela K, Mitsui H, Tajiri M, Yamamoto T, Rino Y, et al. Prognostic value of the IASLC/ATS/ERS classification of lung adenocarcinoma in stage I disease of Japanese cases. Pathol Int. 2012;62(12):785–91. https://doi.org/10.1111/pin.12016.
15. Tancoš V, Grendár M, Farkašová A, Huťka Z, Mičák J, Kviatkovská Z, et al. Programmed death ligand 1 protein expression, histological tumour differentiation and intratumoural heterogeneity in pulmonary adenocarcinoma. Pathology. 2020;52(5):538–45. https://doi.org/10.1016/j.pathol.2020.03.012.
16. Miyazawa T, Marushima H, Saji H, Kojima K, Hoshikawa M, Takagi M, et al. PD-L1 expression in non-small-cell lung cancer including various adenocarcinoma subtypes. Ann Thorac Cardiovasc Surg. 2019;25(1):1–9. https://doi.org/10.5761/atcs.oa.18-00163.
17. Farkašová A, Tancoš V, Kviatkovská Z, Huťka Z, Mičák J, Scheerová K, et al. Clinicopathological analysis of programmed death-ligand 1 testing in tumor cells of 325 patients with non-small cell lung cancer: Its predictive and potential prognostic value. Cesk Patol. 2018;54(3):137–42.
18. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24(5):541–50. https://doi.org/10.1038/s41591-018-0014-x.
19. Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. Cancer Res. 2019;79(18):4557–66. https://doi.org/10.1158/0008-5472.CAN-18-3962.
20. Lei X, Lei Y, Li J-K, Du W-X, Li R-G, Yang J, et al. Immune cells within the tumor microenvironment: biological functions and roles in cancer

immunotherapy. Cancer Lett. 2020;470:126–33. https://doi.org/10.1016/j.ca nlet.2019.11.009.

21. Varn FS, Schaafsma E, Wang Y, Cheng C. Genomic characterization of six virus-associated cancers identifies changes in the tumor immune microenvironment and altered genetic programs. Cancer Res. 2018;78(22): 6413–23. https://doi.org/10.1158/0008-5472.CAN-18-1342.

22. Varn FS, Tafe LJ, Amos CI, Cheng C. Computational immune profiling in lung adenocarcinoma reveals reproducible prognostic associations with implications for immunotherapy. Oncoimmunology. 2018;7(6):e1431084. https://doi.org/10.1080/2162402X.2018.1431084.

23. Petitprez F, de Reyniès A, Keung EZ, Chen TW-W, Sun C-M, Calderaro J, et al. B cells are associated with survival and immunotherapy response in sarcoma. Nature. 2020;577(7791):556–60. https://doi.org/10.1038/s41586-01 9-1906-8.

24. Ye L, Zhang T, Kang Z, Guo G, Sun Y, Lin K, et al. Tumor-infiltrating immune cells act as a marker for prognosis in colorectal cancer. Front Immunol. 2019;10. https://doi.org/10.3389/fimmu.2019.02368.

25. Kinoshita T, Kudo-Saito C, Muramatsu R, Fujita T, Saito M, Nagumo H, et al. Determination of poor prognostic immune features of tumour microenvironment in non-smoking patients with lung adenocarcinoma. Eur J Cancer. 2017;86:15–27. https://doi.org/10.1016/j.ejca.2017.08.026.

26. Stankovic B, Bjørhovde HAK, Skarshaug R, Aamodt H, Frafjord A, Müller E, et al. Immune cell composition in human non-small cell lung cancer. Front Immunol. 2019;9. https://doi.org/10.3389/fimmu.2018.03101.

27. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. N Engl J Med. 2015;373(17):1627–39. https://doi.org/10.1056/ NEJMoa1507643.

28. Huang Q, Zhang H, Hai J, Socinski MA, Lim E, Chen H, et al. Impact of PD-L1 expression, driver mutations and clinical characteristics on survival after anti-PD-1/PD-L1 immunotherapy versus chemotherapy in non-small-cell lung cancer: a meta-analysis of randomized trials. Oncoimmunology. 2018;7(12): e1396403. https://doi.org/10.1080/2162402X.2017.1396403.

29. Xu Y, Wan B, Chen X, Zhan P, Zhao Y, Zhang T, et al. The association of PD-L1 expression with the efficacy of anti-PD-1/PD-L1 immunotherapy and survival of non-small cell lung cancer patients: a meta-analysis of randomized controlled trials. Transl Lung Cancer Res. 2019;8:413–28. https:// doi.org/10.21037/tlcr.2019.08.09.

30. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WEE, Poddubskaya E, et al. Nivolumab versus docetaxel in advanced squamous-cell non–small-cell lung cancer. N Engl J Med. 2015;373(2):123–35. https://doi.org/10.1056/ NEJMoa1504627.

31. Carbognin L, Pilotto S, Milella M, Vaccaro V, Brunelli M, Caliò A, et al. Differential activity of nivolumab, pembrolizumab and MPDL3280A according to the tumor expression of programmed death-ligand-1 (PD-L1): sensitivity analysis of trials in melanoma, lung and genitourinary cancers. PLoS One. 2015;10(6):e0130142. https://doi.org/10.1371/journal. pone.0130142.

32. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511(7511):543–50. https:// doi.org/10.1038/nature13385.

33. Edwards NJ, Oberti M, Thangudu RR, Cai S, McGarvey PB, Jacob S, et al. The CPTAC data portal: a resource for cancer proteomics research. J Proteome Res. 2015;14(6):2707–13. https://doi.org/10.1021/pr501254j.

34. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. N Engl J Med. 2016; 375(12):1109–12. https://doi.org/10.1056/NEJMp1607591.

35. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2012;41(D1):D955–61. https://doi.org/10.1093/nar/gks1111.

36. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30(1):207–10. https://doi.org/10.1093/nar/30.1.207.

37. Zabeck H, Dienemann H, Hoffmann H, Pfannschmidt J, Warth A, Schnabel PA, et al. Molecular signatures in IASLC/ATS/ERS classified growth patterns of lung adenocarcinoma. PLoS One. 2018;13(10):e0206132. https://doi.org/1 0.1371/journal.pone.0206132.

38. Zhu C-Q, Ding K, Strumpf D, Weir BA, Meyerson M, Pennell N, et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non–small-cell lung cancer. J Clin Oncol. 2010;28(29):4417–24. https://doi.org/10.1200/JCO.2009.26.4325.

39. Lee E-S, Son D-S, Kim S-H, Lee J, Jo J, Han J, 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(22):7397–404. https://doi.org/10.1158/1078-0432.CCR-07-4937.

40. Tomida S, Takeuchi T, Shimada Y, Arima C, Matsuo K, Mitsudomi T, et al. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. J Clin Oncol. 2009;27(17):2793–9. https://doi. org/10.1200/JCO.2008.19.7053.

41. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res. 2012;72:100–11. https://doi. org/10.1158/0008-5472.CAN-11-1403.

42. Byers LA, Diao L, Wang J, Saintigny P, Girard L, Peyton M, et al. An epithelial–mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. Clin Cancer Res. 2013;19(1):279–90. https://doi.org/10.1158/1078-0432.CCR-12-1558.

43. Prat A, Navarro A, Paré L, Reguart N, Galván P, Pascual T, et al. Immune-related gene expression profiling after PD-1 blockade in non–small cell lung carcinoma, head and neck squamous cell carcinoma, and melanoma. Cancer Res. 2017;77(13):3540–50. https://doi.org/10.1158/0008-5472.CAN-1 6-3556.

44. Takeuchi T, Tomida S, Yatabe Y, Kosaka T, Osada H, Yanagisawa K, et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol. 2006;24(11):1679–88. https://doi. org/10.1200/JCO.2005.03.8224.

45. Wilkerson MD, Yin X, Walter V, Zhao N, Cabanski CR, Hayward MC, et al. Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. PLoS One. 2012;7(5):e36530. https://doi.org/10.1371/journal. pone.0036530.

46. Schabath MB, Welsh EA, Fulp WJ, Chen L, Teer JK, Thompson ZJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. Oncogene. 2016;35(24):3209–16. https://doi.org/10.1038/ onc.2015.375.

47. Banchereau R, Leng N, Zill O, Sokol E, Liu G, Pavlick D, et al. Molecular determinants of response to PD-L1 blockade across tumor types. Nat Commun. 2021;12(1):3969. https://doi.org/10.1038/s41467-021-24112-w.

48. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The immune landscape of cancer. Immunity. 2018;48:812–830.e14. https:// doi.org/10.1016/j.immuni.2018.03.023.

49. 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(6):1271–82. https://doi.org/10.1158/0008-5472.CAN-1 6-2490.

50. 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(1):10248. https://doi.org/1 0.1038/ncomms10248.

51. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003;34(3):267–73. https://doi.org/10.1038/ng1180.

52. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005; 102(43):15545–50. https://doi.org/10.1073/pnas.0506580102.

53. Hao Q, Cho W. Battle against cancer: an everlasting saga of p53. Int J Mol Sci. 2014;15(12):22109–27. https://doi.org/10.3390/ijms151222109.

54. Liu H, Zhou M. Evaluation of p53 gene expression and prognosis characteristics in uveal melanoma cases. Onco Targets Ther. 2017;10:3429–34. https://doi.org/10.2147/OTT.S136785.

55. Zhao J, Han Y, Li J, Chai R, Bai C. Prognostic value of KRAS/TP53/PIK3CA in non-small cell lung cancer. Oncol Lett. 2019;17(3):3233–40. https://doi.org/1 0.3892/ol.2019.10012.

56. Chen H, Li Y, Long Y, Tang E, Wang R, Huang K, et al. Increased p16 and p53 protein expression predicts poor prognosis in mucosal melanoma. Oncotarget. 2017;8:53226–33. https://doi.org/10.18632/oncotarget.18367.

57. Cheng C, Zhao Y, Schaafsma E, Weng Y, Amos C. An EGFR signature predicts cell line and patient sensitivity to multiple tyrosine kinase inhibitors. Int J Cancer. 2020;147(9):2621–33. https://doi.org/10.1002/ijc.33053.

58. Xu MJ, Johnson DE, Grandis JR. EGFR-targeted therapies in the post-genomic era. Cancer Metastasis Rev. 2017;36(3):463–73. https://doi.org/10.1007/s10555-017-9687-8.

59. Solassol I, Pinguet F, Quantin X. FDA- and EMA-approved tyrosine kinase inhibitors in advanced EGFR-mutated non-small cell lung cancer: safety, tolerability, plasma concentration monitoring, and management. Biomolecules. 2019;9(11):668. https://doi.org/10.3390/biom9110668.

60. Zhang Y, Hunter T. Roles of Chk1 in cell biology and cancer therapy. Int J cancer. 2014;134(5):1013–23. https://doi.org/10.1002/ijc.28226.

61. Grabauskiene S, Bergeron EJ, Chen G, Thomas DG, Giordano TJ, Beer DG, et al. Checkpoint kinase 1 protein expression indicates sensitization to therapy by checkpoint kinase 1 inhibition in non-small cell lung cancer. J Surg Res. 2014;187(1):6–13. https://doi.org/10.1016/j.jss.2013.12.016.

62. Qiu Z, Oleinick NL, Zhang J. ATR/CHK1 inhibitors and cancer therapy. Radiother Oncol. 2018;126(3):450–64. https://doi.org/10.1016/j.radonc.2017.09.043.

63. Kaneto N, Yokoyama S, Hayakawa Y, Kato S, Sakurai H, Saiki I. RAC1 inhibition as a therapeutic target for gefitinib-resistant non-small-cell lung cancer. Cancer Sci. 2014;105(7):788–94. https://doi.org/10.1111/cas.12425.

64. Zou T, Mao X, Yin J, Li X, Chen J, Zhu T, et al. Emerging roles of RAC1 in treating lung cancer patients. Clin Genet. 2017;91(4):520–8. https://doi.org/10.1111/cge.12908.

65. Dong Y-J, Cai Y-R, Zhou L-J, Su D, Mu J, Chen X-J, et al. Association between the histological subtype of lung adenocarcinoma, EGFR/KRAS mutation status and the ALK rearrangement according to the novel IASLC/ATS/ERS classification. Oncol Lett. 2016;11(4):2552–8. https://doi.org/10.3892/ol.2016.4233.

66. Ding Y, Zhang L, Guo L, Wu C, Zhou J, Zhou Y, et al. Comparative study on the mutational profile of adenocarcinoma and squamous cell carcinoma predominant histologic subtypes in Chinese non-small cell lung cancer patients. Thorac Cancer. 2020;11(1):103–12. https://doi.org/10.1111/1759-7714.13208.

67. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453–7. https://doi.org/10.1038/nmeth.3337.

68. Van Wart AT, Durrant J, Votapka L, Amaro RE. Weighted implementation of suboptimal paths (WISP): an optimized algorithm and tool for dynamical network analysis. J Chem Theory Comput. 2014;10(2):511–7. https://doi.org/10.1021/ct4008603.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.