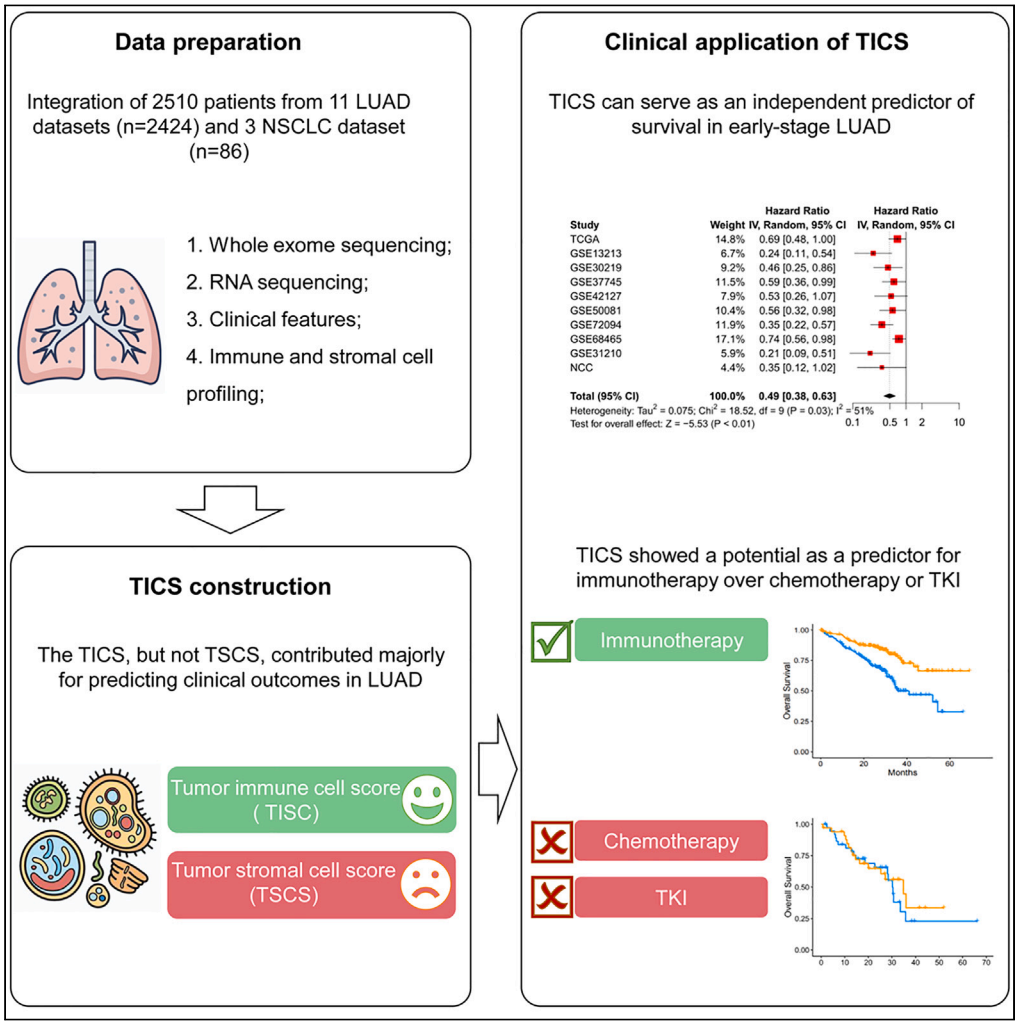
iScience

CellPress
OPEN ACCESS

Article

A prognostic and immunotherapeutic predictive model based on the cell-originated characterization of tumor microenvironment in lung adenocarcinoma



Jiachen Xu, Zhenlin Yang, Wenchuan Xie, ..., Shangli Cai, Jie Wang, Zhijie Wang

jie_969@163.com

Highlights

Evaluation of the prognostic and predictive value of immune cell signatures in LUAD

The TICS, but not TSCS, contributed majorly to predicting clinical outcomes in LUAD

TICS can serve as an independent predictor of survival in the early-stage LUAD

TICS may serve as a predictor for the benefit of immunotherapy in advanced LUAD

Xu et al., iScience 26, 106616
May 19, 2023
https://doi.org/10.1016/j.isci.2023.106616

iScience

CellPress
OPEN ACCESS

Article

A prognostic and immunotherapeutic predictive model based on the cell-originated characterization of tumor microenvironment in lung adenocarcinoma

Jiachen Xu,[1,2,6] Zhenlin Yang,[3,6] Wenchuan Xie,[4,6] Rui Wan,[1] Chengcheng Li,[4] Kailun Fei,[1] Boyang Sun,[1] Xu Yang,[1] Ping Chen,[5] Fanqi Meng,[4] Guoqiang Wang,[4] Jing Zhao,[4] Yusheng Han,[4] Shangli Cai,[4] Jie Wang,[1] and Zhijie Wang[1,7,*]

SUMMARY

Tumor microenvironment (TME) plays a crucial role in predicting prognosis and response to therapy in lung cancer. Our study established a prognostic and immunotherapeutic predictive model, the tumor immune cell score (TICS), by differentiating cell origins in lung adenocarcinoma (LUAD) based on the transcriptomic data of 2,510 patients in 14 independent cohorts, including 12 public datasets and two in-house cohorts. The high TICS was associated with prolonged overall survival (OS), especially in the early-stage LUAD. For the advanced-stage LUAD, high TICS predicted a superior OS in patients who were treated with immunotherapy instead of chemotherapy or TKI. The result suggested that TICS could serve as an indicator for the prognostic stratification management of patients in the early-stage LUAD, and as a potential guide for therapeutic decision-marking in the advanced-stage LUAD. Our findings provided an insight into prognosis stratification and potential guidance for treatment strategy selection.

INTRODUCTION

Lung adenocarcinoma (LUAD) is the major pathological subtype of lung cancer, which is the leading cause of cancer-related death worldwide.[1,2] Although comprehensive treatments including immune checkpoint inhibitors (ICIs) have revolutionized the management of LUAD, bringing unprecedented clinical benefit and improved survival, there are still several unmet clinical needs. For example, the 5-year survival rate remains low to be 19%, and the limited proportion of patients have durable response, partly because of the high heterogeneity of LUAD and incomprehensive characterization based on current prognostic and predictive indicators.[3–5] Therefore, it is vital to explore and develop robust prognostic and predictive models based on the comprehensive cellular and molecular characterization in LUAD.

In recent years, tumor microenvironment (TME), comprising a mass of heterogeneous cell types including tumor cells, immune cells, stromal cells, as well as extracellular matrix (ECM) and inflammatory cytokines, played a crucial role in predicting prognosis and response to therapy.[6–11] Previous studies have shown that several immune cells, such as B cells, $CD8^+$ T cells, $CD4^+$ T cells, M1 macrophages, are associated with favorable clinical outcomes in multiple cancers.[12–17] In addition to immune cells, stromal cells can also regulate tumor immunophenotyping, such as ECM, cancer-associated fibroblast (CAFs), and mesenchymal-epithelial transition (EMT), which promote cancer growth, invasion, and metastasis.[18,19] The expression of PD-L1 on stromal cells was reported to be associated with the progression of colon cancer,[20] and CAFs were correlated with poor survival and could promote tumor progression in the breast cancer and LUAD.[21,22] Furthermore, the aberrant status of immune and stromal cells does not only impact the prognosis, but can also indicate immunotherapeutic efficacy. For example, TME has been used to predict immunotherapeutic responsiveness in melanoma and bladder cancer.[23] However, previous studies seldom classified the specific cell-derived signatures, leading to the unsatisfactory reproducibility of transcriptomics-based models, due to the various environmental conditions including tumor purity, immune cell infiltration or stromal context and so forth, thus making them less robust in the clinical practice.

[1]State Key Laboratory of Molecular Oncology, Department of Medical Oncology, National Cancer Center/ National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100021, P.R. China

[2]Guangdong Provincial People's Hospital/Guangdong Provincial Academy of Medical Sciences, Guangdong Provincial Key Lab of Translational Medicine in Lung Cancer, Guangdong 510317, P.R. China

[3]Department of Thoracic Surgery, National Cancer Center/ National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100021, P.R. China

[4]Burning Rock Biotech, Guangdong 510300, P.R. China

[5]Department of Oncology, Yancheng First Hospital, Affiliated Hospital of Nanjing University Medical School, The First People's Hospital of Yancheng; Jiangsu 224001, P.R. China

[6]These authors contributed equally

[7]Lead contact

*Correspondence: jie_969@163.com

https://doi.org/10.1016/j.isci.2023.106616

Herein, we calculated the enrichment scores of 64 immune and stromal cells based on the xCell algorithm, Subsequently, we established a prognostic and immunotherapeutic predictive model, the tumor immune cell score (TICS), based on the transcriptomic data of 2,510 patients in 14 independent cohorts. The high TICS was associated with a prolonged overall survival (OS), especially in the early-stage LUAD. For the advanced-stage LUAD, high TICS predicted a superior OS in patients who were treated with immunotherapy instead of chemotherapy or TKI. These results suggested that for the early-stage LUAD, TICS could serve as an indicator for the prognostic stratification management of patients, and for the advanced-stage LUAD, TICS shows the potential to guide therapeutic decision-marking. Our study provided an insight into prognosis stratification and potential guidance for treatment strategy selection.

RESULTS

Prognostic potency of tumor immune cell score and tumor stromal cell score in lung adenocarcinoma in the training cohort

In total, 2,510 patients with available clinical information and transcriptome data were obtained from 12 public datasets (TCGA and GEO databases) and 2 in-house cohorts (NCC and NCC-ICIs cohort) in this study. The demographic and clinical features of each public cohort are listed in Table 1 and the overall analysis workflow was illustrated in Figure 1. Then the enrichment scores for 43 immune cell types (including 9 hematopoietic stem cells, 21 lymphoid cells, and 13 myeloid cells) and 21 stromal cell types (including 7 epithelial cells and 14 stroma cells) were generated based on the mRNA expression in each cohort, respectively (Figure S1, Table S1). The enrichment scores of each group were relatively balanced.

The prognostic value of immune and stromal cells was analyzed by univariable Cox regression in six training datasets (Figure 2A). Most cell types showed consistent prognostic performance although with inconsistent statistical significance. Then meta-analyses of the HRs of each cell type were performed in the training cohorts. In total, 26 immune cells and 5 stromal cells were significantly associated with OS in LUAD ($p < 0.050$, Figure 2B, Table S2), among which, 24 (77 · 4%) cell types were associated with better OS ($HR < 1.00$, $p < 0.05$).

To evaluate the prognostic value of stromal cells and immune cells respectively, we constructed TSCS and TICS. The combined score (CS) was calculated by adding them together. The high-TICS and high-CS could predict a superior OS in all six training cohorts, while TSCS was no longer a predictor for OS in 5 out of 6 training cohorts (Figure S2). We further performed a pooled analysis in the six training cohorts. The higher TSCS (HR 0.73, 95% CI 0.60–0.88, $p < 0.010$, Figure 2C), higher TICS (HR 0.55, 95% CI 0.45–0.66, $p < 0.010$, Figure 2D) and higher CS (HR 0.53, 95% CI 0.44–0.65, $p < 0.010$, Figure 2E) exhibited better OS in the meta-analysis by fixed-effect model. Statistical analyses for heterogeneity were insignificant in all pooled estimates ($p > 0.100$), indicating the consistency of the association between TSCS or TICS and OS across these cohorts. However, the higher CS score exhibited a slightly higher HR compared with TICS (0.55 vs. 0.53, Figure 2E), implying that the TICS instead of TSCS made the major contributions to the prognosis in LUAD. Therefore, the subsequent analysis mainly focused on TICS.

Prognostic potency of tumor immune cell score in the public validation cohorts

To validate the prognostic performance of TICS, three GEO datasets were included as independent validation cohorts. Consistently, patients in the high TICS group had a remarkable survival benefit than those in the low TICS group in the validation cohorts, respectively (GSE72094, HR 0.44, 95% CI 0.29–0.64, $p < 0.001$; GSE68465, HR 0.76, 95% CI 0.59–0.98, $p = 0.035$; GSE31210, HR 0.21, 95% CI 0.09–0.51, $p < 0.001$; Figures 3A–3C). Pooled analysis revealed consistent prolonged OS in the high TICS group (HR 0.46, 95% CI 0.25–0.85, $p = 0.010$; Figure 3D).

To further evaluate the independent prognostic ability of the TICS signature, we performed multivariable Cox regression analyses in the training and validation cohorts. After adjusting for clinicopathological features including age, sex, stage, smoking status and mutations in driver genes, the association between TICS and OS remained significant in each cohort (Figure 3E; Tables S3–S11). These results consistently indicated that the TICS could serve as an independent prognostic predictor for OS in LUAD patients.

Prognostic potency of tumor immune cell score in the in-house validation cohort

To further validate the prognostic performance of TICS, the NCC cohort was included as independent in-house validation cohort. The NCC cohort included 203 patients with stage I-III LUAD treated with surgery.

Table 1. The clinicopathological characteristics of patients in 12 cohorts

	TCGA	GSE13213	GSE30219	GSE37745	GSE42127	GSE50081	GSE31210	GSE68465	GSE72094	GSE61676	GSE135222	GSE126044	NCC	NCC-ICIs
N	492	117	85	106	133	127	246	443	442	43	27	16	203	30
Cancer Type	LUAD	LUAD	LUAD	LUAD	LUAD	LUAD	LUAD	LUAD	LUAD	NSCLC	NSCLC	NSCLC	LUAD	LUAD
Age[a]	66 (59, 72)	61 (55, 67)	60 (55, 69)	64 (55, 70)	66 (59, 74)	70 (63, 76)	61 (55, 65)	65 (58, 72)	70 (64, 76)	61 (54, 66)	62 (58, 68)	65 (55, 67)	60 (54, 67)	58 (51, 62)
Sex														
Female	266 (54%)	57 (49%)	19 (22%)	60 (57%)	65 (49%)	62 (49%)	130 (53%)	220 (50%)	240 (54%)	24 (56%)	5 (19%)	3 (17%)	116 (57%)	11 (37%)
Male	226 (46%)	60 (51%)	66 (78%)	46 (43%)	68 (51%)	65 (51%)	116 (47%)	223 (50%)	202 (46%)	19 (44%)	22 (81%)	15 (83%)	87 (43%)	19 (63%)
TNM Stage														
I	267 (54%)	79 (68%)	71 (84%)	70 (66%)	89 (67%)	92 (72%)	168 (74%)	114 (26%)	265 (64%)				87 (43%)	
II	119 (24%)	13 (11%)	13 (15%)	19 (18%)	22 (17%)	35 (28%)	58 (26%)	291 (66%)	69 (17%)				51 (25%)	
III	79 (16%)	25 (21%)	1 (1.2%)	13 (12%)	20 (15%)			38 (8.6%)	63 (15%)	4 (9%)			65 (32%)	4 (13%)
IV	26 (5.3%)			4 (3.8%)	1 (0.8%)				17 (4.1%)	39 (91%)				26 (87%)
Smoking														
Ever	409 (86%)	61 (52%)				92 (80%)	123 (50%)	300 (86%)	335 (91%)				65 (32%)	15 (50%)
Never	69 (14%)	56 (48%)				23 (20%)	123 (50%)	49 (14%)	33 (9.0%)				138 (68%)	15 (50%)
TP53														
MUT	182 (37%)	38 (33%)							111 (25%)		17 (63%)			11 (37%)
WT	310 (63%)	78 (67%)							331 (75%)		10 (37%)			19 (63%)
EGFR														
MUT	30 (6.1%)	45 (38%)					127 (52%)		47 (11%)		1 (3.7%)		127 (63%)	7 (23.3%)
WT	462 (94%)	72 (62%)					119 (48%)		395 (89%)		26 (96%)		75 (37%)	23 (76.7%)

(*Continued on next page*)

Table 1. ***Continued***

	TCGA	GSE13213	GSE30219	GSE37745	GSE42127	GSE50081	GSE31210	GSE68465	GSE72094	GSE61676	GSE135222	GSE126044	NCC	NCC-ICIs
KRAS														
MUT	121 (25%)	15 (13%)					20 (8.1%)		154 (35%)		0 (0%)		16 (7.9%)	2 (6.7%)
WT	371 (75%)	102 (87%)					226 (92%)		288 (65%)		27 (100%)		186 (92%)	28 (93.3%)
ALK														
MUT	16 (3.3%)						11 (4.5%)				1 (3.7%)		7 (3.5%)	1 (3.3%)
WT	476 (97%)						235 (96%)				26 (96%)		195 (97%)	29 (96.7%)
Adjuvant therapy														
NO													136 (67%)	1 (3.3%)
YES													67 (33%)	29 (96.7%)
TKI therapy														
NO													188 (93%)	
YES										43 (100%)			15 (7%)	
Platform	Illu.HiSeq V2	Agilent. 4x44K	Affy. Plus 2	Affy. Plus 2	Illu.WG-6 V3	Affy. Plus 2	Affy. Plus 2	Affy. U133A	Affy. 2.0	Affymetrix HuEx-1_0-st	Illu.HiSeq 2500	Illu.HiSeq 2500		
Ref.	TCGA, 2018	Tomida et al., 2009[24]	Rousseaux et al., 2013[25]	Botling et al., 2013[26]	Tang et al., 2013[27]	Der et al., 2014[28]	Okayama et al., 2012[29]	Shedden et al., 2008[30]	Schabath et al., 2016[31]	Baty et al., 2017[32]	Jung et al., 2019[33]	Cho et al., 2020[34]		

Abbreviations: LUAD, lung adenocarcinoma; NSCLC, non-small-cell lung cancer; TP53, Tumor Protein P53; MUT, mutation; WT, wild type; EGFR, epidermal growth factor receptor; KRAS, kirsten rat sarcoma viral oncogene; ALK, anaplastic lymphoma kinase; TKI, tyrosine kinase inhibitor; N/A = Not Applicable.

[a]Median (IQR); n (%).

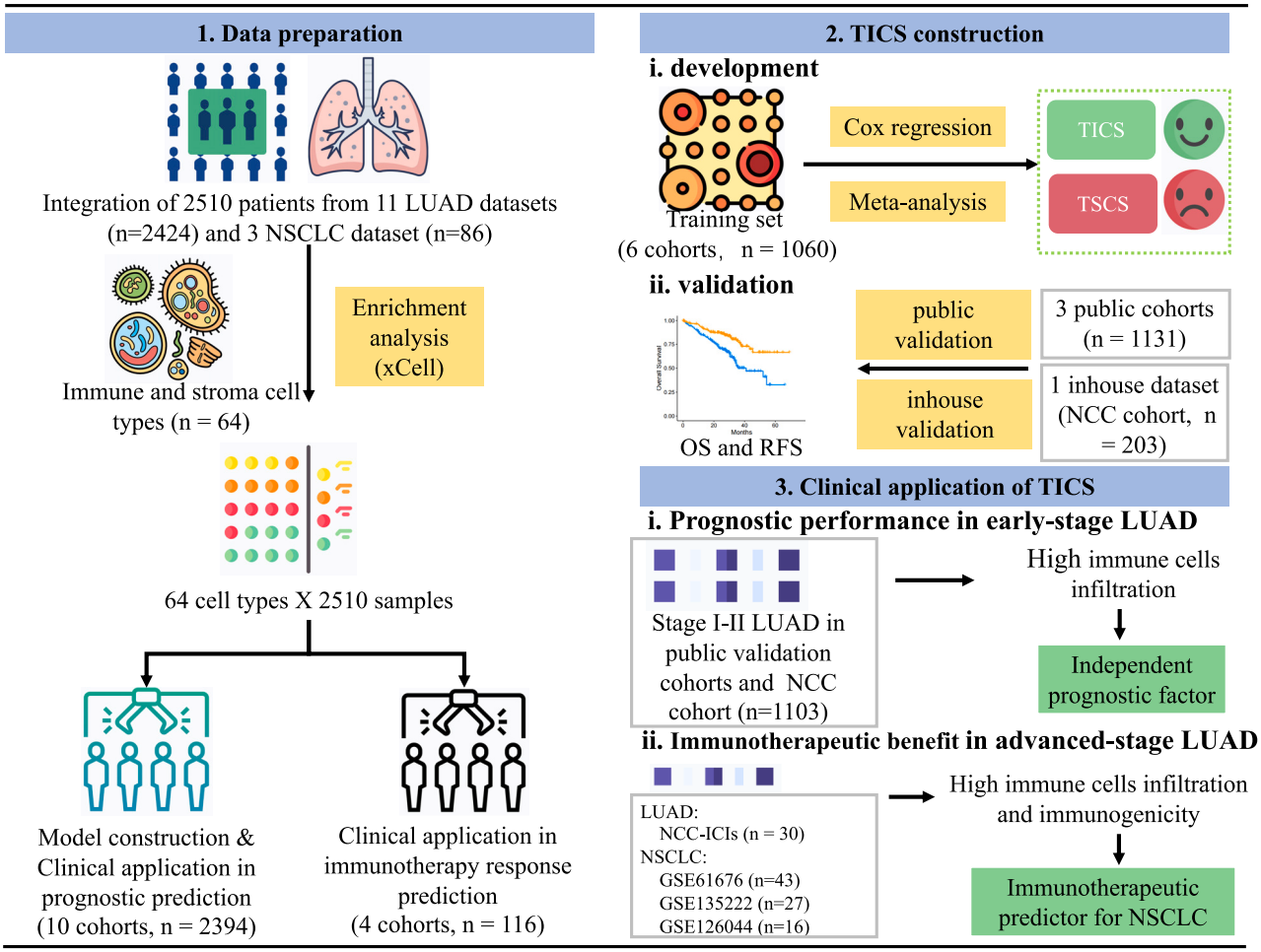


Figure 1. The overview of the development, validation, and clinical utility of TICS

Abbreviations: LUAD, Lung adenocarcinoma; NSCLC, non-small-cell lung cancer; HR, hazard ratio; TSCS, tumor stromal cell score; TICS, tumor immune cell score; OS, overall survival.

The median age was 60 (IQR, 54–67) years old and 87 patients (42.8%) were male. The median follow-up time was 65 · 3 months. Consistently, we observed patients with high TICS had a longer RFS (HR 0.72, 95% CI 0.47–1.10, p = 0.143, Figure S3) and OS (HR 0.42, 95% CI 0.21–0.84, p = 0.012; Figure 3F; Table S12)

The prognostic performance of tumor immune cell score in early- and advanced-stage lung adenocarcinoma

Since the anti-tumor immunity in the advanced stage of tumor is prone to compromising immunologic tolerance, tumor invasion, and metastasis, and confounded by various treatment regimens, we postulated that TICS would perform better in predicting prognosis in the early-stage LUAD than that in the advanced-stage LUAD. Meta analyses were conducted, and the results showed that patients with high TICS were associated with a longer OS in the early-stage LUAD in all datasets (6 training sets, 3 public validation sets, and NCC cohort), when compared with those with low TICS (Pool analyses, HR = 0.49, 95% CI: 0.38–0.63, p < 0.010, Figure 4A). However, the statistical differences of OS were not significant in 5 of 6 the datasets of advanced-stage LUAD stratified by TICS (Figure 4B). These results confirmed that the prognostic value of the TICS was more pivotal in early-stage LUAD.

To further confirm the independent prognostic performance of TICS in early-stage LUAD, univariable and multivariable Cox regression analyses were performed in an integrated dataset consisting of early-stage LUAD patients (n = 1975). The results revealed that TICS, age, and *EGFR* mutation were independent risk factors for

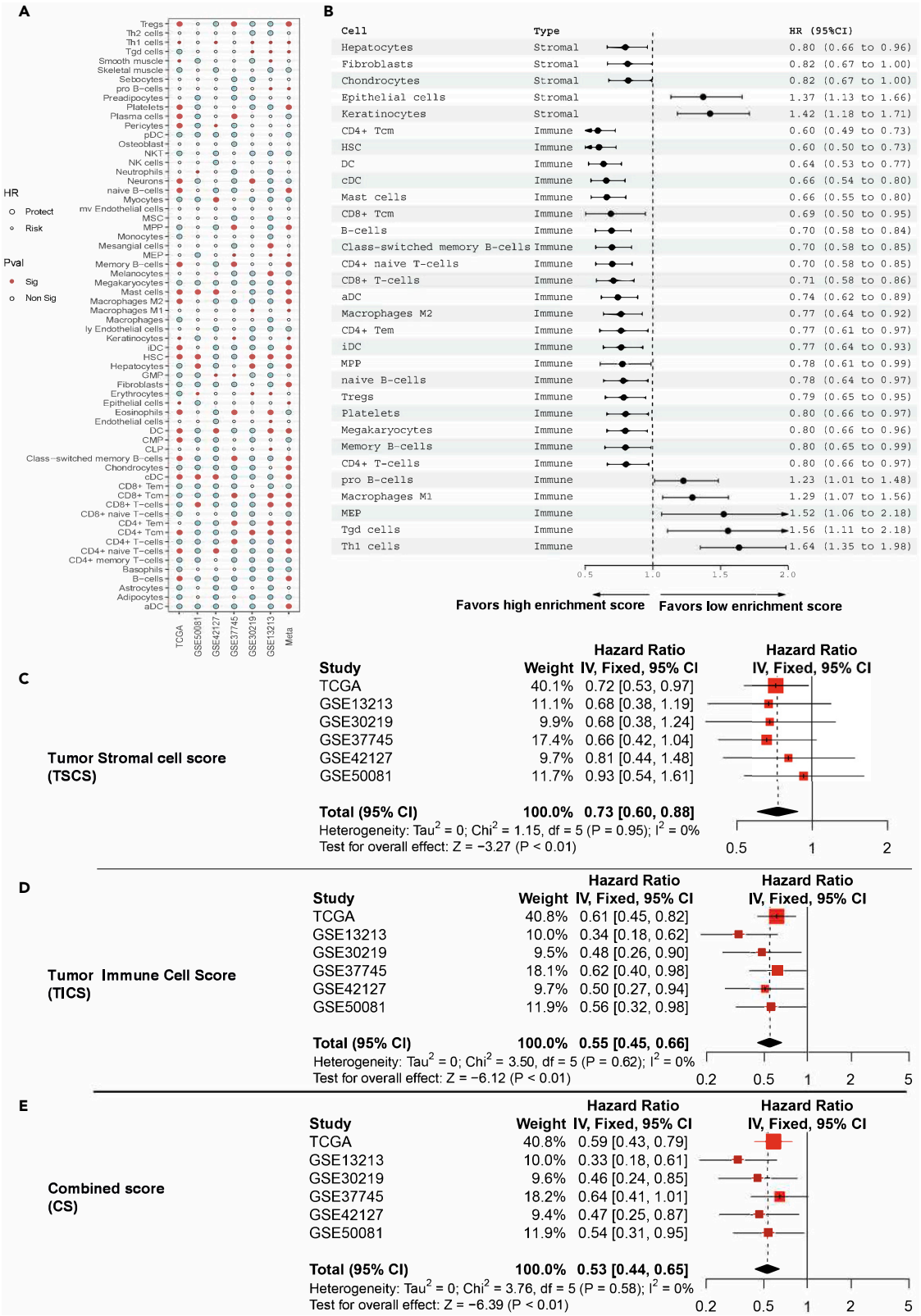

A
Tregs
Th2 cells
Th1 cells
Tgd cells
Smooth muscle
Skeletal muscle
Sebocytes
pro B−cells
Preadipocytes
Platelets
Plasma cells
Pericytes
pDC
Osteoblast
NKT
NK cells
Neutrophils
Neurons
naive B−cells
Myocytes
mv Endothelial cells
MSC
MPP
Monocytes
Mesangial cells
MEP
Memory B−cells
Melanocytes
Megakaryocytes
Mast cells
Macrophages M2
Macrophages M1
Macrophages
ly Endothelial cells
Keratinocytes
iDC
HSC
Hepatocytes
GMP
Fibroblasts
Erythrocytes
Epithelial cells
Eosinophils
Endothelial cells
DC
CMP
CLP
Class−switched memory B−cells
Chondrocytes
cDC
CD8+ Tem
CD8+ Tcm
CD8+ T−cells
CD8+ naive T−cells
CD4+ Tem
CD4+ Tcm
CD4+ T−cells
CD4+ naive T−cells
CD4+ memory T−cells
Basophils
B−cells
Astrocytes
Adipocytes
aDC
TCGA
GSE50081
GSE42127
GSE37745
GSE30219
GSE13213
Meta
HR
Protect
Risk
Pval
Sig
Non Sig
B
Cell
Type
HR (95%CI)
Hepatocytes Stromal 0.80 (0.66 to 0.96)
Fibroblasts Stromal 0.82 (0.67 to 1.00)
Chondrocytes Stromal 0.82 (0.67 to 1.00)
Epithelial cells Stromal 1.37 (1.13 to 1.66)
Keratinocytes Stromal 1.42 (1.18 to 1.71)
CD4+ Tcm Immune 0.60 (0.49 to 0.73)
HSC Immune 0.60 (0.50 to 0.73)
DC Immune 0.64 (0.53 to 0.77)
cDC Immune 0.66 (0.54 to 0.80)
Mast cells Immune 0.66 (0.55 to 0.80)
CD8+ Tcm Immune 0.69 (0.50 to 0.95)
B−cells Immune 0.70 (0.58 to 0.84)
Class−switched memory B−cells Immune 0.70 (0.58 to 0.85)
CD4+ naive T−cells Immune 0.70 (0.58 to 0.85)
CD8+ T−cells Immune 0.71 (0.58 to 0.86)
aDC Immune 0.74 (0.62 to 0.89)
Macrophages M2 Immune 0.77 (0.64 to 0.92)
CD4+ Tem Immune 0.77 (0.61 to 0.97)
iDC Immune 0.77 (0.64 to 0.93)
MPP Immune 0.78 (0.61 to 0.99)
naive B−cells Immune 0.78 (0.64 to 0.97)
Tregs Immune 0.79 (0.65 to 0.95)
Platelets Immune 0.80 (0.66 to 0.97)
Megakaryocytes Immune 0.80 (0.66 to 0.96)
Memory B−cells Immune 0.80 (0.65 to 0.99)
CD4+ T−cells Immune 0.80 (0.66 to 0.97)
pro B−cells Immune 1.23 (1.01 to 1.48)
Macrophages M1 Immune 1.29 (1.07 to 1.56)
MEP Immune 1.52 (1.06 to 2.18)
Tgd cells Immune 1.56 (1.11 to 2.18)
Th1 cells Immune 1.64 (1.35 to 1.98)
0.5
1.0
1.5
2.0
Favors high enrichment score
Favors low enrichment score
C
Tumor Stromal cell score (TSCS)
Study
Weight
Hazard Ratio IV, Fixed, 95% CI
Hazard Ratio IV, Fixed, 95% CI
TCGA 40.1% 0.72 [0.53, 0.97]
GSE13213 11.1% 0.68 [0.38, 1.19]
GSE30219 9.9% 0.68 [0.38, 1.24]
GSE37745 17.4% 0.66 [0.42, 1.04]
GSE42127 9.7% 0.81 [0.44, 1.48]
GSE50081 11.7% 0.93 [0.54, 1.61]
Total (95% CI) 100.0% 0.73 [0.60, 0.88]
Heterogeneity: Tau² = 0; Chi² = 1.15, df = 5 (P = 0.95); I² = 0%
Test for overall effect: Z = −3.27 (P < 0.01)
0.5
1
2
D
Tumor Immune Cell Score (TICS)
Study
Weight
Hazard Ratio IV, Fixed, 95% CI
Hazard Ratio IV, Fixed, 95% CI
TCGA 40.8% 0.61 [0.45, 0.82]
GSE13213 10.0% 0.34 [0.18, 0.62]
GSE30219 9.5% 0.48 [0.26, 0.90]
GSE37745 18.1% 0.62 [0.40, 0.98]
GSE42127 9.7% 0.50 [0.27, 0.94]
GSE50081 11.9% 0.56 [0.32, 0.98]
Total (95% CI) 100.0% 0.55 [0.45, 0.66]
Heterogeneity: Tau² = 0; Chi² = 3.50, df = 5 (P = 0.62); I² = 0%
Test for overall effect: Z = −6.12 (P < 0.01)
0.2
0.5
1
2
5
E
Combined score (CS)
Study
Weight
Hazard Ratio IV, Fixed, 95% CI
Hazard Ratio IV, Fixed, 95% CI
TCGA 40.8% 0.59 [0.43, 0.79]
GSE13213 10.0% 0.33 [0.18, 0.61]
GSE30219 9.6% 0.46 [0.24, 0.85]
GSE37745 18.2% 0.64 [0.41, 1.01]
GSE42127 9.4% 0.47 [0.25, 0.87]
GSE50081 11.9% 0.54 [0.31, 0.95]
Total (95% CI) 100.0% 0.53 [0.44, 0.65]
Heterogeneity: Tau² = 0; Chi² = 3.76, df = 5 (P = 0.58); I² = 0%
Test for overall effect: Z = −6.39 (P < 0.01)
0.2
0.5
1
2
5

Figure 2. Prognostic performance of immune cells and stromal cells in the training cohorts

(A) Bubble plot of the prognostic effect of 64 cell types across six LUAD datasets in the training cohorts. The size and color of the circles indicate the HR for OS and statistical significance (red) or non-significance (blue), respectively. Larger circles represent decreasing hazard for death and vice versa.

(B) The pooled HRs for each cell type in the meta-analysis in six LUAD datasets in the training cohorts.

(C–E) Forest plot illustrating the association between TSCS (C), TISC (D) and CS (E), and OS in the meta-analysis in the training cohorts. AbbreviationsLUAD, Lung adenocarcinoma; HR, hazard ratio; TSCS, tumor stromal cell score; TICS, tumor immune cell score; CS, combined score; OS, overall survival.

predicting better OS in early-stage LUAD patients (HR = 0.52, 95% CI: 0.39–0.69, $p < 0.001$, Table 2). Consistent with the results in the integrated dataset, the high-TICS lead to significantly longer OS compared with low-TICS in each cohort as shown in univariable or multivariable Cox regression analyses (Table 3). Taken together, these results demonstrated that the TICS can serve as an independent predictor of survival in early-stage LUAD.

The immunologic characteristics based on the tumor immune cell score

Next, we further analyzed the immunologic characteristics based on the TICS in early- and advanced-stage LUAD to investigate whether the difference in prognostic power of TICS was due to the difference in immune traits of TICS subgroups in the early and advance-stage LUAD. We used the NCC cohort and TCGA cohort to investigate the immunologic characteristics in different TICS subgroups due to its comprehensive information, including transcriptome, genetic mutations, and survival data. Gene Set Enrichment Analysis (GSEA) analysis was performed on the differentially expressed genes to explore the potential mechanisms of TICS in LUAD. The results demonstrated that the B cell/T cell receptor and Th17 cell differentiation were enriched in the high TICS group both in the early- and advanced-stage LUAD (FDR adjust $p < 0.010$, Figures 5A and 5B) in the NCC cohort, indicating anti-tumor immune signaling activation in the high TICS group. The similar signaling pathway enrichment results were further achieved in the GO analysis that the immune response pathway was significantly enriched in the high TICS group including T cell differentiation, positive regulation of immune response, inflammatory response, and activation of immune response (FDR adjusted $p < 0.050$) compared with the low TICS group (Figures 5C and 5D) in the NCC cohort. Consistent tendency was validated in the TCGA cohort, suggesting immune response activation in both early- and advanced stage LUAD (Figures S4A–S4D).

To demonstrate the association between the TICS and potential immune response, infiltrated immune cells in the high TICS and low TICS groups were compared in the NCC cohort and TCGA cohort, respectively. As a result, the immune cells infiltration and immune signatures were higher in the high TICS group compared with the low TICS group both in the early- and advanced-stage LUAD, that most immune killing cells including activated B cells, activated CD8 T cells, memory CD8 T cells, T helper cells were dramatically increased in the high TICS group compared with those in the low TICS group (FDR adjusted $p < 0.001$, Figures 5E and 5F). Similar results were also observed in the TCGA cohort (Figures S4E and S4F).

We further estimated the potential clinical response of immunotherapy with Tumor Immune Dysfunction and Exclusion (TIDE) algorithm, which has been reported with T cell functional inactivation and positively associated with immune infiltration, respectively. The high TICS group was characterized as lower TIDE scores, consisting of significantly higher TIDE dysfunction signatures and low TIDE exclusion scores, compared to the low TICS group both in the early- and advanced-stage LUAD (Mann-Whitney, $p < 0.001$. Figures 5G and 5H). The above results were also consistently validated in the TCGA cohort (Figures S4G and S4H). Moreover, somatic copy number alterations (SCNA)[35] and total mRNA expression (TmS)[36] were compared between the high and the low TICS group in the early- and advanced-stage LUAD (Figures S5A and S5B). Overall, the TIDE score, SCNA, and TmS were lower in the high TICS group, suggesting relatively reduced risk of cancer progression and potentially higher immunogenicity, compared with the low TICS group.

Tumor immune cell score may serve as an immunotherapeutic predictor for non-small-cell lung cancer (NSCLC)

Considering the consistent higher immunogenicity but inconsistent prognostic prediction associated with high TICS in early- and advanced-stage LUAD, we further investigated the potential mechanism. We first compared the cancer hallmarks (n = 41) between early-stage LUAD and advance-stage LUAD in the TCGA cohort, and multiple signaling pathways involved in cancer proliferation and metastasis were increased in the advance-stage LUAD compared with those in the early-stage LUAD, including DNA damage repair, Wnt signaling, Hedgehog signaling, mTOR signaling (Mann-Whitney, $p < 0 \cdot 050$) (Figure S6A). Additionally, the expression of antigen processing, IL-1 family signaling, and TGF-beta signaling was significantly increased in the advanced-stage LUAD (Figure S6B), indicating the potential immune tolerance.

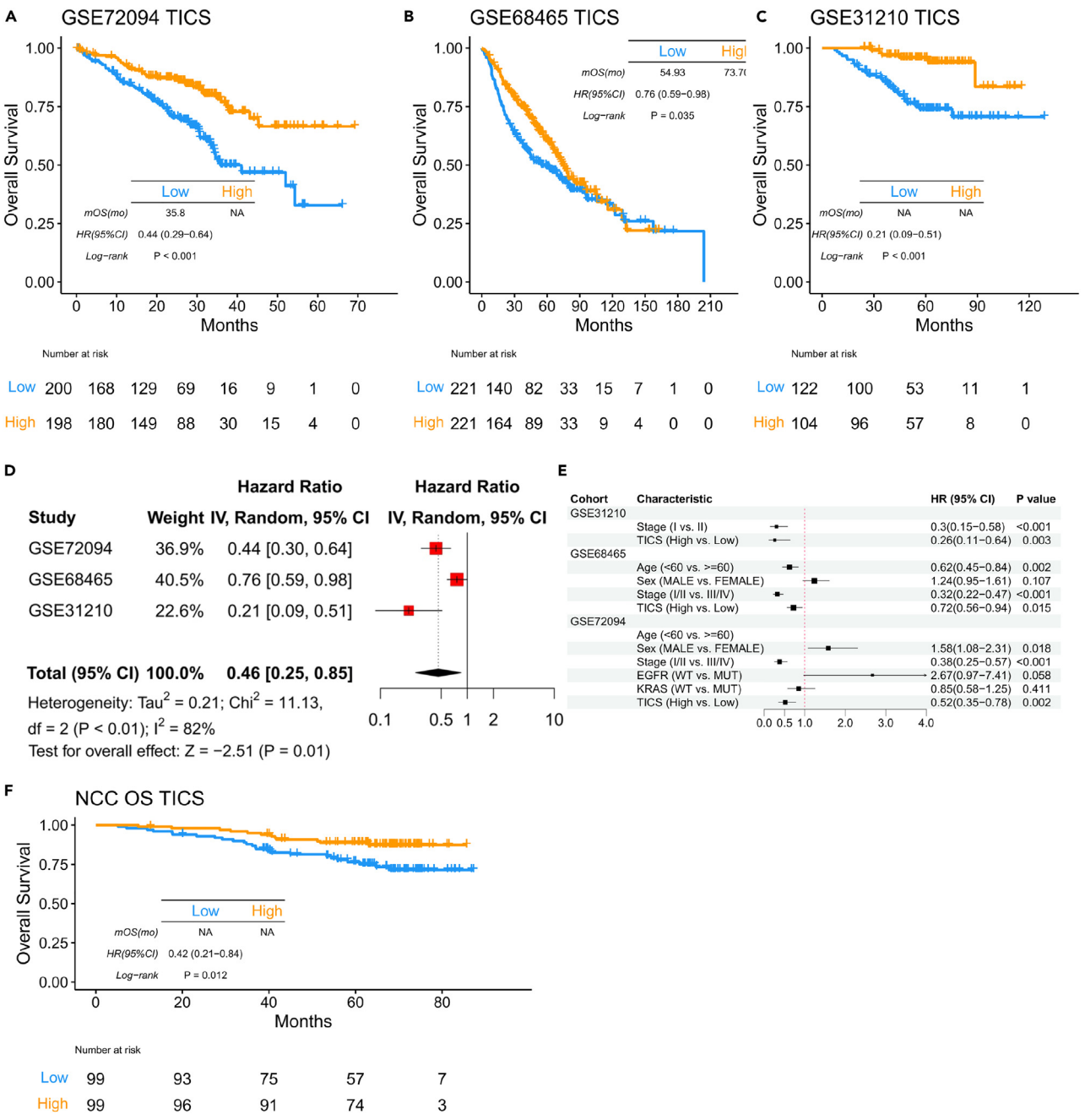


Figure 3. Prognostic performance of the TICS in the validation cohorts

(A–C) Kaplan-Meier survival curves comparing OS between the high- and low-TICS group in the three public validation cohorts, GSE72094 (A), GSE68465 (B), and GSE31210 (C) datasets.

(D) Forest plot illustrating the association between TISC and OS in meta-analysis in the public validation cohorts.

(E) Multivariable analyses of the TICS, age, sex, stage, and/or driver gene mutation in three public validation cohorts.

(F) Kaplan-Meier survival curves comparing OS between the high- and low-TICS group in the NCC cohorts. AbbreviationsTICS, tumor immune cell score; OS, overall survival; HR, hazard ratio; CI, confidence interval.

Next, higher immunogenicity and immune cell infiltrations associated with high TICS inspired our hypothesis into the predictive efficacy of TICS in the immunotherapeutic benefit in the advanced-stage LUAD. We further investigated the immunotherapeutic predictive efficacy of TICS in NCC-ICIs cohort. The NCC-ICIs cohort included 30 stage III-IV LUAD patients who were treated with anti-PD-(L)1 treatment and had

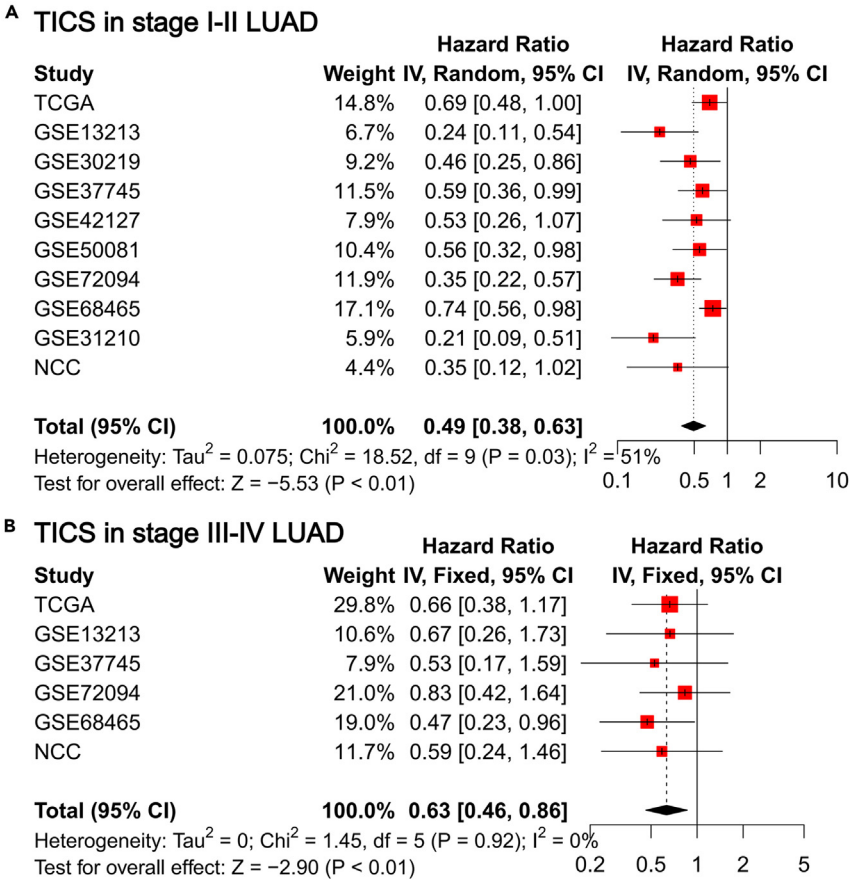


Figure 4. Performance of the TICS in predicting OS in stage I-II and stage III-IV LUAD

(A) Forest plot of the association between the TICS and OS in meta-analysis in stage I-II LUAD.

(B) Forest plot of the association between TICS and OS in meta-analysis in stage III-IV LUAD; AbbreviationsLUAD, Lung adenocarcinoma; TICS, tumor immune cell score; OS, overall survival; CI, confidence interval.

baseline tissue samples sequenced by WES and mRNA (Table 1). The median age was 58 (IQR, 51–62) years old and 19 patients (63.3%) were male. Consistently, patients in the high TICS group had a favorable PFS and OS (PFS, HR = 0.20; 95% CI, 0.04–1.00, p = 0.039; OS, HR = 0.06; 95% CI, 0.01–0.51, p = 0.001; Figures 6A and 6B), compared with those in the low TICS group. It was further validated in two public cohorts of patients with advanced-stage NSCLC (GSE135222 n = 27, and GSE126044 n = 16), that patients with high TICS showed longer PFS than those with low TICS (GSE135222, HR = 0.09; 95% CI, 0.02–0.32, p < 0.001; Figure 6C), and a significantly higher TICS score was observed in ICI-responders than in ICI-non-responders (GSE126044, Mann-whitney, p = 0.005, Figure S7A). In the multivariable Cox regression analysis adjusting by age, smoking, TMB, and driver gene mutation, high TICS remained an independent predictor of superior PFS (Figure 6D and Table 4), and OS (Table S13) in NSCLC. Additionally, there was no significant difference in TMB (Mann-Whitney, p = 0.366) and neoantigen (Mann-Whitney, p = 0.712) between high- and low-risk groups, suggesting the predictive efficacy was not confounded by TMB and neoantigen (Figures S7B–S7E). Additionally, no significant difference in driver gene mutations between high and low TICS groups (Figures 6E and S7F).

However, in the advanced-stage LUAD who received TKI or chemotherapy, high TICS was no longer associated with OS in the NCC cohort (chemotherapy: HR = 0.18; 95% CI, 0.02–1.50, p = 0.076; TKI: HR = 1.10; 95% CI, 0.55–2.20, p = 0.772; Figures S7G and S7H), and no significant difference in TICS score (Mann-Whitney, p = 0.857, Figure S7I) between ICI-responders- and ICI-non-responders, suggesting that the predictive efficacy of TICS may be compromised by different therapeutic regimens in the advanced-stage LUAD. Moreover, TICS showed potential as a predictor for immunotherapy over chemotherapy or other regimens.

Table 2. Univariable analysis and Multivariable Cox regression analyses of OS in early stage (I-II) LUAD in all cohorts

		Univariable Cox		Multivariable Cox	
Characteristic	Size	HR (95% CI)	p value	HR (95% CI)	p value
Age (<60 vs. >=60)	1966	0.66(0.55–0.78)	<0.001	0.7(0.5–0.96)	0.028
Sex (Male vs. Female)	1975	1.36(1.16–1.58)	<0.001	1.2(0.9–1.6)	0.211
Smoking (Never vs. Ever)	1540	0.46(0.36–0.58)	<0.001	0.66(0.44–0.98)	0.039
EGFR (WT vs. Mut)	1159	2.99(2.16–4.12)	<0.001	2.08(1.34–3.23)	0.001
KRAS (WT vs. Mut)	1159	0.6(0.45–0.78)	<0.001	0.9(0.65–1.23)	0.505
ALK (WT vs. Mut)	746	1.07(0.48–2.43)	0.862		
TP53 (WT vs. Mut)	932	0.76(0.59–0.99)	0.043	0.8(0.6–1.07)	0.131
Chemo treatment (No vs. Yes)	134	0.43(0.16–1.17)	0.098		
TICS (High vs. Low)	1975	0.55(0.47–0.64)	<0.001	0.52(0.39–0.69)	<0.001

Abbreviations: OS, overall survival; LUAD, Lung adenocarcinoma; HR, hazard ratio; CI, confidence interval; EGFR, epidermal growth factor receptor; MUT, mutation; WT, wild type; TP53, Tumor Protein P53; KRAS, kirsten rat sarcoma viral oncogene; ALK, anaplastic lymphoma kinase; TICS, tumor immune cell score.

DISCUSSION

In this study, we developed a prognostic and immunotherapeutic predictive model, TICS, based on the comprehensive cell-originated characterization of TME in 2,451 participants from 12 independent cohorts including 10 public datasets and 2 in-house cohorts, demonstrating the association between high- or low-TICS and prognosis, immune contexture and immunotherapeutic responsiveness in patients with LUAD, providing the insight into prognosis stratification and potential guidance for treatment strategy selection. To our knowledge, this study included the largest cohort to evaluate the value of TME in LUAD by differentiating cell origins.

Numerous studies have explored the association between immune cells and OS in LUAD. For example, high $CD4^+$ T cells in TME were associated with longer OS and disease-specific survival.[37] Moreover, a higher density of mast cells in TME was reported to be associated with improved survival in LUAD.[38] High B-cell and $CD8^+$ T cell infiltrations were associated with a favorable prognosis in LUAD.[39] Although infiltrated immune cells have been widely shown to play an important role in anti-tumor immune activity and be associated with a better prognosis, the effect of stromal cells on the prognosis of tumor remained insufficient.[7] Currently, CAFs have been studied in several tumors,[18] demonstrating the association with tumor growth, invasion and poor prognosis.[7] However, other stromal cells are scarcely studied regarding prognosis in LUAD. Furthermore, previous studies based on bulk transcriptomics seldomly discriminated the stromal or immune cell-derived signatures, so the robustness and reproducibility of previous transcriptomics-based models would be influenced by tumor heterogeneity and immune cell infiltration status. Although single-cell sequencing can distinguish different cell origins, its exorbitant price and demanding technique impeded the application in clinical practice. Therefore, a robust, stable, and relatively cost-effective model was urgently warranted.

We identified 64 immune and stromal cell types based on the transcriptomic data of LUAD, demonstrating the association with OS of 26 immune cells and 5 stromal cells. Further, we constructed TICS, TSCS, and CS based on immune cells, stromal cells, and the resultant to evaluate the prognostic stratification in LUAD, respectively. We demonstrated that the TICS, instead of TSCS, were the major contributor for predicting superior clinical outcomes.

Noteworthily, TICS performed better in predicting prognosis in the stage I-II LUAD compared with the stage III-IV LUAD, however, the immune-related pathways, including B cell/T cell receptor and Th17 cell differentiation, in the high TICS group were increased compared with the low TICS group no matter in the early-stage or in the advanced-stage LUAD. Considering the consistent higher immunogenicity but inconsistent prognostic prediction associated with high TICS in early- and advanced-stage LUAD, the potential mechanisms were further investigated. During the early stage of cancer development, immune killing cells are the key players in immune defense against tumors.[40] Whereas at advanced stage, these tumors subsequently

Table 3. Univariable analysis and Multivariable Cox regression analyses of OS in early-stage (I-II) LUAD in separate cohort

			Univariate Cox		Multivariate Cox	
Cohort	Characteristic	Size	HR (95% CI)	p value	HR (95% CI)	p value
GSE30219						
	Age(<60 vs. ≥60)	92	0.72(0.33–1.57)	0.409		
	Sex(MALE vs. FEMALE)	92	1.86(0.89–3.87)	0.100		
	Smoking(Never vs. Ever)	92	0.73(0.36–1.48)	0.378		
	EGFR(WT vs. MUT)	92	1.24(0.59–2.59)	0.567		
	KRAS(WT vs. MUT)	92	0.56(0.21–1.45)	0.230		
	TP53(WT vs. MUT)	91	0.62(0.3–1.26)	0.187		
	TICS(High vs. Low)	92	0.24(0.11–0.54)	0.001		
GSE30219						
	Age(<60 vs. ≥60)	84	0.75(0.41–1.39)	0.360		
	Sex(MALE vs. FEMALE)	84	1.05(0.51–2.19)	0.887		
	TICS(High vs. Low)	84	0.46(0.25–0.86)	0.015		
GSE37745						
	Age(<60 vs. ≥60)	89	0.73(0.44–1.22)	0.229		
	Sex(MALE vs. FEMALE)	89	1.16(0.70–1.92)	0.557		
	TICS(High vs. Low)	89	0.59(0.36–0.99)	0.045		
GSE42127						
	Age(<60 vs. ≥60)	111	0.55(0.24–1.26)	0.157		
	Sex(MALE vs. FEMALE)	111	1.75(0.87–3.52)	0.117		
	TICS(High vs. Low)	111	0.52(0.26–1.07)	0.075		
GSE50081						
	Age(<60 vs. ≥60)	127	0.68(0.29–1.6)	0.376		
	Sex(MALE vs. FEMALE)	127	1.41(0.81–2.46)	0.228		
	Smoking(Never vs. Ever)	115	0.59(0.26–1.33)	0.207		
	TICS(High vs. Low)	127	0.56(0.32–0.98)	0.042		
TCGA						
	Age(<60 vs. ≥60)	377	0.82(0.53–1.26)	0.360		
	Sex(MALE vs. FEMALE)	386	1.04(0.73–1.5)	0.816		
	Smoking(Never vs. Ever)	376	1.12(0.67–1.88)	0.673		
	EGFR(WT vs. MUT)	386	0.41(0.22–0.74)	0.003	0.37(0.2–0.68)	0.001
	KRAS(WT vs. MUT)	386	1.13(0.72–1.77)	0.594		
	ALK(WT vs. MUT)	386	1.64(0.52–5.16)	0.399		
	TP53(WT vs. MUT)	386	0.79(0.54–1.15)	0.216		
	TICS(High vs. Low)	386	0.69(0.48–1)	0.047	0.65(0.45–0.95)	0.024
GSE31210						
	Age(<60 vs. ≥60)	226	0.72(0.36–1.43)	0.346		
	Sex(MALE vs. FEMALE)	226	1.52(0.78–2.96)	0.219		
	Smoking(Never vs. Ever)	226	0.61(0.31–1.19)	0.150		
	EGFR(MUT vs. WT)	226	0.47(0.24–0.93)	0.030	0.56(0.28–1.11)	0.097
	KRAS(MUT vs. WT)	226	0.87(0.26–2.85)	0.817		
	ALK(MUT vs. WT)	226	1.49(0.36–6.24)	0.582		
	TICS(High vs. Low)	226	0.21(0.09–0.51)	0.001	0.23(0.10–0.56)	0.001

(*Continued on next page*)

Table 3. *Continued*

Cohort	Characteristic	Size	Univariate Cox		Multivariate Cox	
			HR (95% CI)	p value	HR (95% CI)	p value
GSE68465						
	Age(<60 vs. ≥60)	405	0.64(0.46–0.88)	0.007	0.66(0.47–0.91)	0.011
	Sex(MALE vs. FEMALE)	405	1.36(1.03–1.79)	0.031	1.27(0.96–1.68)	0.097
	Smoking(Never vs. Ever)	326	0.87(0.54–1.39)	0.565		
	TICS(High vs. Low)	405	0.74(0.56–0.98)	0.033	0.76(0.58–1.01)	0.061
GSE72094						
	Age(<60 vs. ≥60)	321	0.69(0.35–1.39)	0.304		
	Sex(MALE vs. FEMALE)	321	1.48(0.94–2.31)	0.089		
	Smoking(Never vs. Ever)	271	0.72(0.26–1.99)	0.527		
	EGFR(WT vs. MUT)	321	11.3(1.57–81.3)	0.016	7.81(1.07–57.00)	0.043
	KRAS(WT vs. MUT)	321	0.54(0.35–0.85)	0.007	0.76(0.48–1.20)	0.233
	TP53(WT vs. MUT)	321	0.68(0.42–1.11)	0.121		
	TICS(High vs. Low)	321	0.35(0.22–0.58)	0.000	0.42(0.25–0.69)	0.001
NCC						
	Age(<60 vs. ≥60)	134	1.36(0.51–3.61)	0.543		
	Sex(Male vs. Female)	134	2.66(0.96–7.31)	0.059		
	Smoking(Never vs. Ever)	134	0.24(0.09–0.65)	0.005	0.26(0.09–0.72)	0.010
	EGFR(Mut vs. WT)	134	0.38(0.14–1.03)	0.058		
	KRAS(Mut vs. WT)	134	1.46(0.33–6.44)	0.616		
	ALK(Mut vs. WT)	134	1.93(0.26–14.6)	0.524		
	Visceral_pleural_invasion(No vs. Yes)	134	1.33(0.48–3.66)	0.579		
	Lymphvascular_invasion(No vs. Yes)	134	2.07(0.27–15.7)	0.481		
	Chemo_treatment(No vs. Yes)	134	0.43(0.16–1.17)	0.098		
	TKI_treatment(No vs. Yes)	134	0.47(0.11–2.08)	0.320		
	TICS(High vs. Low)	134	0.35(0.12–1.02)	0.054	0.40(0.14–1.17)	0.090

Abbreviations: OS, overall survival; TCGA, The Cancer Genome Atlas; NCC, National Cancer Center; HR, hazard ratio; CI, confidence interval; TICS, tumor immune cell score; EGFR, epidermal growth factor receptor; MUT, mutation; WT, wild type; TP53, Tumor Protein P53; KRAS, kirsten rat sarcoma viral oncogene; ALK, anaplastic lymphoma kinase; TKI, tyrosine kinase inhibitor.

evolve neutrally, thereby maximizing intratumoral heterogeneity and increasing the probability of immune tolerance.[41] The anticancer ability of immune cells may be gradually compromised by genomic instability and evolutionary tumor growth. Based on the above-mentioned hypothesis, we compared the cancer hallmarks between early-stage and advance-stage LUAD. Consistent with the preconception, in our study, multiple signaling pathways involved in cancer proliferation and metastasis, as well as immune tolerance, were increased in the advance-stage LUAD. Further, treatment-naïve patients in the high TICS group had a lower TIDE score, SCNA, and TmS levels, indicating weaker immune escape and improved response to ICB treatment. However, diversified treatment strategies are applied in patients with advanced NSCLC, such as TKI, immunotherapy, chemotherapy, and anti-angiogenesis therapy, and so forth,[42] attributing to the possible confounded predictive ability of TICS caused by various treatment regimens. Nevertheless, we still assumed that advanced-stage patients with high TICS might be susceptible to immunotherapy-based regimens, due to the association between higher TICS and higher immunity in advanced-stage LUAD.

As expected, in our study, patients in the high TICS group who received ICIs had a favorable PFS and durable clinical benefit compared with those in the low TICS group. Moreover, the predictive potency of TICS was not confounded by clinical covariates, driver mutations, and TMB, showing its potential to be an independent predictor for better clinical outcomes of ICIs in NSCLC. Moreover, TICS could not predict the survival benefit in patients receiving chemotherapy or TKI treatment, suggesting a specific immunotherapeutic predictive potency of TICS in LUAD. However, there was a relatively small size of

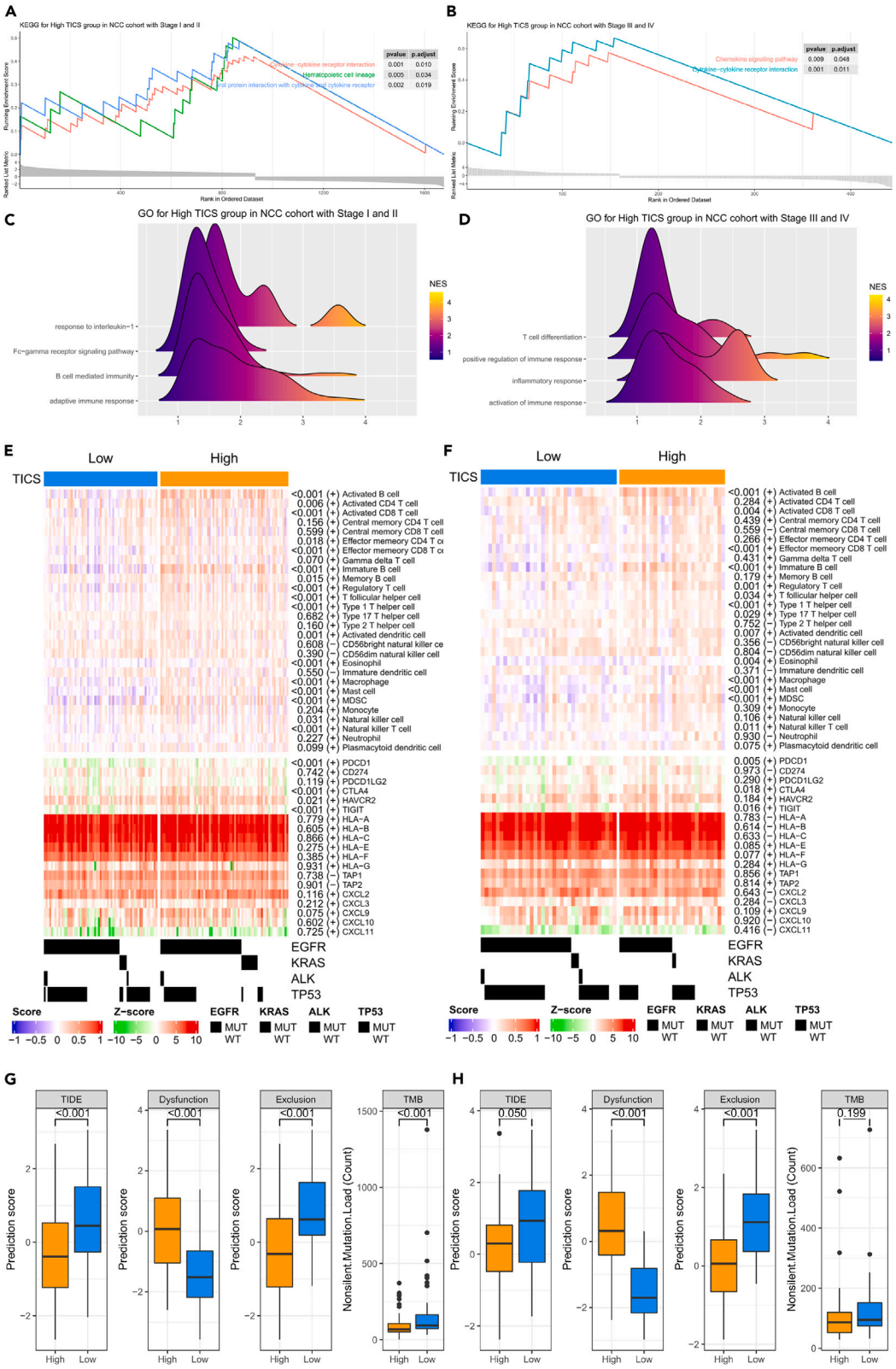

A
KEGG for High TICS group in NCC cohort with Stage I and II
pvalue p.adjust
0.001 0.010
0.005 0.034
0.002 0.019
Cytokine−cytokine receptor interaction
Hematopoietic cell lineage
Cell adhesion molecules
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset
B
KEGG for High TICS group in NCC cohort with Stage III and IV
pvalue p.adjust
0.009 0.048
0.001 0.011
Chemokine signaling pathway
Cytokine−cytokine receptor interaction
C
GO for High TICS group in NCC cohort with Stage I and II
response to interleukin−1
Fc−gamma receptor signaling pathway
B cell mediated immunity
adaptive immune response
NES
D
GO for High TICS group in NCC cohort with Stage III and IV
T cell differentiation
positive regulation of immune response
inflammatory response
activation of immune response
NES
E
Low
High
TICS
<0.001 (+) Activated B cell
0.006 (+) Activated CD4 T cell
<0.001 (+) Activated CD8 T cell
0.156 (+) Central memory CD4 T cell
0.599 (+) Central memory CD8 T cell
0.018 (+) Effector memeory CD4 T ce
<0.001 (+) Effector memeory CD8 T ce
0.070 (+) Gamma delta T cell
<0.001 (+) Immature B cell
0.015 (+) Memory B cell
<0.001 (+) Regulatory T cell
<0.001 (+) T follicular helper cell
<0.001 (+) Type 1 T helper cell
0.682 (+) Type 17 T helper cell
0.160 (+) Type 2 T helper cell
0.001 (+) Activated dendritic cell
0.608 (−) CD56bright natural killer ce
0.390 (−) CD56dim natural killer cell
<0.001 (+) Eosinophil
0.550 (−) Immature dendritic cell
<0.001 (+) Macrophage
<0.001 (+) Mast cell
<0.001 (+) MDSC
0.204 (+) Monocyte
0.031 (+) Natural killer cell
<0.001 (+) Natural killer T cell
0.227 (+) Neutrophil
0.099 (+) Plasmacytoid dendritic cell
<0.001 (+) PDCD1
0.742 (+) CD274
0.119 (+) PDCD1LG2
<0.001 (+) CTLA4
0.021 (+) HAVCR2
<0.001 (+) TIGIT
0.779 (+) HLA−A
0.605 (+) HLA−B
0.866 (+) HLA−C
0.275 (+) HLA−E
0.385 (+) HLA−F
0.931 (+) HLA−G
0.738 (−) TAP1
0.901 (−) TAP2
0.116 (+) CXCL2
0.212 (+) CXCL3
0.075 (+) CXCL9
0.602 (+) CXCL10
0.725 (+) CXCL11
EGFR
KRAS
ALK
TP53
Score
Z−score
EGFR KRAS ALK TP53
MUT
WT
F
Low
High
TICS
<0.001 (+) Activated B cell
0.284 (+) Activated CD4 T cell
0.004 (+) Activated CD8 T cell
0.439 (+) Central memory CD4 T cell
0.559 (−) Central memory CD8 T cell
0.266 (+) Effector memeory CD4 T cell
<0.001 (+) Effector memeory CD8 T cell
0.431 (+) Gamma delta T cell
<0.001 (+) Immature B cell
0.179 (+) Memory B cell
0.001 (+) Regulatory T cell
0.034 (+) T follicular helper cell
<0.001 (+) Type 1 T helper cell
0.029 (+) Type 17 T helper cell
0.752 (−) Type 2 T helper cell
0.007 (+) Activated dendritic cell
0.356 (−) CD56bright natural killer cell
0.804 (−) CD56dim natural killer cell
0.004 (+) Eosinophil
0.371 (−) Immature dendritic cell
<0.001 (+) Macrophage
<0.001 (+) Mast cell
<0.001 (+) MDSC
0.309 (+) Monocyte
0.106 (+) Natural killer cell
0.011 (+) Natural killer T cell
0.930 (−) Neutrophil
0.075 (+) Plasmacytoid dendritic cell
0.005 (+) PDCD1
0.973 (−) CD274
0.290 (+) PDCD1LG2
0.018 (+) CTLA4
0.184 (+) HAVCR2
0.016 (+) TIGIT
0.783 (−) HLA−A
0.614 (−) HLA−B
0.633 (−) HLA−C
0.085 (+) HLA−E
0.077 (+) HLA−F
0.284 (+) HLA−G
0.856 (+) TAP1
0.814 (+) TAP2
0.643 (−) CXCL2
0.284 (−) CXCL3
0.109 (+) CXCL9
0.920 (−) CXCL10
0.416 (−) CXCL11
EGFR
KRAS
ALK
TP53
Score
Z−score
EGFR KRAS ALK TP53
MUT
WT
G
TIDE
Dysfunction
Exclusion
TMB
<0.001
Prediction score
Nonsilent.Mutation.Load (Count)
High
Low
H
TIDE
0.050
Dysfunction
<0.001
Exclusion
<0.001
TMB
0.199
Prediction score
Nonsilent.Mutation.Load (Count)
High
Low

Figure 5. Immune characteristics in the high- and low-TICS group in the NCC cohort

(A and B) GSEA analysis illustrating the enrichment of immune signaling in the high-TICS groups in the early- (A) and advanced-stage (B) LUAD in the NCC cohort.

(C and D) Gene Ontology analysis illustrating the NES of immune and tumor signaling in the high-TICS groups in the early- (C) and advanced-stage (D) LUAD in the NCC cohort.

(E and F) Heatmap depicting the different infiltrated immune cell and immune-related genes in the high- and low-TICS groups in the early- (E) and advanced-stage (F) LUAD in the NCC cohort.

(G and H) The boxplot of TIDE, dysfunction score and exclusion sore, and TMB in the high- and low-TICS groups in the early- (G) and advanced-stage (H) LUAD in the NCC cohort. AbbreviationsLUAD, Lung adenocarcinoma; TICS, tumor immune cell score; GSEA, Gene Set Enrichment Analysis; TIDE, tumor immune dysfunction and exclusion; TMB, tumor mutation burden; NES, normalized enrichment score; OS, overall survival.

patients with advanced-stage LUAD, and the predictive performance on treatment-choices warrants further validation in future randomized controlled trails.

To sum up, the model we established surpassed the majority of previous models by distinguishing the contribution of stromal cells and immune cells in TME to get rid of the influence of tumor purity and micro-environmental heterogeneity. Moreover, we included multiple independent datasets as meta-cohorts for training, and four independent datasets, including in-house cohorts, for validation, ensuring the robustness of the model. Furthermore, we deeply explored the different prognosis and the immune contexture stratified by our model in different clinical stages, providing clues for future multiple application scenarios in clinical practice.

In conclusion, we constructed a prognostic model, the tumor immune cell score (TICS), based on the tumor-infiltrating immune cell signature in a large scale of patients in multi-cohorts, revealing that TICS is associated with survival in patients with LUAD especially in early stage, and may serve as a specific predictor for the benefit of ICIs in advanced LUAD.

Limitations of the study

As for limitations, the retrospective setting and pooled-estimate methodology of this study might introduce multiple biases. The limitation of the retrospective setting can be greatly minimized by the large sample size, by which the experimental features might be balanced, such as race, stage, and the platform of mRNA testing, and so forth. Notably, there was a relatively small size of patients with advanced-stage LUAD, and the predictive performance on treatment choices warrants further validation in future randomized controlled trails.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patient recruitment & sample acquisition
- METHOD DETAILS
  - Public lung datasets
  - RNA sequencing
  - Whole exome sequencing
  - Public data acquisition and pre-processing
  - Estimates of cells enrichment scores
  - Prognosis-related markers selection and signature construction
  - Evaluation and characteristics identification of TICS
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106616.

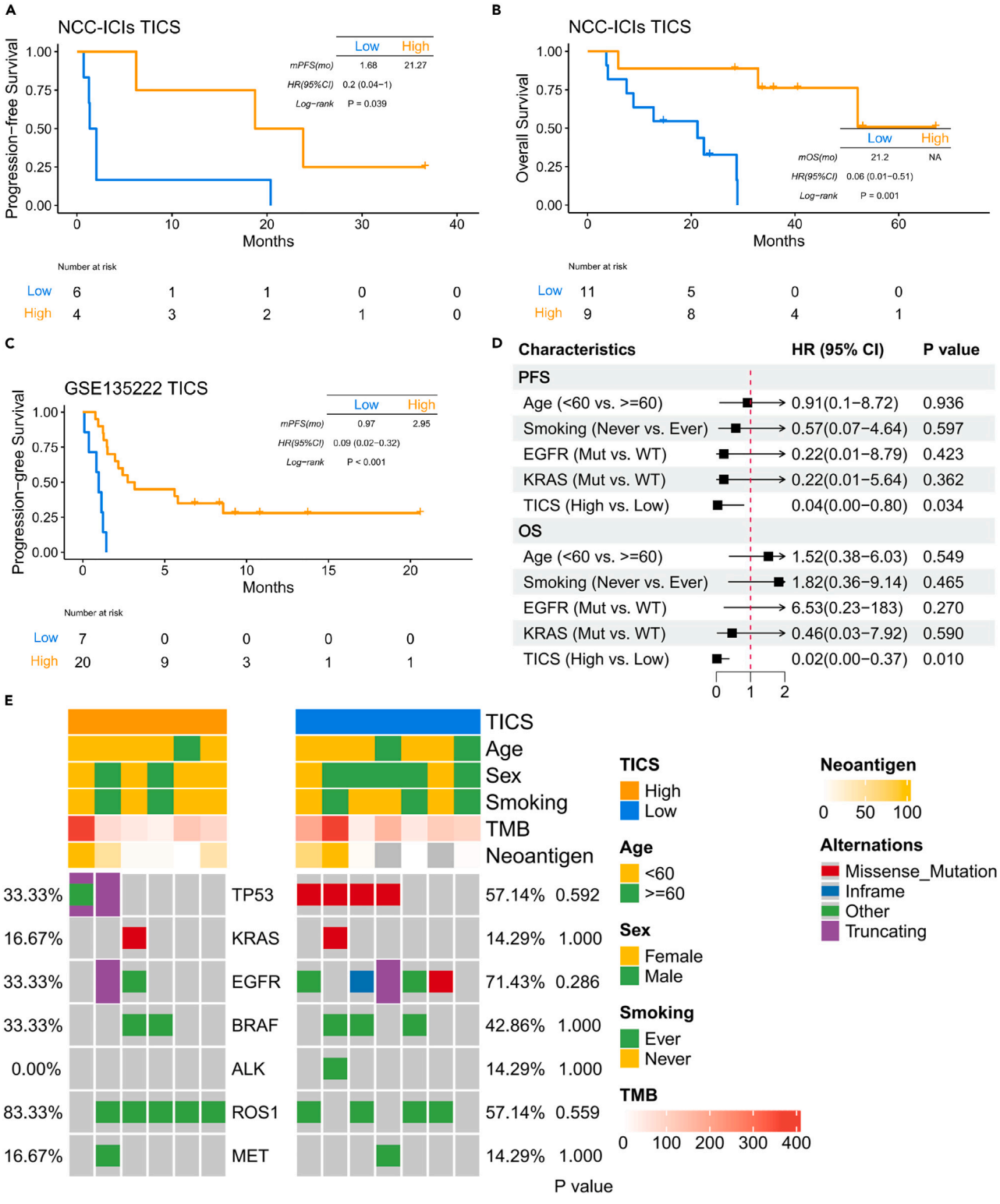


Figure 6. The performance of TICS in predicting the immunotherapeutic efficacy in LUAD

(A–C) Kaplan-Meier survival curves comparing PFS (A, NCC-ICI cohort), OS (B, NCC-ICI cohort), and PFS (C, GSE135222) between the high- and low-TISC groups in the patients with LUAD who received ICIs treatments.

(D) Multivariable analyses of the TICS, age, smoking, and driver gene mutation in the NCC-ICIs cohort.

(E) OncoPrint displaying the mutation spectrum, clinical characteristics in the high- and low-TISC groups in the NCC-ICIs cohort. Abbreviations LUAD, Lung adenocarcinoma; TICS, tumor immune cell score; PFS, progression-free survival; OS, overall survival; HR, hazard ratio; CI, confidence interval.

Table 4. Multivariable Cox regression analyses of PFS and OS in NCC-ICIs cohort

		Multivariable Cox	
Outcomes	Characteristic	HR (95% CI)	p value
PFS			
	Age (<60 vs. ≥60)	0.91(0.1–8.72)	0.936
	Smoking (Never vs. Ever)	0.57(0.07–4.64)	0.597
	TP53 (Mut vs. WT)	0.57(0.04–8.3)	0.680
	EGFR (Mut vs. WT)	0.22(0.01–8.79)	0.423
	KRAS (Mut vs. WT)	0.22(0.01–5.64)	0.362
	TICS (High vs. Low)	0.04(0–0.8)	0.034
OS			
	Age (<60 vs. ≥60)	1.52(0.38–6.03)	0.549
	Smoking (Never vs. Ever)	1.82(0.36–9.14)	0.465
	TP53 (Mut vs. WT)	0.13(0.01–2.43)	0.172
	EGFR (Mut vs. WT)	6.53(0.23–183)	0.270
	KRAS (Mut vs. WT)	0.46(0.03–7.92)	0.590
	TICS (High vs. Low)	0.02(0–0.37)	0.010

Abbreviations: PFS, progression-free survival; OS, overall survival; NCC, National Cancer Center; ICI, immune checkpoint inhibitor; HR, hazard ratio; CI, confidence interval; TICS, tumor immune cell score; EGFR, epidermal growth factor receptor; MUT, mutation; WT, wild type; TP53, Tumor Protein P53; KRAS, kirsten rat sarcoma viral oncogene.

ACKNOWLEDGMENTS

This work was supported by the National Natural Sciences Foundation (82102886 for JX); Beijing Nova Program (20220484119 for JX, Z211100002121055 for ZY); the Special Research Fund for Central Universities, Peking Union Medical College (3332021029 for JX); National Natural Sciences Foundation (82203827 for ZY); the Special Research Fund for Central Universities, Peking Union Medical College (3332021028 for ZY); Beijing Natural Science Foundation (7222146 for ZY); Guangdong Association of Clinical Trials (GACT)/Chinese Thoracic Oncology Group (CTONG) and Guangdong Provincial Key Lab of Translational Medicine in Lung Cancer (Grant No. 2017B030314120); Beijing Hope Run Special Fund of Cancer Foundation of China (LC2021B22, LC2020B09, LC2020A33); the National Natural Sciences Foundation (81871889 and 82072586 for ZW) and Beijing Natural Science Foundation (7212084 for ZW).

AUTHOR CONTRIBUTIONS

JX: conceptualization, data curation, project administration, writing - review & editing, and funding acquisition. ZY: conceptualization, data curation, methodology, and writing - review & editing. WX: conceptualization, formal analysis, methodology, and roles/writing - original draft. RW: data curation and roles/writing - original draft. CL: formal analysis and roles/writing - original draft. KF: data curation and roles/writing - original draft. BS: data curation and roles/writing - original draft. XY: data curation and roles/writing - original draft. PC: data curation and roles/writing - original draft. FM: formal analysis and roles/writing - original draft. GW: formal analysis, and roles/writing - original draft. JZ: formal analysis and roles/writing - original draft. YH: formal analysis and roles/writing - original draft. SC: supervision. JW: supervision. ZW: conceptualization, data curation, project administration, funding acquisition, and supervision. All authors read and approved the final article.

DECLARATION OF INTERESTS

WX, CL, FM, GW, JZ, and SC are employees of Burning Rock Biotech. YH is a founder of Burning Rock Biotech. The other authors declare that no competing interests.

Received: January 6, 2023
Revised: February 24, 2023
Accepted: April 4, 2023
Published: April 8, 2023

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J. Clin. *71*, 209–249.

2. Thai, A.A., Solomon, B.J., Sequist, L.V., Gainor, J.F., and Heist, R.S. (2021). Lung cancer. Lancet *398*, 535–554.

3. Chansky, K., Detterbeck, F.C., Nicholson, A.G., Rusch, V.W., Vallières, E., Groome, P., Kennedy, C., Krasnik, M., Peake, M., Shemanski, L., et al. (2017). 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. *12*, 1109–1121.

4. Hua, X., Zhao, W., Pesatori, A.C., Consonni, D., Caporaso, N.E., Zhang, T., Zhu, B., Wang, M., Jones, K., Hicks, B., et al. (2020). Genetic and epigenetic intratumor heterogeneity impacts prognosis of lung adenocarcinoma. Nat. Commun. *11*, 2459.

5. Hussaini, S., Chehade, R., Boldt, R.G., Raphael, J., Blanchette, P., Maleki Vareki, S., and Fernandes, R. (2021). Association between immune-related side effects and efficacy and benefit of immune checkpoint inhibitors - a systematic review and meta-analysis. Cancer Treat Rev. *92*, 102134.

6. Barnes, T.A., and Amir, E. (2017). HYPE or HOPE: the prognostic value of infiltrating immune cells in cancer. Br. J. Cancer *117*, 451–460.

7. Quail, D.F., and Joyce, J.A. (2013). Microenvironmental regulation of tumor progression and metastasis. Nat. Med. *19*, 1423–1437.

8. Bense, R.D., Sotiriou, C., Piccart-Gebhart, M.J., Haanen, J., van Vugt, M., de Vries, E.G.E., Schröder, C.P., and Fehrmann, R.S.N. (2017). Relevance of tumor-infiltrating immune cell composition and functionality for disease outcome in breast cancer. J. Natl. Cancer Inst. *109*, djw192.

9. Fridman, W.H., Zitvogel, L., Sautès-Fridman, C., and Kroemer, G. (2017). The immune contexture in cancer prognosis and treatment. Nat. Rev. Clin. Oncol. *14*, 717–734.

10. Hanahan, D., and Coussens, L.M. (2012). Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell *21*, 309–322.

11. Turley, S.J., Cremasco, V., and Astarita, J.L. (2015). Immunological hallmarks of stromal cells in the tumour microenvironment. Nat. Rev. Immunol. *15*, 669–682.

12. Hernández-Prieto, S., Romera, A., Ferrer, M., Subiza, J.L., López-Asenjo, J.A., Jarabo, J.R., Gómez, A.M., Molina, E.M., Puente, J., González-Larriba, J.L., et al. (2015). A 50-gene signature is a novel scoring system for tumor-infiltrating immune cells with strong correlation with clinical outcome of stage I/II non-small cell lung cancer. Clin. Transl. Oncol. *17*, 330–338.

13. Hiraoka, K., Miyamoto, M., Cho, Y., Suzuoki, M., Oshikiri, T., Nakakubo, Y., Itoh, T., Ohbuchi, T., Kondo, S., and Katoh, H. (2006). Concurrent infiltration by CD8+ T cells and CD4+ T cells is a favourable prognostic factor in non-small-cell lung carcinoma. Br. J. Cancer *94*, 275–280.

14. Kawai, O., Ishii, G., Kubota, K., Murata, Y., Naito, Y., Mizuno, T., Aokage, K., Saijo, N., Nishiwaki, Y., Gemma, A., et al. (2008). Predominant infiltration of macrophages and CD8(+) T Cells in cancer nests is a significant predictor of survival in stage IV nonsmall cell lung cancer. Cancer *113*, 1387–1395.

15. Stankovic, B., Bjørhovde, H.A.K., Skarshaug, R., Aamodt, H., Frafjord, A., Müller, E., Hammarström, C., Beraki, K., Bækkevold, E.S., Woldbæk, P.R., et al. (2018). Immune cell composition in human non-small cell lung cancer. Front. Immunol. *9*, 3101.

16. Geng, Y., Shao, Y., He, W., Hu, W., Xu, Y., Chen, J., Wu, C., and Jiang, J. (2015). Prognostic role of tumor-infiltrating lymphocytes in lung cancer: a meta-analysis. Cell. Physiol. Biochem. *37*, 1560–1571.

17. Zheng, X., Hu, Y., and Yao, C. (2017). The paradoxical role of tumor-infiltrating immune cells in lung cancer. Intractable Rare Dis. Res. *6*, 234–241.

18. Bremnes, R.M., Dønnem, T., Al-Saad, S., Al-Shibli, K., Andersen, S., Sirera, R., Camps, C., Marinez, I., and Busund, L.T. (2011). The role of tumor stroma in cancer progression and prognosis: emphasis on carcinoma-associated fibroblasts and non-small cell lung cancer. J. Thorac. Oncol. *6*, 209–217.

19. Denton, A.E., Roberts, E.W., and Fearon, D.T. (2018). Stromal cells in the tumor microenvironment. Adv. Exp. Med. Biol. *1060*, 99–114.

20. O'Malley, G., Treacy, O., Lynch, K., Naicker, S.D., Leonard, N.A., Lohan, P., Dunne, P.D., Ritter, T., Egan, L.J., and Ryan, A.E. (2018). Stromal cell PD-L1 inhibits CD8(+) T-cell antitumor immune responses and promotes colon cancer. Cancer Immunol. Res. *6*, 1426–1441.

21. Bochet, L., Lehuédé, C., Dauvillier, S., Wang, Y.Y., Dirat, B., Laurent, V., Dray, C., Guiet, R., Maridonneau-Parini, I., Le Gonidec, S., et al. (2013). Adipocyte-derived fibroblasts promote tumor progression and contribute to the desmoplastic reaction in breast cancer. Cancer Res. *73*, 5657–5668.

22. Min, K.W., Kim, D.H., Noh, Y.K., Son, B.K., Kwon, M.J., and Moon, J.Y. (2021). Cancer-associated fibroblasts are associated with poor prognosis in solid type of lung adenocarcinoma in a machine learning analysis. Sci. Rep. *11*, 16779.

23. Bagaev, A., Kotlov, N., Nomie, K., Svekolkin, V., Gafurov, A., Isaeva, O., Osokin, N., Kozlov, I., Frenkel, F., Gancharova, O., et al. (2021). Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell *39*, 845–865.e7.

24. Tomida, S., Takeuchi, T., Shimada, Y., Arima, C., Matsuo, K., Mitsudomi, T., Yatabe, Y., and Takahashi, T. (2009). Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. J. Clin. Oncol. *27*. https://doi.org/10.1200/JCO.2008.19.7053.

25. Rousseaux, S., Debernardi, A., Jacquiau, B., Vitte, A.L., Vesin, A., Nagy-Mignotte, H., Moro-Sibilot, D., Brichon, P.Y., Lantuejoul, S., Hainaut, P., et al. (2013). Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. Sci. Transl. Med. *5*. https://doi.org/10.1126/scitranslmed.3005723.

26. Botling, J., Edlund, K., Lohr, M., Hellwig, B., Holmberg, L., Lambe, M., Berglund, A., Ekman, S., Bergqvist, M., Pontén, F., et al. (2013). Biomarker discovery in non-small cell lung cancer: Integrating gene expression profiling, meta-analysis, and tissue microarray validation. Clin. Cancer Res. *19*. https://doi.org/10.1158/1078-0432.CCR-12-1139.

27. Tang, H., Xiao, G., Behrens, C., Schiller, J., Allen, J., Chow, C.W., Suraokar, M., Corvalan, A., Mao, J., White, M.A., et al. (2013). A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. Clin. Cancer Res. *19*. https://doi.org/10.1158/1078-0432.CCR-12-2321.

28. Der, S.D., Sykes, J., Pintilie, M., Zhu, C.Q., Strumpf, D., Liu, N., Jurisica, I., Shepherd, F.A., and Tsao, M.S. (2014). Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. J. Thorac. Oncol. *9*. https://doi.org/10.1097/JTO.0000000000000042.

29. Okayama, H., Kohno, T., Ishii, Y., Shimada, Y., Shiraishi, K., Iwakawa, R., Furuta, K., Tsuta, K., Shibata, T., Yamamoto, S., et al. (2012). Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res *72*. https://doi.org/10.1158/0008-5472.CAN-11-1403.

30. Shedden, K., Taylor, J.M.G., Enkemann, S.A., Tsao, M.S., Yeatman, T.J., Gerald, W.L., Eschrich, S., Jurisica, I., Giordano, T.J., Misek, D.E., et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. Nat. Med. *14*. https://doi.org/10.1038/nm.1790.

31. Schabath, M.B., Welsh, E.A., Fulp, W.J., Chen, L., Teer, J.K., Thompson, Z.J., Engel, B.E., Xie, M., Berglund, A.E., Creelan, B.C., et al. (2016). Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. Oncogene *35*. https://doi.org/10.1038/onc.2015.375.

32. Baty, F., Joerger, M., Früh, M., Klingbiel, D., Zappa, F., and Brutsche, M. (2017). 24h-gene variation effect of combined bevacizumab/erlotinib in advanced non-squamous non-small cell lung cancer using exon array blood profiling. J. Transl. Med. (15) https://doi.org/10.1186/s12967-017-1174-z.

33. Jung, H., Kim, H.S., Kim, J.Y., Sun, J.M., Ahn, J.S., Ahn, M.J., Park, K., Esteller, M., Lee, S.H., and Choi, J.K. (2019). DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. Nat. Commun. *10*, 4278.

34. Cho, J.W., Hong, M.H., Ha, S.J., Kim, Y.J., Cho, B.C., Lee, I., and Kim, H.R. (2020). Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-PD-1 therapy in non-small cell lung cancer. Exp. Mol. Med. *52*, 1550–1563.

35. Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science *355*, eaaf8399.

36. Cao, S., Wang, J.R., Ji, S., Yang, P., Dai, Y., Guo, S., Montierth, M.D., Shen, J.P., Zhao, X., Chen, J., et al. (2022). Estimation of tumor cell total mRNA expression in 15 cancer types predicts disease progression. Nat. Biotechnol. *40*, 1624–1633.

37. Wakabayashi, O., Yamazaki, K., Oizumi, S., Hommura, F., Kinoshita, I., Ogura, S., Dosaka-Akita, H., and Nishimura, M. (2003). CD4+ T cells in cancer stroma, not CD8+ T cells in cancer cell nests, are associated with favorable prognosis in human non-small cell lung cancers. Cancer Sci. *94*, 1003–1009.

38. Tomita, M., Matsuzaki, Y., and Onitsuka, T. (1999). Correlation between mast cells and survival rates in patients with pulmonary adenocarcinoma. Lung Cancer *26*, 103–108.

39. Iglesia, M.D., Parker, J.S., Hoadley, K.A., Serody, J.S., Perou, C.M., and Vincent, B.G. (2016). Genomic analysis of immune cell infiltrates across 11 tumor types. J. Natl. Cancer Inst. *108*, djw144.

40. Anichini, A., Perotti, V.E., Sgambelluri, F., and Mortarini, R. (2020). Immune escape mechanisms in non small cell lung cancer. Cancers *12*, 3605.

41. West, J., Schenck, R.O., Gatenbee, C., Robertson-Tessi, M., and Anderson, A.R.A. (2021). Normal tissue architecture determines the evolutionary course of cancer. Nat. Commun. *12*, 2060.

42. Herbst, R.S., Morgensztern, D., and Boshoff, C. (2018). The biology and management of non-small cell lung cancer. Nature *553*, 446–454.

43. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.

44. Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., Li, Z., Traugh, N., Bu, X., Li, B., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat. Med. *24*, 1550–1558.

45. Owens, N.D., De Domenico, E., and Gilchrist, M.J. (2019). An RNA-seq protocol for differential expression analysis. Cold Spring Harb. Protoc. *2019*. pdb.prot098368.

46. Hundal, J., Carreno, B.M., Petti, A.A., Linette, G.P., Griffith, O.L., Mardis, E.R., and Griffith, M. (2016). pVAC-Seq: a genome-guided in silico approach to identifying tumor neoantigens. Genome Med. *8*, 11.

47. Hundal, J., Kiwala, S., Feng, Y.Y., Liu, C.J., Govindan, R., Chapman, W.C., Uppaluri, R., Swamidass, S.J., Griffith, O.L., Mardis, E.R., and Griffith, M. (2019). Accounting for proximal variants improves neoantigen prediction. Nat. Genet. *51*, 175–179.

48. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678.

49. Xie, Y., Wang, X., and Story, M. (2009). Statistical methods of background correction for Illumina BeadArray data. Bioinformatics *25*, 751–757.

50. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell *173*, 400–416.e11.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
LUAD patient tumor tissues	National Cancer Center/Cancer Hospital and Chinese Academy of Medical Sciences	N/A
Deposited data		
Somatic variants, processed RNA-seq data, the immune cell fractions of tumor microenvironment and clinical information of TCGA LUAD samples	Genomic Data Commons (GDC) data portal	https://gdc.cancer.gov/
Processed microarray data of human LUAD	Tomida et al.,	GEO: GSE13213
Processed microarray data of human LUAD	Rousseaux et al.,	GEO: GSE30219
Processed microarray data of human LUAD	Botling et al.,	GEO: GSE37745
Processed microarray data of human LUAD	Tang et al.,	GEO: GSE42127
Processed microarray data of human LUAD	Der et al.,	GEO: GSE50081
Processed microarray data of human LUAD	Okayama et al.,	GEO: GSE31210
Processed microarray data of human LUAD	Shedden et al.,	GEO: GSE68465
Processed microarray data of human LUAD	Schabath et al.,	GEO: GSE72094
Processed microarray data of human LUAD	Baty et al.,	GEO: GSE61676
Processed RNA-seq data of human NSCLC	Jung et al.[33]	GEO: GSE135222
Processed RNA-seq data of human NSCLC	Cho et al.[34]	GEO: GSE126044
Immune and stroma cell types	Aran et al.[43]	N/A
Critical commerical assays		
TIANamp Genomic DNA kit	Tiangen Biotech, Beijing, China	N/A
Quanti-IT dsDNA HS Assay Kit	Thermo Fisher Scientific, MA, USA	N/A
DNBSEQ-T7R platform	MGI, Shenzhen, China	N/A
Software and algorithms		
Burrows-Wheeler Alignment tool (BWA; version 0.7.17)	N/A	https://github.com/lh3/bwa
netMHC (version 4.034)	N/A	N/A
Human genome (hg19)	Genome Reference ConsortiumHuman Build	genome.ucsc.edu
R (version 4.2)	N/A	N/A
Rstudio	N/A	https://support–rstudio-com.netlify.app/products/rstudio/
xCell	Aran et al.[43]	https://xcell.ucsf.edu/
The Tumor Immune Dysfunction and Exclusion (TIDE)	Jiang et al.[44]	http://tide.dfci.harvard.edu/
survival	R package	N/A
survminer	R package	N/A
ggplot2	R package	N/A
survivalROC	R package	N/A
Meta	R package	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
maftools	R package	N/A
forester	R package	N/A
Szcox	R package	N/A
SubgrPlots	R package	N/A
ComplexHeatmap	R package	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Zhijie Wang (jie_969@163.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data from in-house cohorts are available from the corresponding author on reasonable request. Data from publicly archive datasets are available from the Cancer Genome Atlas (TCGA), gene expression omnibus (GEO) database, as publications cited in the manuscript. These accession numbers for the datasets are also listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient recruitment & sample acquisition

Two in-house datasets containing 203 stage I-III LUAD patients (female: 116 (57%)) who received surgery and treated with adjuvant chemotherapy or TKI and 30 advanced-stage LUAD patients (female: 11 (37%)) who treated with anti-PD-(L)1 antibodies at National Cancer Center/Cancer Hospital and Chinese Academy of Medical Sciences from April 2014 to April 2022 (named NCC cohort, or Wang cohort, and NCC-ICIs cohort) were also included. All patients have informed consent, and this study was approved by the ethics committees of the National Cancer Center (NCC-22/250-3454, NCC-22/429-3631).

METHOD DETAILS

Public lung datasets

The public datasets with mRNA expression and clinical information of LUAD were searched from The Cancer Genome Atlas (TCGA), gene expression omnibus (GEO) database. Overall, 12 public cohorts were gathered for this study, including 10 LUAD cohorts (TCGA-LUAD, GSE13213, GSE30219, GSE31210, GSE37745, GSE42127, GSE50081, GSE68465, GSE72094, GSE61676 and two advanced stage non-small-cell lung cancer (NSCLC) cohort with immune checkpoint inhibitor (ICI) treatment (GSE135222 and GSE126044). Patients without follow-up information were removed from further evaluation. The demographic and clinical features of each cohort are listed in Table 1.

RNA sequencing

198 and 22 patients respectively from NCC cohort and NCC-ICIs cohort underwent RNA sequencing. RNA extraction, sequencing library construction, sequencing and FASTQ data quality control were performed in accordance with the protocol by Nick D.L. Owens et al.[45]

Whole exome sequencing

202 and 20 patients respectively from NCC cohort and NCC-ICIs cohort underwent whole exome sequencing (WES). Genomic DNA was extracted using the TIANamp Genomic DNA kit (Tiangen Biotech,

Beijing, China) following manufacturer's instruction. DNA concentration and purity were estimated using Nanodrop 2000 spectrophotometer and Qubit 2.0 Fluorometer with Quanti-IT dsDNA HS Assay Kit (Thermo Fisher Scientific, MA, USA). Library construction was performed using a custom 53M length capturing probe, made by Integrated DNA Technologies (IDT, IA, USA), and covering the coding regions of all genes and partial non-coding regions. Captured libraries were then pair-end sequenced in 100bp lengths with DNBSEQ-T7R platform (MGI, Shenzhen, China) following the manufacturer's guidance. Raw data was filtered to remove low-quality reads and adaptor sequence. Reads were further mapped to the reference human genome (hg19) utilizing BWA aligner (version 0.7.10) for mutation calling.

TMB was defined as the number of nonsynonymous SNVs and indels in examined coding regions with the variated allele frequency (VAF) $\geq$ 1% in tumor tissues. TNB is defined as the number of neoantigens. To screen the neoantigen, we employed depth-based filters as follows: any variants with normal coverage <= 5× and normal VAF of $\geq$2% were filtered out. The normal coverage cutoff can be increased up to 20× to eliminate occasional misclassification of germline variants as somatic. For tumor coverage from DNA, a cutoff is placed at $\geq$10x with a VAF of $\geq$40%. To further evaluate the effect of relevant nearby variants on neoantigen identification, we used netMHC- 4.034 an updated version of the pVAC tools software to assess the binding affinities of the neoantigens with the corrected mutant peptide sequence.[46] TNB is defined as the number of neoantigens obtained through the above prediction process.[47]

Public data acquisition and pre-processing

Level 3 RNA sequencing data (FPKM format) of TCGA-LUAD were downloaded from UCSC Xena browser (https://xenabrowser.net/datapages/), and the FPKM values were log2-transformed into log transcripts per kilobase million (log2(TPM+1)) values.[48] For GEO datasets, the microarray data sets from GSE13213 and GSE42127 generated by Agilent and Illumina platform, were processed using locally weighted scatterplot smoothing (LOWESS) normalization and Model-Based Background Correction (MBCB) method,[49] respectively. The other microarray data sets from Affymetrix were processed using the robust multichip average (RMA) algorithm in the 'affy' R package, including background adjustment, quantile normalization, and final summarization of oligonucleotides per transcript using the median polish algorithm. The RNA sequencing data (TPM format) of GSE135222 cohort was download from GEO database. The RNA sequencing data (Count format) of GSE126044 cohort was downloaded from GEO database and were converted into log2(CPM+0.001).

The clinicopathological data of these data sets were also collected. For TCGA data, clinical and genomic data were obtained from the Genomic Data Commons (https://portal.gdc.cancer.gov/) using the R package "TCGA biolinks". Complete survival information of TCGA-LUAD was obtained from the supplementary data of the published research.[50] The clinical data of GEO data sets were downloaded from the corresponding dataset page in the GEO website (https://www.ncbi.nlm.nih.gov/geo/) and analyzed with the "GEOquery" package. To minimize the bias generating from different platforms or sequencing methods in different datasets, meta-analysis consisting of six independent cohorts has been used in the training phase to construct and internally validate the discriminative power of the prognostic-related signature. Otherwise, three larger public datasets were used as independent validation sets, to confirm the robustness of the model, respectively (GSE72094, GSE68465 and GSE31210). Another public cohort (GSE135222) consisting of 27 patients with advanced-stage NSCLC who received PD-1 antibody with whole-exome, transcriptomes and clinical survival was collected from a previous study.[33] Another public cohort (GSE126044) consisting of 16 patients with advanced-stage NSCLC who received PD-1 antibody regimen with transcriptomes and drug response was collected from a previous study.[34]

Estimates of cells enrichment scores

To quantify proportions of immune cells in LUAD samples, we used the xCell algorithm, which could convert mRNA expression profiles to enrichment scores of 64 immune and stromal cell types across samples, including multiple adaptive and innate immunity cells, hematopoietic progenitors, epithelial cells, and extracellular matrix cells (Table S2). The normalized Z-score matrix of microarray data was uploaded to the xCell website (https://xcell.ucsf.edu/). Subsequently, 64 cell types including stromal cells and immune cells with enrichment scores were generated. The pre-calculated enrichment scores by xCell were obtained for TCGA-LUAD samples from a previous study.[43]

Prognosis-related markers selection and signature construction

Each cell was transformed into binary variables with the cutoff of median enrichment score in the six independent training data sets. Then univariable Cox regression analysis was applied to identify the prognostic efficacy between high and low tumor-infiltrating cell enrichment score in each training set. The hazard ratios (HRs) were generated from a meta-analysis of six training sets for each cell type. Subsequently, the cell type which was supposed to be associated with prognosis ($P<0.05$) was selected to develop the tumor stromal cell score (TSCS) and TICS. The TSCS and TICS was calculated by the formula:

$$TICS\ or\ TSCS = \sum_{i=1}^{n} \frac{HR_i - 1}{se(HR)} * \text{xCell enrichment score}$$

where *HRi* was the hazard ratio of *ith* tumor-infiltrating cell from meta-analysis and xCell enrichment score represented the enrichment score of each cell. The median value of cell scores in different cohorts was selected to stratify patients into high- and low-risk subgroups. The Kaplan-Meier (KM) method was used to generate survival curves for the subgroups, and the log-rank (Mantel-Cox) test was used to determine the statistical significance of differences.

Evaluation and characteristics identification of TICS

To validate the prognostic performance of TICS, three public datasets and one in-house cohort were included as independent validation cohorts. Briefly, TICS was calculated for each patient. Stratification analysis was applied to compare survival between high- and low-risk groups regarding to age, sex, tumor stage, smoking, *EGFR, KRAS, ALK* and *TP53* mutation. Then gene set enrichment analysis (GSEA) of Gene Ontology and KEGG were applied to investigate the potential difference in the biological function between high- and low-risk subgroups using the clusterProfiler R package. Additionally, single-sample gene set enrichment analysis (ssGSEA) method which based on 28 immune gene sets was used to quantify the relative abundance of these immune gene sets in TCGA and NCC cohort. Meanwhile, we explored the different mRNA expression of chemokine and immune checkpoints. The immune context and genetic characteristics were compared in the high- and low-risk groups, including TMB, tumor immune dysfunction and exclusion (TIDE),[44] somatic copy number alterations (SCNA) and total mRNA expression (TmS).

QUANTIFICATION AND STATISTICAL ANALYSIS

Unpaired Student t test was performed to compare normally distributed variables between two groups, and Mann-Whitney U test was used to compare non-normally distributed variables between two groups. Kaplan-Meier curves were used to generate survival curves. Log-rank test was used to compare the difference between survival curves, and Cox regression analysis was used to determine the HR and corresponding 95% confidence interval (95% CI). The prognostic values of the genes in TICS were accessed by the "szcox" function in R package 'ezcox'. R package 'SubgrPlots' was used for subgroup analysis, which was visualized by the forester package. The ComplexHeatmap package was used to visualize the mutation landscape in TCGA-LUAD dataset. All statistical analyses were conducted using R (https://www.r-project.org/), and the P values were two-sided. Unless otherwise stated, P values of less than 0.05 were considered statistically significant.

ADDITIONAL RESOURCES

Clinical trial registry number: NCT03301688. Cinical trial URL: https://clinicaltrials.gov/ct2/show/NCT03301688?term=NCT03301688&draw=1&rank=1.