iScience, Volume *26*

Supplemental information

Machine learning-based risk model incorporating tumor immune and stromal contexture predicts cancer prognosis and immunotherapy efficacy

Li-Na He, Haifeng Li, Wei Du, Sha Fu, Linfeng luo, Tao Chen, Xuanye Zhang, Chen Chen, Yongluo Jiang, Yixing Wang, Yuhong Wang, Hui Yu, Yixin Zhou, Zuan Lin, Yuanyuan Zhao, Yan Huang, Hongyun Zhao, Wenfeng Fang, Yunpeng Yang, Li Zhang, and Shaodong Hong

Supplemental information

This file includes Figures S1-S12, Table S1, and Table S9-S14.

Tables S2-S8 that accompany this paper are provided in xlsx format.

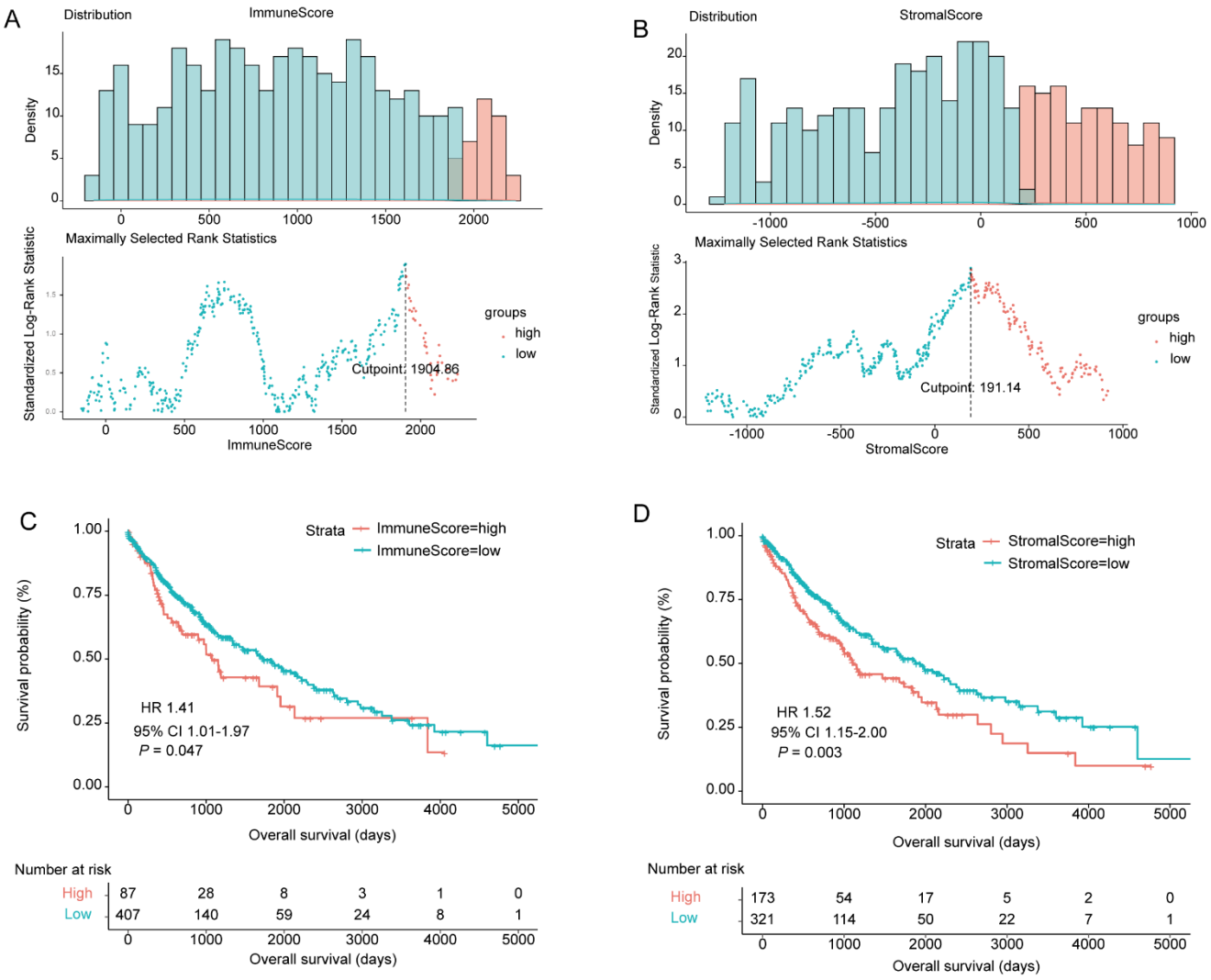
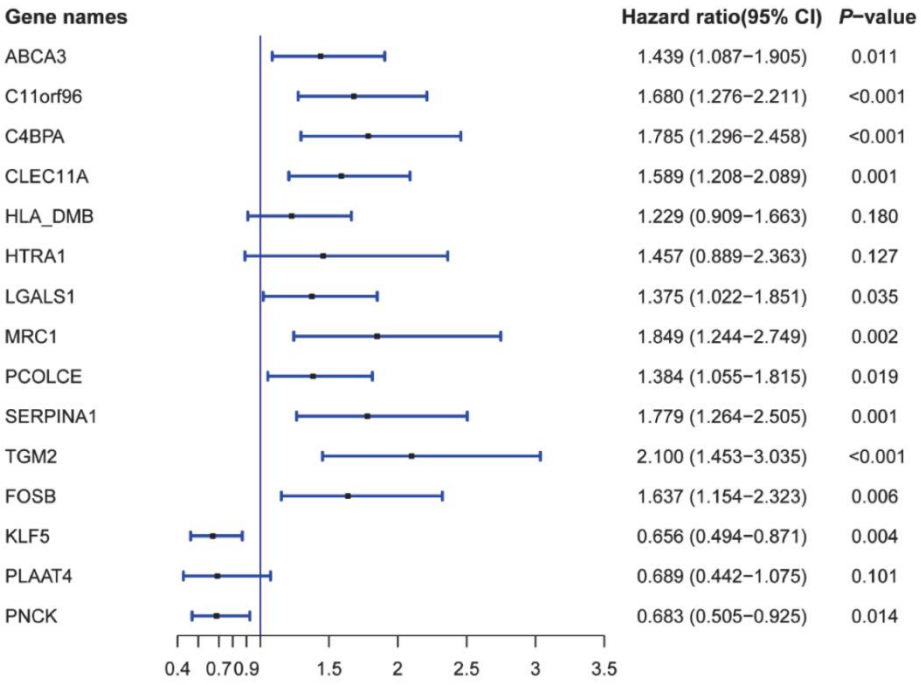


Figure S1. Identification of cut-points of immune scores and stromal scores and their association with prognosis, related to Figure 2. (A-B) Optimal cut-offs for immune score (A) and stromal score (B) identified by the maximally selected rank statistics. Scatters in the bottom panel indicate corresponding standardized log-rank statistics of every gene expression cut-point. The vertical dashed line indicates the optimal cut-point. The upper panel presents the density distribution histogram of the low- and high-immune/stromal score groups. (C-D) Kaplan–Meier survival analysis based on immune score (C) and stromal score (D) strata.



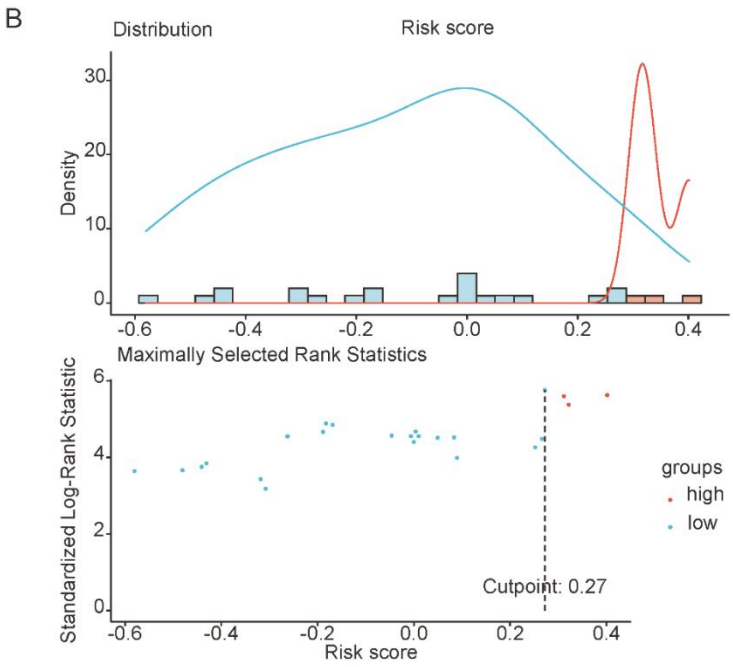


Figure S2. Univariate analyses of candidate mode genes and identification of cut-point for TMErisk, related to Figure 2 and Figure 3. (A) Forest plot for hazard ratios of OS in the TCGA-LUSC training cohort according to 15 candidate mode genes. (B) Optimal cut-off for TMErisk identified by the maximally selected rank statistics. Scatters in the bottom panel indicate corresponding standardized log-rank statistics of every gene expression cut-point. The vertical dashed line indicates the optimal cut-point. The upper panel presents the density distribution histogram of the low- and high-TMErisk groups. OS, overall survival; TCGA, the cancer genome atlas; LUSC, lung squamous cell carcinoma.

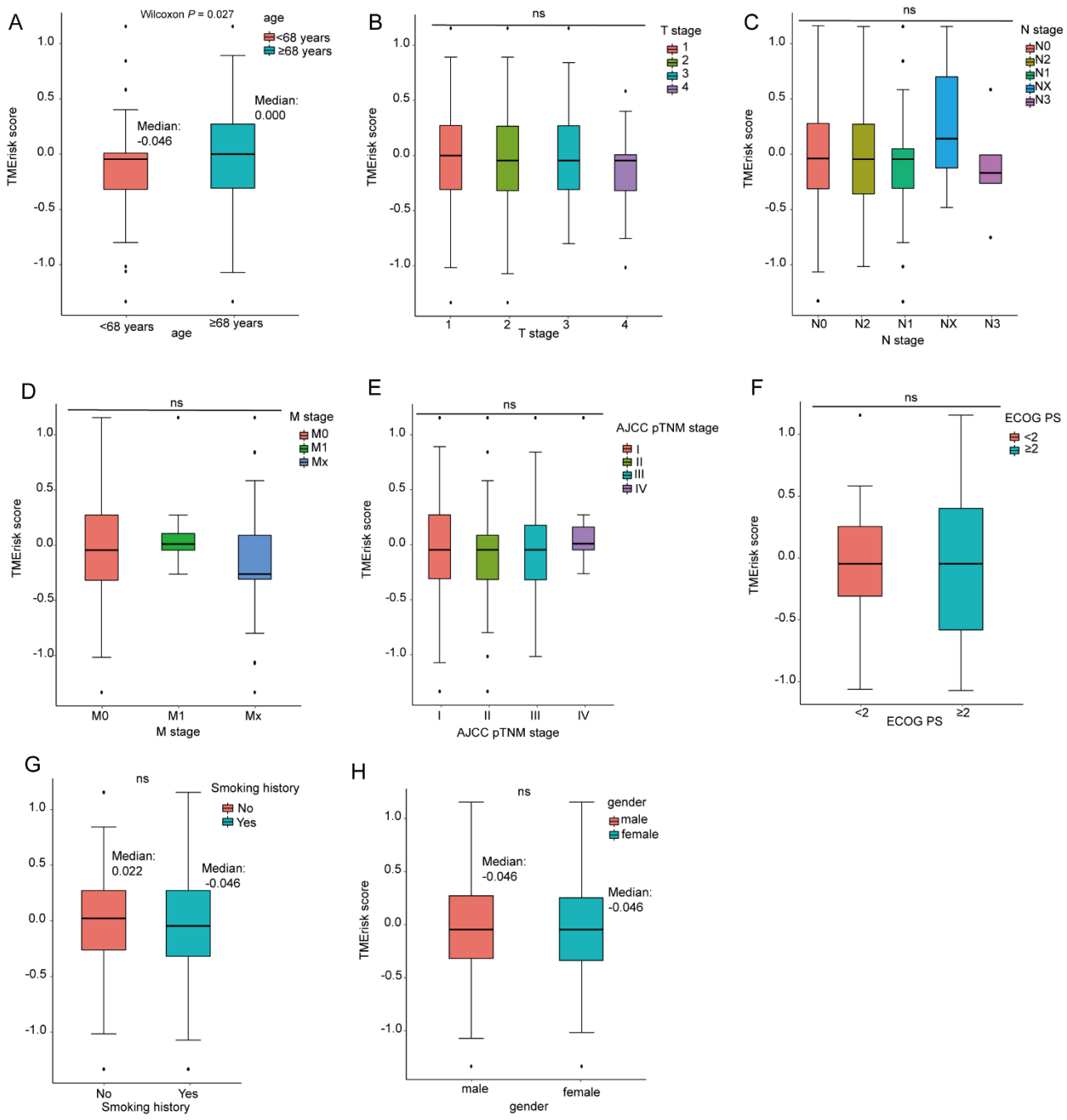


Figure S3. Distribution of TMErisk according to different baseline clinicopathological characteristics in the TCGA-LUSC cohort, related to Figure 2. Statistical difference between two groups was tested by Wilcoxon rank sum test, and Kruskal-Wallis test was used for comparison among three or more groups. *P < 0.05; ns, not significant; ECOG, Eastern Cooperative Oncology Group; PS, performance status.

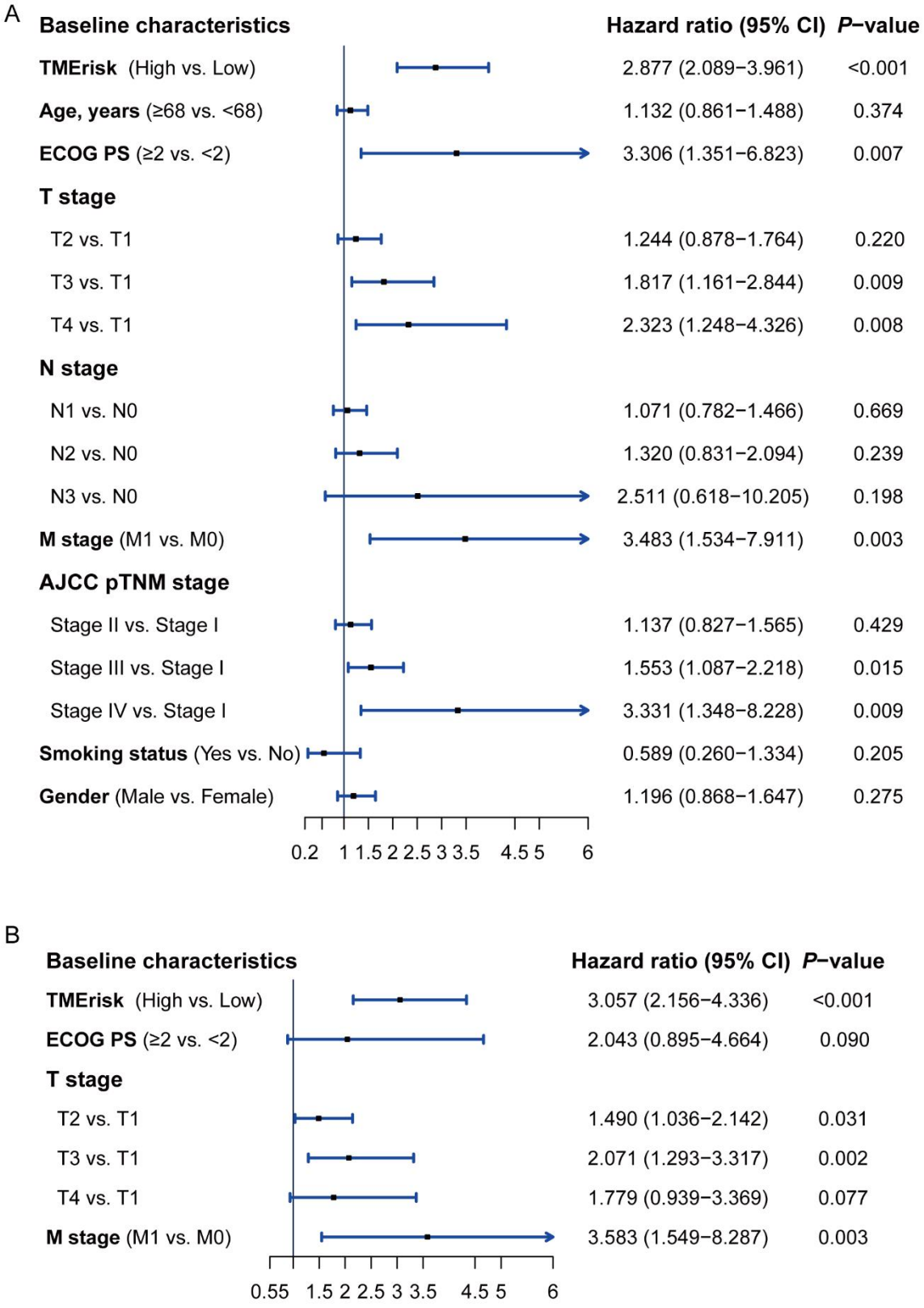


Figure S4. Forest plot for hazard ratios of OS in the TCGA-LUSC training cohort according to TMErisk and other baseline clinicopathological characteristics, related to Figure 2. (A) Univariate Cox analysis for OS according to different variables. (B) Multivariate Cox analysis for OS according to different variables. OS, overall survival.

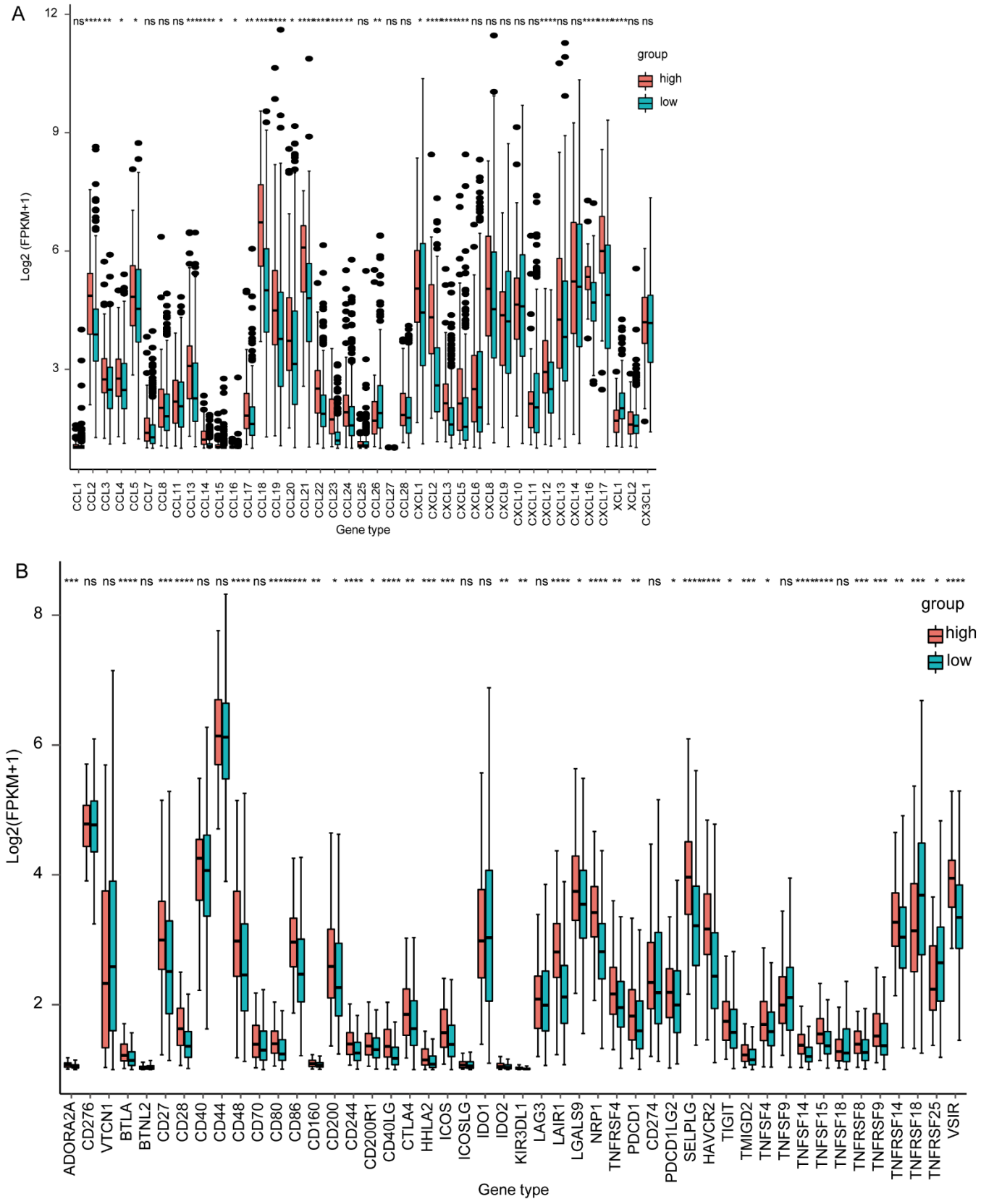


Figure S5. Association of TMErisk with chemokine genes and immune-related genes, related to Figure 4. (A) Expression level of chemokine genes in the low- and high-TMErisk groups. (B) Expression level of immune-related genes in the low- and high-TMErisk groups. Differences were examined by Wilcoxon test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; ns, not significant.

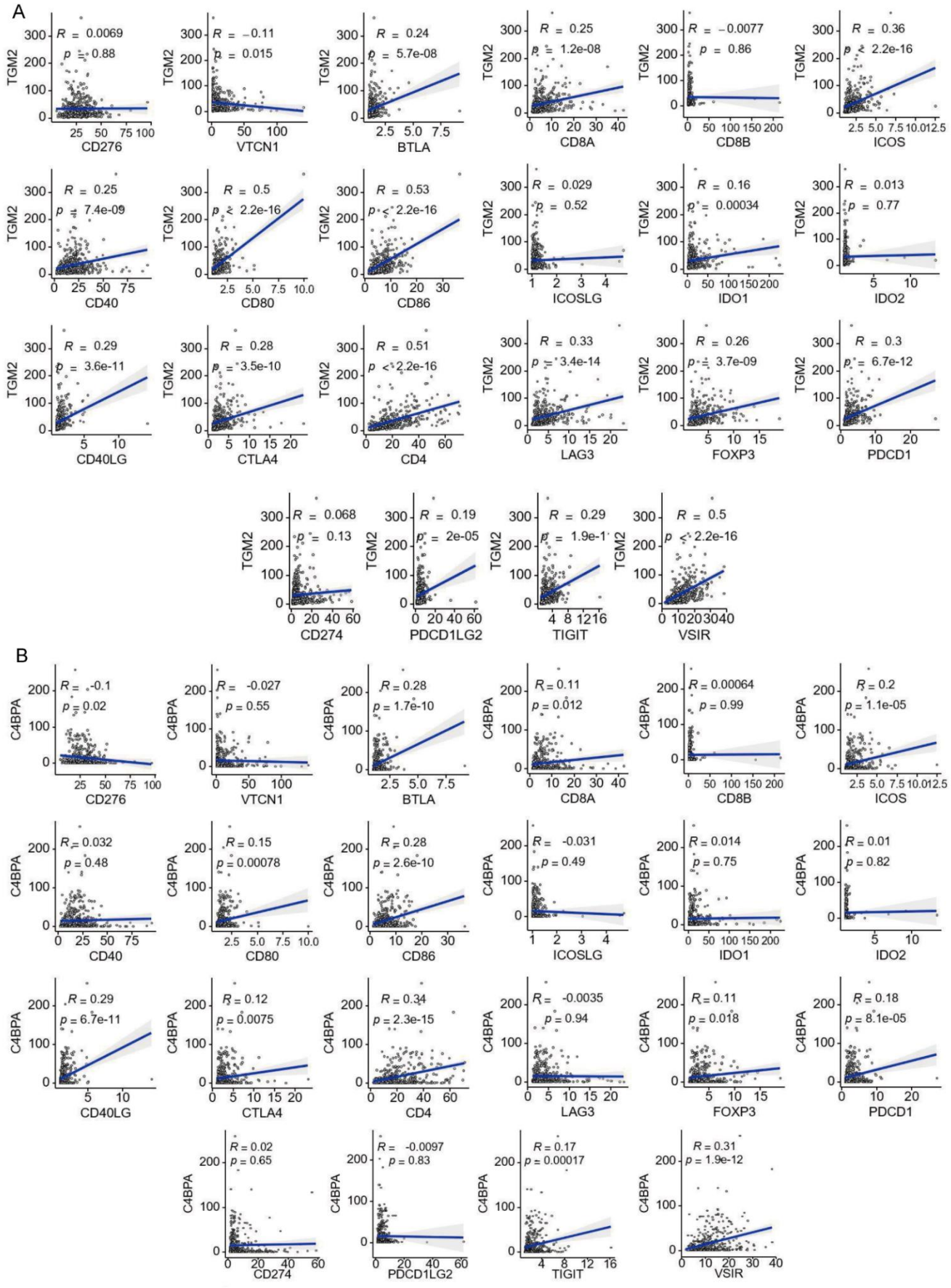


Figure S6. Pearson correlation analysis between selected immune-related genes and TGM2 (A) as well as C4BPA (B), related to Figure 4.

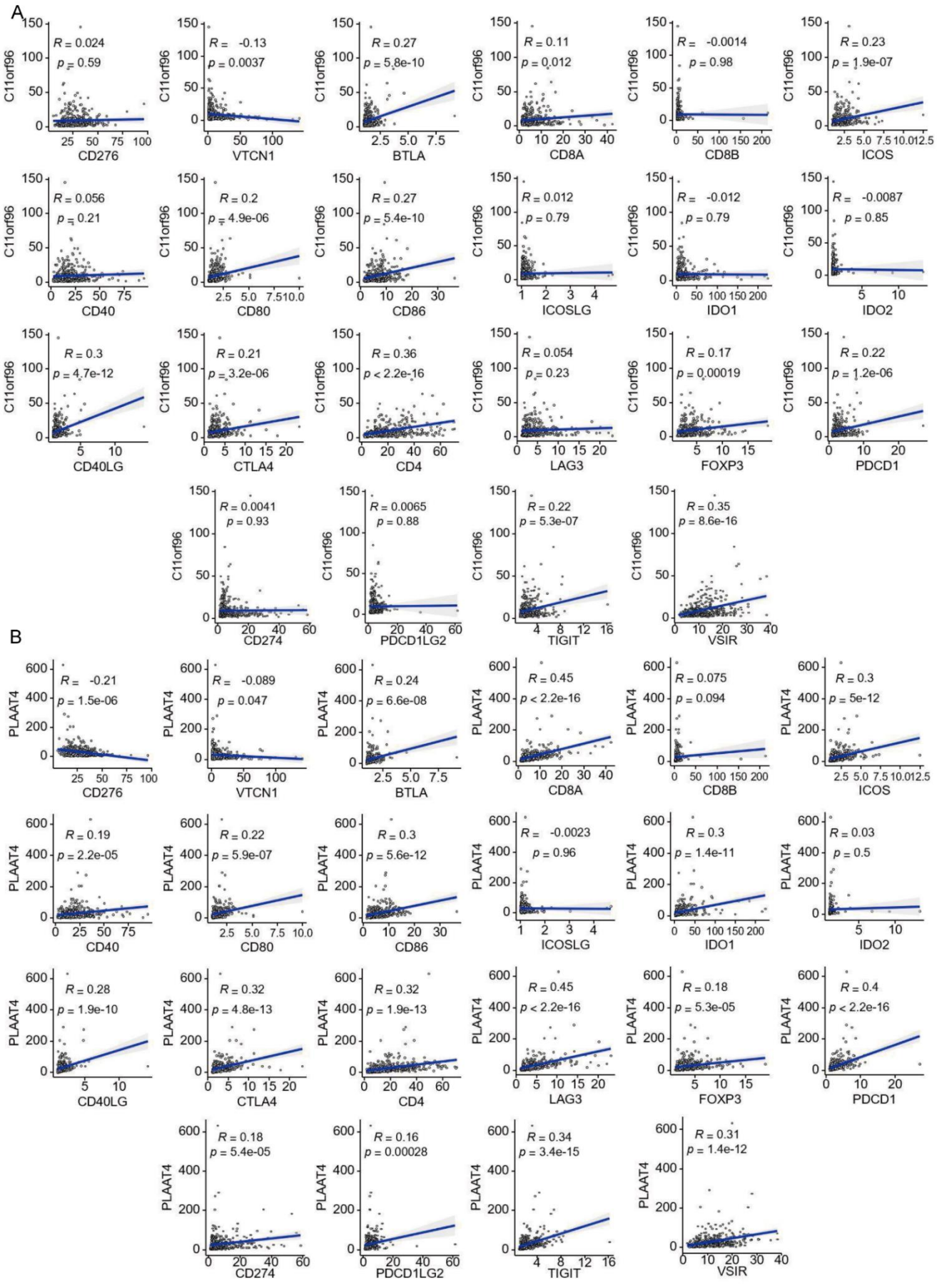


Figure S7. Pearson correlation analysis between selected immune-related genes and C11orf96 (A) as well as PLAAT4 (B), related to Figure 4.

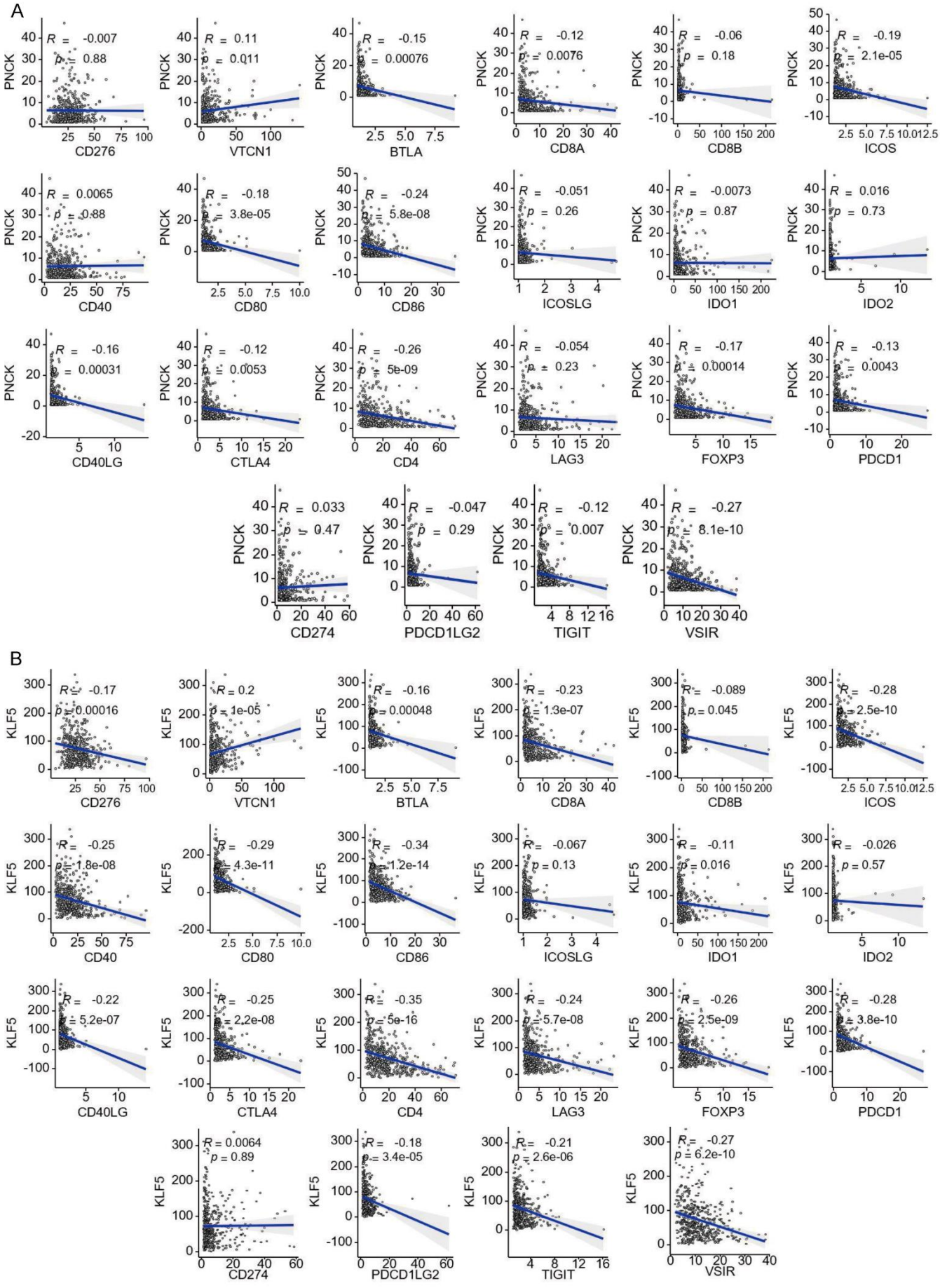


Figure S8. Pearson correlation analysis between selected immune-related genes and PNCK (A) as well as KLF5 (B), related to Figure 4.

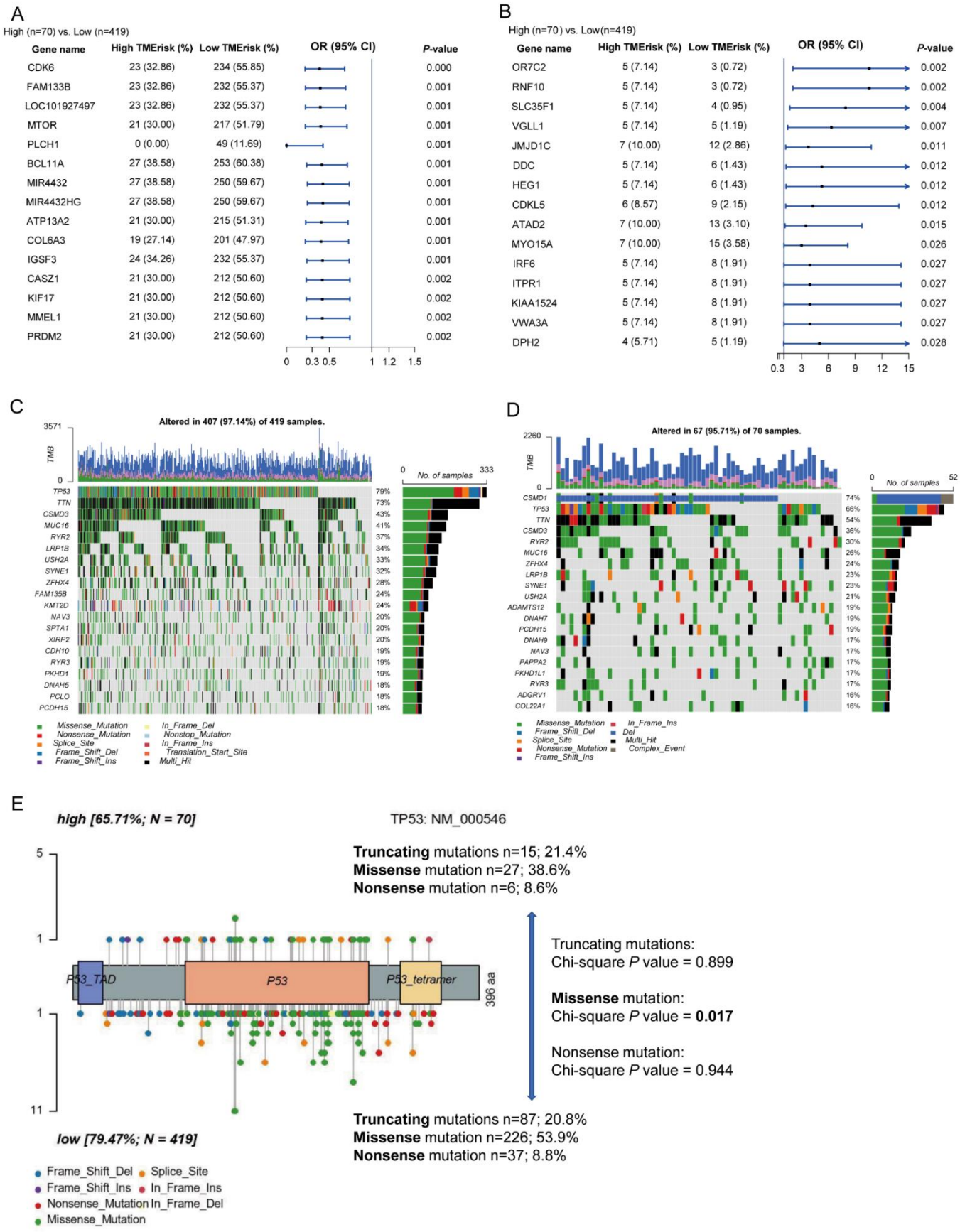


Figure S9. Mutation status between the low- and high-TMErisk groups, related to STAR Methods. (A-B) Comparison of gene mutation frequencies between the low- and high-TMErisk groups. Forest plots show the top 15 genes mutated more frequently in the low-TMErisk (A) and high-TMErisk (B) group. (C) The top 20 mutated genes and distribution of mutation types in the low-TME risk group. (D) The top 20 mutated genes and distribution of mutation types in the high-TME risk group. (E) The lollipop plot shows the mutation types and mutation sites of TP53 according to TMErisk strata. OR, odds ratio.

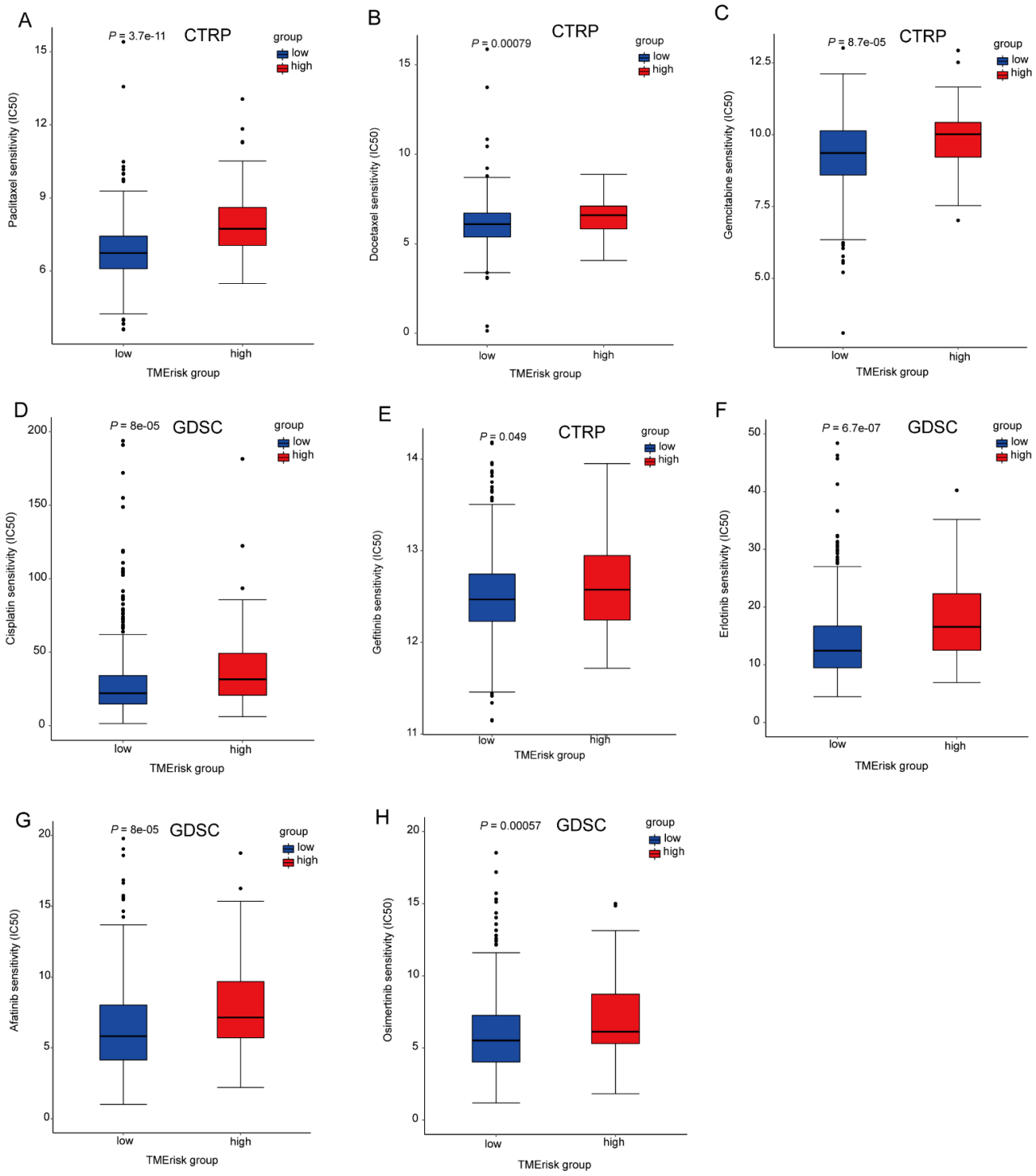


Figure S10. Prediction of drug sensitivity for the low- and high-TMErisk groups, related to Figure 5 and Figure 6. Prediction was generated based on GDSC and CTRP cell line databases. Statistical difference was examined by Wilcoxon test. CTRP, Cancer Therapeutics Response Portal; GDSC, Genomics of Drug Sensibility in Cancer.

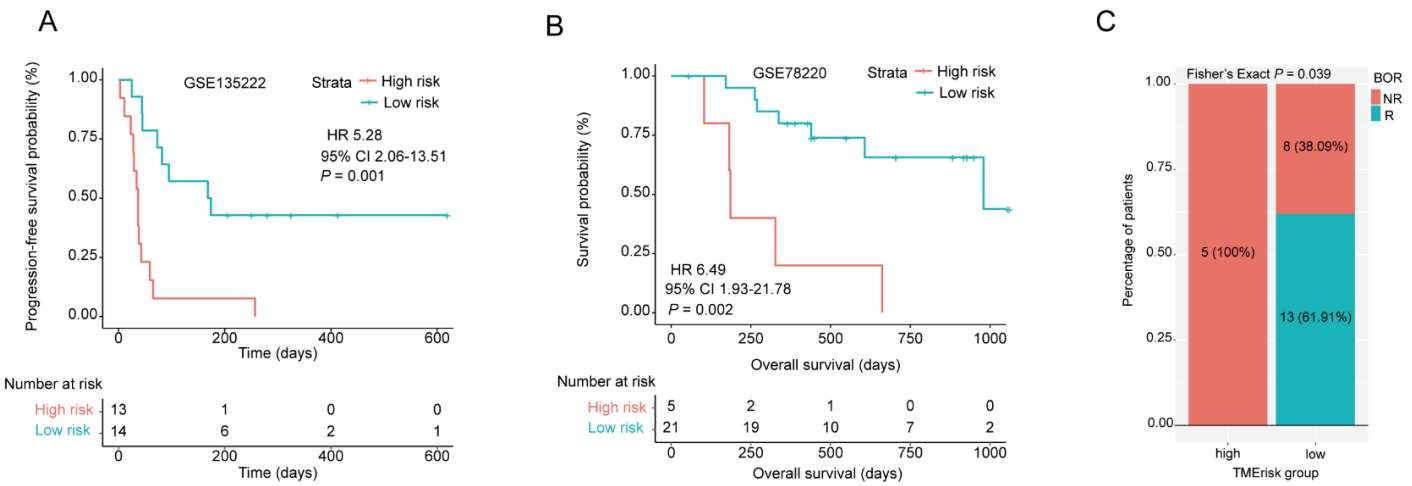


Figure S11. Validation of predictive value of TMErisk in GEO datasets, related to Figure 5. (A) Kaplan–Meier analysis for progression-free survival based on TMErisk strata in the GSE135222 cohort. (B) Kaplan–Meier analysis for overall survival based on TMErisk strata in the GSE78220 cohort. (C) Association of TMErisk with respond to immunotherapy in the GSE78220 cohort. BOR, best overall response; R, responders, patients with BOR of complete response or partial response; NR, non-responders, patients with BOR of stable disease or progressive disease.

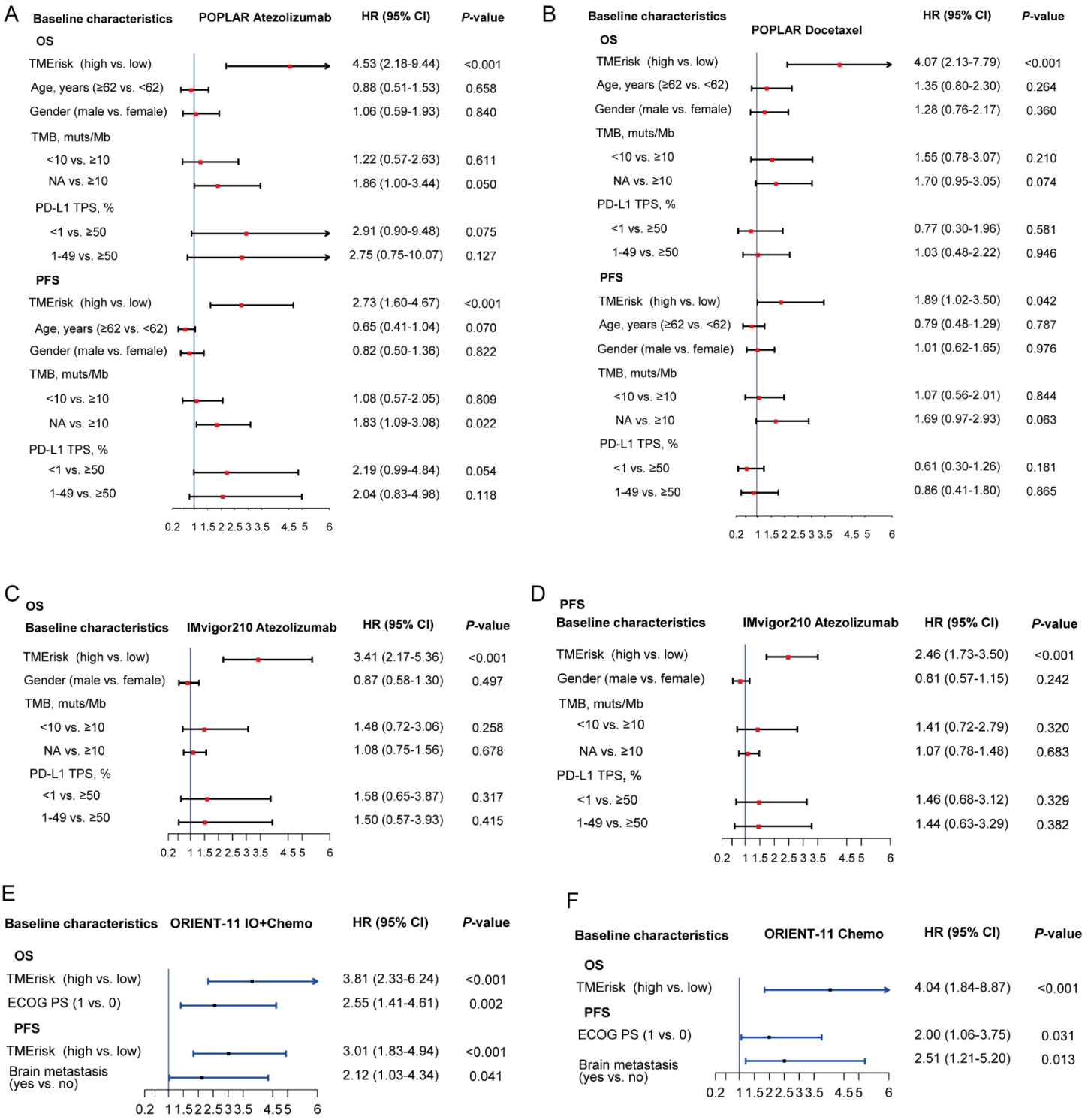


Figure S12. Forest plot for hazard ratios of OS and PFS in the POPLAR, IMvigor210 and ORIENT-11 cohorts according to TMErisk and other baseline clinicopathological characteristics, related to Figure 6 and Figure 7. (A-B) Univariate Cox analysis for OS and PFS according to different variables in the POPLAR immunotherapy (A) and chemotherapy (B) datasets. (C-D) Univariate Cox analysis for OS (C) and PFS (D) according to different variables in the IMvigor210 cohort. (E-F) Multivariate Cox analysis for OS and PFS in the immunochemotherapy (E) and chemotherapy (F) groups from ORIENT-11 cohort.

HR, hazard ratio; CI, confidence interval; OS, overall survival; PFS, progression-free survival; TMB, tumor mutation burden; PD-L1, programmed cell death ligand 1; TPS, tumor proportion score; Chemo, chemotherapy; IO+Chemo, immunotherapy plus chemotherapy; ECOG, Eastern Cooperative Oncology Group; PS, performance status

Table S1. Baseline patient characteristics in the TCGA-LUSC cohort, related to STAR Methods

Patient characteristics	TCGA-LUSC cohort (N = 494)
Age, years	
Median (range)	68 (39-90)
< 68, n (%)	229 (46.4)
≥ 68, n (%)	260 (52.6)
NA, n (%)	5 (1.0)
Gender, n (%)	
Male	366 (74.1)
Female	128 (25.9)
ECOG PS, n (%)	
< 2	88 (17.8)
≥ 2	21 (4.3)
NA	385 (77.9)
Smoking status, n (%)	
Never smoker	18 (3.6)
Current or former smoker	464 (93.9)
NA	12 (2.4)
T stage, n (%)	
T1	114 (23.1)
T2	287 (58.1)
T3	70 (14.2)
T4	23 (4.6)
N stage, n (%)	
N0	316 (64.0)
N1	127 (25.7)
N2	40 (8.1)
N3	5 (1.0)
Nx	6 (1.2)
M stage, n (%)	
M0	405 (82.0)
M1	8 (1.6)
Mx	77 (15.6)
NA	4 (0.8)

AJCC pTNM stage, n (%)	
Stage I	242 (49.0)
Stage II	158 (32.0)
Stage III	83 (16.8)
Stage IV	7 (1.4)
NA	4 (0.8)
TMErisk score	
Median (range)	-0.45 (-1.33-1.15)
Low risk, n (%)	421 (85.2)
High risk, n (%)	73 (14.8)

Abbreviations: ECOG, Eastern Cooperative Oncology Group; PS, performance status; AJCC, American Joint Committee on Cancer; NA, not available.

Table S9. Infiltration levels of immune and stromal cells in the low- and high-TMErisk groups, related to Figure 4

Program	Enriched in low-TMErisk group	Enriched in high-TMErisk group
CIBERSORT	Plasma cells, Monocytes, Dendritic cells resting, Mast cells activated	B cells memory, T cells regulatory (Tregs), Macrophages M2, Eosinophils, Neutrophils
EPIC	T cell $CD8^+$	B cell, Cancer associated fibroblast, T cell $CD4^+$, Endothelial cell
xCell	T cell $CD4^+$ central memory, T cell $CD8^+$ naive, Common lymphoid progenitor, T cell $CD4^+$ Th1, T cell $CD4^+$ Th2	Myeloid dendritic cell activated, T cell $CD4^+$ effector memory, Class-switched memory B cell, Myeloid dendritic cell, Endothelial cell, Eosinophil, Cancer associated fibroblast, Granulocyte-monocyte progenitor, Hematopoietic stem cell, Macrophage, Macrophages M1, Macrophages M2, Mast cell, Monocyte, Neutrophil, T cell NK, Tregs

Table S10. Univariate Cox analyses for OS and PFS in the immunochemotherapy group from ORIENT-11, related to Figure 7

Patient characteristics	OS		PFS	
	HR (95% CI)	***P* value**	**HR (95% CI)**	***P* value**
Age, years				
> 60 vs. ≤ 60	1.45 (0.90-2.38)	0.121	1.53 (0.92-2.53)	0.101
Gender				
Male vs. Female	1.48 (0.81-2.67)	0.200	1.66 (0.89-3.09)	0.110
ECOG PS				
1 vs. 0	2.16 (1.20-3.90)	0.010	1.51 (0.87-2.63)	0.146
Smoking status				
Current/former smoker vs. Never smoker	1.07 (0.65-1.75)	0.791	1.10 (0.66-1.84)	0.720
PD-L1 TPS, %, n (%)				
1-49 vs. < 1	0.79 (0.42-1.47)	0.448	1.06 (0.57-1.98)	0.843
≥ 50 vs. < 1	0.64 (0.36-1.13)	0.122	0.73 (0.40-1.33)	0.301
Brain metastases, n (%)				
Yes vs. No	1.67 (0.85-3.28)	0.137	2.14 (1.04-4.35)	0.036
TMErisk score				
High vs. Low	3.42 (2.10-5.56)	< 0.001	3.01 (1.83-4.94)	< 0.001

Abbreviations: ECOG, Eastern Cooperative Oncology Group; PS, performance status; EGFR, epidermal growth factor receptor； ALK，anaplastic lymphoma kinase；PD-L1, programmed cell death ligand 1; TPS, tumor proportion score； HR, hazard ratio；CI, confidence interval；OS, overall survival; PFS, progression-free survival.

Table S11. Univariate Cox analyses for OS and PFS in the chemotherapy group from ORIENT-11, related to Figure 7

Patient characteristics	OS		PFS	
	HR (95% CI)	***P* value**	**HR (95% CI)**	***P* value**
Age, years				
> 60 vs. ≤ 60	0.75 (0.41-1.39)	0.365	1.06 (0.60-1.87)	0.838
Gender				
Male vs. Female	1.72 (0.79-3.73)	0.172	1.34 (0.69-2.62)	0.389
ECOG PS				
1 vs. 0	1.03 (0.54-2.00)	0.921	1.84 (0.99-3.43)	0.054
Smoking status				
Current/former smoker vs. Never smoker	1.25 (0.63-2.45)	0.523	0.99 (0.54-1.81)	0.968
PD-L1 TPS, %, n (%)				
1-49 vs. < 1	0.77 (0.30-1.96)	0.581	1.22 (0.56-2.69)	0.617
≥ 50 vs. < 1	1.45 (0.72-2.95)	0.295	1.06 (0.57-1.98)	0.860
Brain metastases, n (%)				
Yes vs. No	0.82 (0.38-1.78)	0.613	2.24 (1.10-4.56)	0.026
TMErisk score				
High vs. Low	4.04 (1.84-8.87)	< 0.001	1.27 (0.71-2.26)	0.423

Abbreviations: ECOG, Eastern Cooperative Oncology Group; PS, performance status; EGFR, epidermal growth factor receptor； ALK，anaplastic lymphoma kinase；PD-L1, programmed cell death ligand 1; TPS, tumor proportion score； HR, hazard ratio；CI, confidence interval；OS, overall survival; PFS, progression-free survival.

Table S12. Baseline patient characteristics in the POPLAR and IMvigor210 cohorts, related to STAR Methods

Patient characteristics	POPLAR		IMvigor210
	Atezolizumab (N = 81)	Docetaxel (N = 75)	(N = 208)
Age, years			
Median (range)	61 (42-82)	63 (36-80)	NA
< 62, n (%)	41 (50.6)	32 (42.7)	NA
≥ 62, n (%)	40 (49.4)	43 (57.3)	NA
Gender, n (%)			
Male	56 (69.1)	44 (58.7)	162 (77.9)
Female	25 (30.9)	31 (41.3)	46 (22.1)
TMB,muts/Mb			
< 10	15 (18.5)	16 (21.3)	10 (4.8)
≥ 10	35 (43.2)	34 (45.3)	134 (64.4)
Not available	31 (38.3)	25 (33.3)	64 (30.8)
PD-L1 TPS, %, n (%)			
< 1	55 (69.7)	37 (49.3)	164 (78.8)
1-49	17 (21.0)	27 (36.0)	35 (16.8)
≥ 50	9 (11.1)	11 (14.7)	9 (4.3)
TMErisk score			
Median (range)	0.25 (-1.53 – 1.27)	0.67 (-0.11 – 2.20)	0.46 (-1.13 – 1.42)

Abbreviations: TMB, tumor mutation burden; PD-L1, programmed cell death ligand 1; TPS, tumor proportion score.

Table S13. Baseline patient characteristics in the ORIENT-11 cohort, related to STAR Methods

Patient characteristics	IO+Chemo cohort (N = 113)	Chemo cohort (N = 58)
Age, years		
Median (range)	61 (30-74)	60 (37-74)
≤ 60, n (%)	52 (46.0)	30 (51.7)
> 60, n (%)	61 (54.0)	28 (48.3)
Gender, n (%)		
Male	87 (77.0)	45 (77.6)
Female	26 (23.0)	13 (22.4)
ECOG PS		
0	35 (31.0)	17 (29.3)
1	78 (69.0)	41 (70.7)
Smoking status		
Never smoker	42 (37.2)	16 (27.6)
Current or former smoker	71 (62.8)	42 (72.4)
Histology		
Adenocarcinoma	111 (98.2)	55 (94.8)
Other types	2 (1.8)	3 (5.2)
EGFR mutation status		
Positive	2 (1.8)	1 (1.7)
Negative	111 (98.2)	57 (98.3)
ALK translocation		
Positive	0 (0)	0 (0)
Negative	113 (100)	58 (100)
TNM stage,		
IIIB	7 (6.2)	4 (6.9)
IIIC	2 (1.8)	5 (8.6)
IV	104 (92.0)	49 (84.5)
PD-L1 TPS, %, n (%)		
< 1	31 (27.4)	19 (32.8)
1-49	33 (29.2)	13 (22.4)
≥ 50	49 (43.4)	26 (44.8)
Brain metastases, n (%)	12 (10.6)	11 (19.0)
TMErisk score		
Median (range)	0.00 (-1.68 – 1.93)	-2.00 (-3.56 – 1.16)

Abbreviations: ECOG, Eastern Cooperative Oncology Group; PS, performance status; EGFR, epidermal growth factor receptor； ALK，anaplastic lymphoma kinase；PD-L1, programmed cell death ligand 1; TPS, tumor proportion score; IO+Chemo, immunotherapy plus chemotherapy.

Table S14. Cut-points used in each dataset, related to STAR Methods

Datasets	C11orf96	C4BPA	PLAAT4	TGM2	KLF5	PNCK	TMErisk
TCGA-LUSC	5.952	23.190	51.288	72.246	73.092	5.914	0.27
TCGA-LUAD	19.323	16.440	15.327	93.926	63.064	0.060	-0.02
GSE81089	7.261	56.762	97.238	142.426	47.928	10.799	-0.38
GSE30219	9.348	7.942	9.022	8.105	8.346	5.949	0.00
GSE37745	9.192	10.207	8.509	8.302	10.701	6.521	-1.47
GSE157011	9.641	9.768	9.208	8.168	9.064	6.689	-0.68
GSE31210	6518.450	9749.500	3058.135	2259.950	1328.665	19.605	-1.35
GSE135222	10.570	0.390	119.120	38.630	46.380	5.750	0.67
GSE78220	24.298	0.021	NA	12.825	1.108	0.468	-2.15
OAK_LUSC_immunotherapy	3.000	0.569	4.682	5.881	8.701	2.369	-0.80
OAK_LUAD_immunotherapy	6.149	6.837	5.559	8.160	7.700	0.119	-0.34
POPLAR_immunotherapy	5.450	6.560	6.600	6.580	7.130	0.400	-0.10
IMvigor210_immunotherapy	9.425	1.535	97.605	40.615	682.025	3.905	0.06
OAK_LUSC_chemotherapy	2.480	0.380	5.383	5.550	6.460	1.590	-2.07
OAK_LUAD_chemotherapy	3.465	3.600	6.114	8.187	7.600	1.650	-0.61
POPLAR_chemotherapy	4.440	4.825	4.150	9.120	7.350	2.230	1.25
ORIENT-11_combo	105.955	89.240	55.889	409.449	161.353	2.110	0.03
ORIENT-11_chemotherapy	15.558	2.250	27.786	109.750	43.179	1.844	-2.04