Supplementary methods

Gene expression profiling

The RNA immune oncology (IO) panel was performed as previously described [1]. Briefly, total RNA was extracted from the FFPE samples and reverse transcribed, amplified, and ligated to fluorescent barcodes. After purification, the libraries were pooled and sequenced on the Ion S5 530 chip (Thermo Fisher Scientific, Waltham, MA). For each sample, we obtained 1-2 M reads to conduct data processing. For housekeeping (HK) gene normalization, ten HK genes were considered as endogenous controls. The absolute readout of each HK gene was compared against a predetermined HK reads per million (RPM) profile. The baseline HK RPM profile was established by measuring the average RPM of a number of replicates of GM12878 cell line samples across different sequencing runs. The fold-change ratio for each HK gene was calculated as follows: ratio of HK = absolute read count of HK/RPM profile of HK. The average value of all HK ratios was then applied as the normalization ratio for the pending sample: normalization ratio = median of all HK ratios. The normalized RPM (nRPM) of all genes of each sample was designated as follows: nRPM of sample S, gene G = absolute read count of sample S, gene G/normalization ratio of sample S.

Construction of a linear SVM classifier

In the data pre-processing step (online supplementary figure S1), we first

excluded the features with more than 80% NULL values, and filled in the missing data with the median value. All data were then subjected to log2 (raw data+1) transformation. For feature selection, briefly, we used multiple statistical methods to remove the noise features (online supplementary figure S1): (1) variance greater than or equal to 0.0005 and the top 80%; (2) mutual information between feature and binary label greater than or equal to 0.0005 and the top 90%; (3) AUC greater than or equal to 0.59; and (4) p values of five statistical tests (including chi-squared state, ANOVA F-value, t-test, Wilcoxon rank-sum test, and Mann-Whitney rank test) between two subgroups (DCB vs. NDB) less than or equal to 0.1. A certain feature was retained when it met the criteria for significance in 4 of the 8 statistical methods. After data standardization and normalization, recursive feature elimination cross-validation (RFECV) was performed to select features with the best accuracy score. We repeated the stratified shuffled-split cross-validator with 10 splitting iterations and a range of 0.15-0.34 test size to obtain the consensus features. The classic estimator, linear SVM, was implemented. We used the 13-fold cross-validation procedure (10 stratified-shuffle-split + 3 stratified-K-fold) to average the prediction error from the randomization of individuals to folds. In each fold, hyperparameter optimization was tuned by a cross-validation grid exhaustive search with a 0.3 test set. The prediction of the independent validation cohort was generated with the training-test set procedure, and the performance was evaluated by calculating the accuracy

and AUC. The analysis was conducted using Python 3.6, and code used for analyses is available at https://github.com/WellJoea/MLkit. We used the Youden index as the cut-off point (high IO-score > 0.52) to stratify patients with different prognoses and DCB rates.

References

1. Paluch BE, Glenn ST, Conroy JM, Papanicolau-Sengos A, Bshara W, Omilian AR, Brese E, Nesline M, Burgher B, Andreas J, et al: Robust detection of immune transcripts in FFPE samples using targeted RNA sequencing. Oncotarget. 2017;8:3197-3205.

Supplementary tables

Table S1. Baseline characteristics (N = 96)

Factors	Discovery cohort (N=72)	Validation cohort (N=24)	All (N=96)
Age			
median (range)	58 (15.0-76.0)	62.5 (26.0-75.0)	59.0 (15.0-76.0)
<=60, N (%)	42 (58.3)	11 (45.8)	53 (55.2)
>60, N (%)	30 (41.7)	13 (54.2)	43 (44.8)
Gender, N (%)			
Male	51 (70.8)	15 (62.5)	66 (68.8)
Female	21 (29.2)	9 (37.5)	30 (31.3)
Cancer type, N (%)			
Gastric cancer (GC)	29 (40.3)	9 (37.5)	38 (39.6)
Esophageal cancer (EC)	21 (29.2)	4 (16.7)	25 (26.0)
Colorectal cancer (CRC)	11 (15.3)	6 (25.0	17 (17.7)
Others	11 (15.3)	5 (20.8)	16 (16.7)
Histopathology, N (%)			
Adenocarcinoma	40 (55.6)	15 (62.5)	55 (57.3)
Squamous cell cancer	20 (27.8)	4 (16.7)	24 (25.0)
Neuroendocrine tumor	10 (13.9)	4 (16.7)	14 (14.6)
Others	2 (2.8)	1 (4.2)	3 (3.1)
Treatment option, N (%)			
Anti-PD-1 therapy	49 (68.1)[a]	16 (66.7)[b]	65 (67.7)
Anti-PD-L1 therapy	12 (16.7)	6 (25.0)	18 (18.8)
Anti-PD-1/PD-L1 + Anti-CTLA-4	11 (15.3)	2 (8.3)	13 (13.5)
Best response, N (%)			
Complete Response	0 (0.0)	1 (4.2)	1 (1.0)
Partial Response	14 (19.4)	7 (29.2)	21 (21.9)
Stable Disease	19 (26.4)	7 (29.2)	26 (27.1)
Progressive Disease	39 (54.2)	9 (37.5)	48 (50.0)
Response group, N (%)			
Durable clincial benefit (DCB)	25 (34.7)	9 (37.5)	34 (35.4)
No durable benefit (NDB)	47 (65.3)	15 (62.5)	62 (64.6)
MSI/MMR status, N (%)			
MSI-H/dMMR	11 (15.3)	7 (29.2)	18 (18.8)

MSS/MSI-L/pMMR	39 (54.2)	12 (50.0)	51 (53.1)
Not Available	22 (30.6)	5 (20.8)	27 (28.1)
PD-L1 expression, N (%)			
Positive	27 (37.5)	8 (33.3)	35 (36.5)
Negative	24 (33.3)	6 (25.0)	30 (31.3)
Not Available	21 (29.2)	10 (41.7)	31 (32.3)

[a] two patients received PD-1 inhibitor plus chemotherapy, [b] one patient received PD-1 inhibitor plus chemotherapy, and one patient received PD-1 inhibitor plus apatinib. Abbreviations: CTLA-4, cytotoxic T-lymphocyte–associated protein 4; PD-1, programmed cell death 1; PD-L1, Program death ligand 1; MSI-H, Microsatellite Instability high; MSI-L, Microsatellite Instability low ; MSS, Microsatellite Stability; dMMR, deficient-Mismatch Repair; pMMR, proficient-Mismatch Repair.

Table S2. Functional characterization of the 24 immune oncology-related genes (IO-score signature)

Gene	Gene Discriptive	Pathway	Function
IDO1	Indoleamine 23-dioxygenase 1	Checkpoint pathway	Checkpoint pathway
CCL22	C-C motif chemokine ligand 22	Chemokine signaling	Cytokine signaling
IL13	Interleukin 13	Cytokine signaling	Cytokine signaling
TNFSF9	Tumor necrosis factor superfamily member 9	Cytokine signaling	Cytokine signaling
IFITM1	Interferon induced transmembrane protein 1	Type I interferon signaling	Cytokine signaling
IFITM2	Interferon induced transmembrane protein 2	Type I interferon signaling	Cytokine signaling
STAT1	Signal transducer and activator of transcription 1	Type II interferon signaling	Cytokine signaling
IL1B	Interleukin 1 beta	Type II interferon signaling	Cytokine signaling
TAP1	Transporter 1, ATP-binding cassette subfamily B member	Type II interferon signaling	Cytokine signaling
NRP1	Neuropilin 1	Dendritic cell	Lymphocyte marker
STAT6	Signal transducer and activator of transcription 6	Helper T cell	Lymphocyte marker
CD163	CD163 molecule	Macrophage	Lymphocyte marker
KREMEN1	Kringle containing transmembrane protein 1	Neutrophil	Lymphocyte marker
VCAM1	Vascular cell adhesion molecule 1	Leukocyte migration	Lymphocyte regulation
CCL2	C-C motif chemokine ligand 2	Lymphocyte infiltrate	Lymphocyte regulation
LAPTM5	Lysosomal protein transmembrane 5	Lymphocyte infiltrate	Lymphocyte regulation
M6PR	Mannose-6-phosphate receptor cation dependent	T cell regulation	Lymphocyte regulation
BAGE	B melanoma antigen	Tumor antigen	Tumor characterization
MAGEA3	MAGE family member A3	Tumor antigen	Tumor characterization
MLANA	Melan-A	Tumor antigen	Tumor characterization
BRCA2	BRCA2 DNA repair associated	Tumor suppressor	Tumor characterization
CDKN2A	Cyclin-dependent kinase inhibitor 2A	Tumor suppressor	Tumor characterization
EFNA4	Ephrin A4	Oncogene	Tumor characterization
PTEN	Phosphatase and tensin homolog	Tumor suppressor	Tumor characterization

Table S3. Discriminative performance of IO-score in discovery and validation cohorts

Cohort	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Discovery	0.97	0.94	0.92	0.96	0.92	0.96
Validation	0.74	0.83	0.67	0.93	0.86	0.82

AUC, area under the receiver operating characteristic curve; PPV, positive predicetive value; NPV, negative predicetive value.

Table S4. Association between durable clinical benefit and candidate biomarkers

Characteristics	No. of patients	Response		*P* value	OR value (95% CI)
		NDB (%)	DCB (%)		
MSI/MMR status (combined cohort)	69				
MSI-H/dMMR	18	8 (44.4)	10 (55.6)	0.10	2.50 (0.84-7.49)
MSS/MSI-L/pMMR	51	34 (66.7)	17 (33.3)		
PD-L1 status[a] (combined cohort)	65				
PD-L1(-)	30	20 (66.7)	10 (33.3)	0.75	1.18 (0.43-3.29)
PD-L1(+)	35	22 (62.9)	13 (37.1)		
TMB level[b] (combined cohort)	90				
TMB-low	57	43 (75.4)	14 (24.6)	0.002	4.17 (1.67-10.43)
TMB-high	33	14 (42.4)	19 (57.6)		
IO-score (combined cohort)	96				
IO-score low	64	59 (92.2)	5 (7.8)	< 0.001	114.07 (25.48-510.59)
IO-score high	32	3 (9.4)	29 (90.6)		
IO-score (discovery cohort)	72				
IO-score low	47	45 (95.7)	2 (4.3)	< 0.001	258.75 (34.21-1957.04)
IO-score high	25	2 (8.0)	23 (92.0)		
IO-score (validation cohort)	24				
IO-score low	17	14 (82.4)	3 (17.6)	0.004[c]	28.00 (2.40-326.74)
IO-score high	7	1 (14.3)	6 (85.7)		

[a]PD-L1 positive was defined as >=1% of the tumor/stromal cell membrane staining.

[b]TMB-low group was defined as TMB < 5 mutations/mb and TMB-high group was defined as >= 5 mutations/mb.

[c]Fisher's exact test.

PD-L1, program death ligand 1; MSI-H, Microsatellite Instability high; MSI-L, Microsatellite Instability low ; MSS, Microsatellite Stability; dMMR, deficient-Mismatch Repair; pMMR, proficient-Mismatch Repair; TMB-high, Tumor mutation burden high; TMB-L, Tumor mutation burden low; OR, odds ratio.

Supplementary figures

Figure S1. Flow diagram for machine learning.

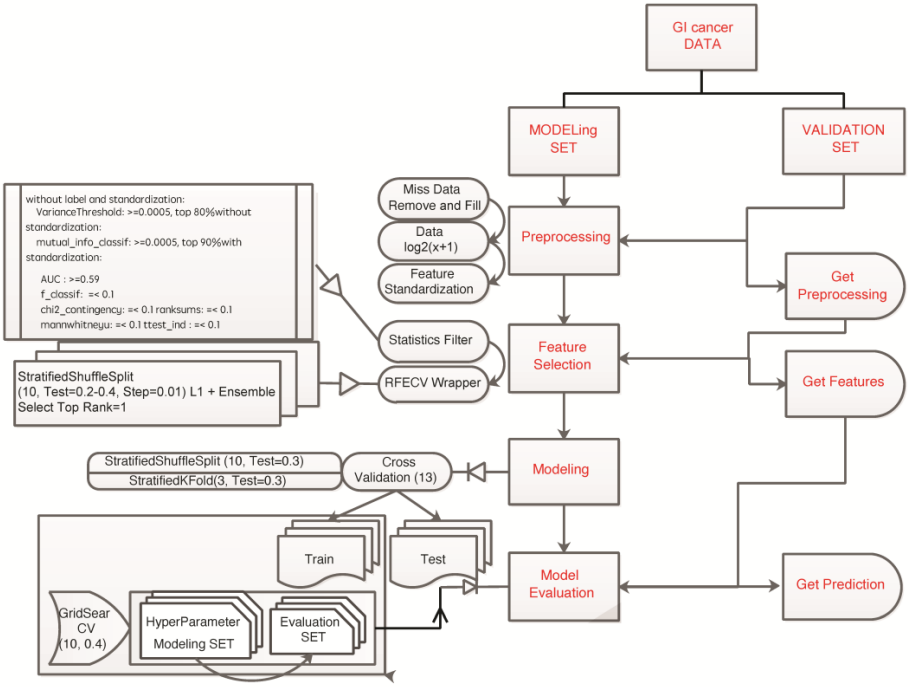

Figure S2. Predictive and prognostic value of the TMB level in the combined cohort (N=90). A. ROC curve of TMB in predicting clinical benefit. B. Comparison of the DCB rates between the TMB-high and TMB-low groups. C and D, Kaplan–Meier curves comparing PFS (C) and OS (D) between the TMB-high and TMB-low subgroups.

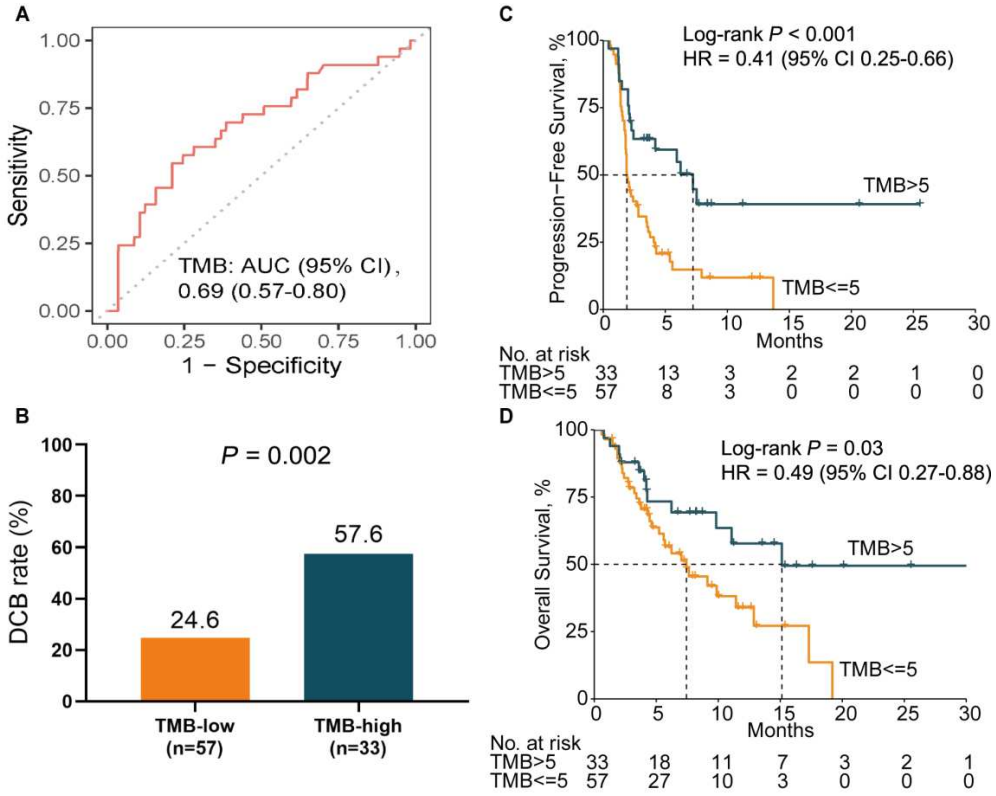

Figure S3. Predictive and prognostic values of MSI/MMR status in the combined cohort (N=69). A. ROC curve of MSI-H/dMMR in predicting clinical benefit. B. Comparison of the DCB rates between the MSI-H/dMMR and MSS/MSI-L/pMMR groups. C and D, Kaplan–Meier curves comparing progression-free survival (C) and overall survival (D) between the MSI-H/dMMR and MSS/MSI-L/pMMR subgroups.

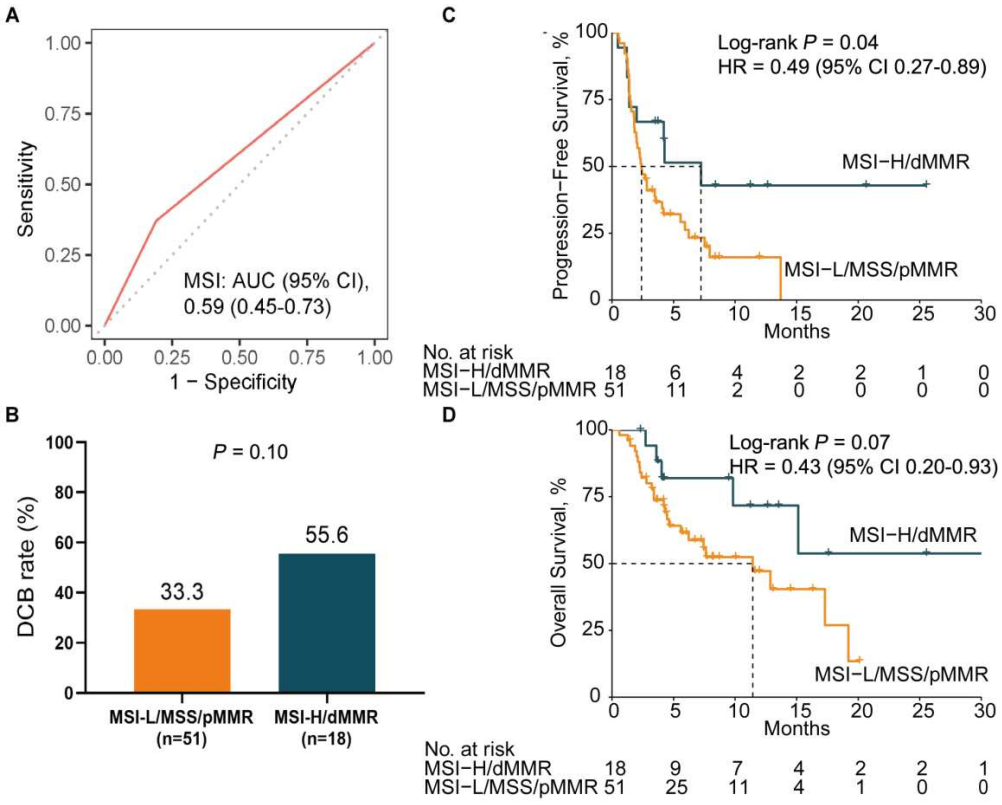

Figure S4. Predictive and prognostic value of PD-L1 expression in the combined cohort (N=65). A. ROC curve of PD-L1 expression in predicting clinical benefit. B. Comparison of the DCB rates between the PD-L1-positive and -negative groups. C and D, Kaplan–Meier curves comparing PFS (C) and OS (D) between the PD-L1-positive and-negative subgroups.

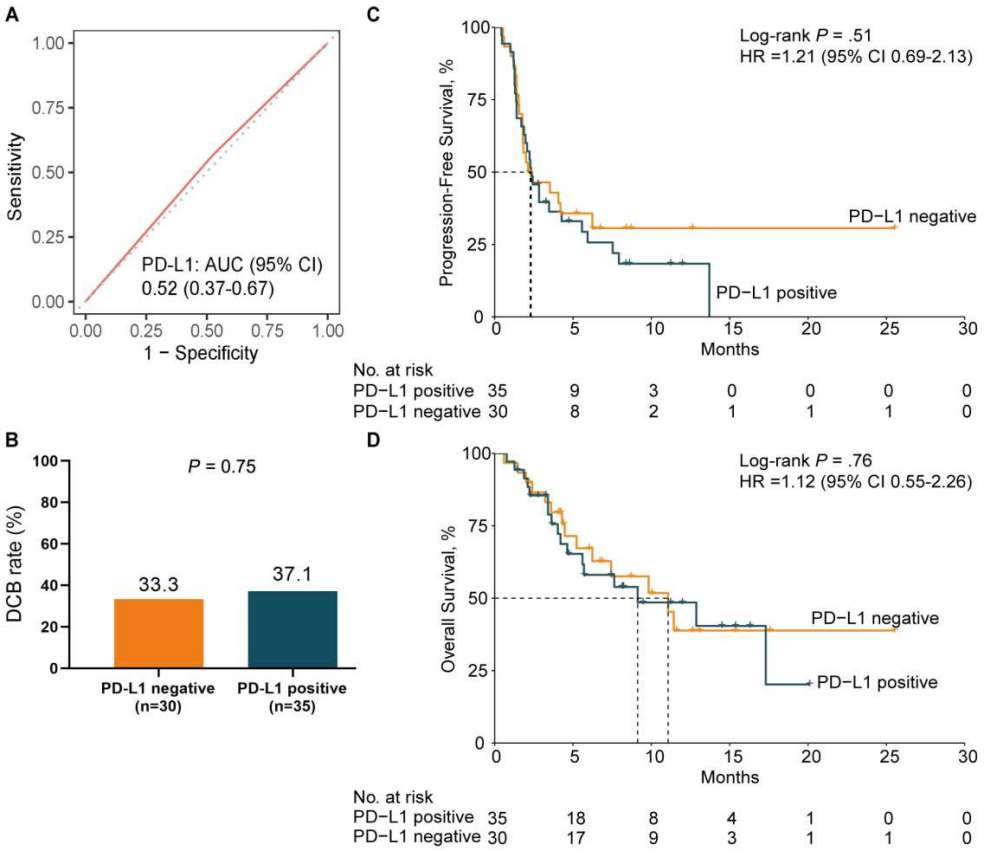

Figure S5. Venn diagram depicting the correlations among the IO-score, TMB, MSI/MMR status and PD-L1 expression. A, Venn diagram showing the correlations among IO-score high, TMB-high, MSI-H/dMMR and PD-L1 positivity. B, Distributions of the TMB level, MSI/MMR status and PD-L1 expression in the IO-score-high subgroup. C, Venn diagram showing the correlations among IO-score low, TMB-high, MSI-H/dMMR and PD-L1 positivity. D, Distributions of the TMB level, MSI/MMR status and PD-L1 expression in the IO-score-low subgroup.

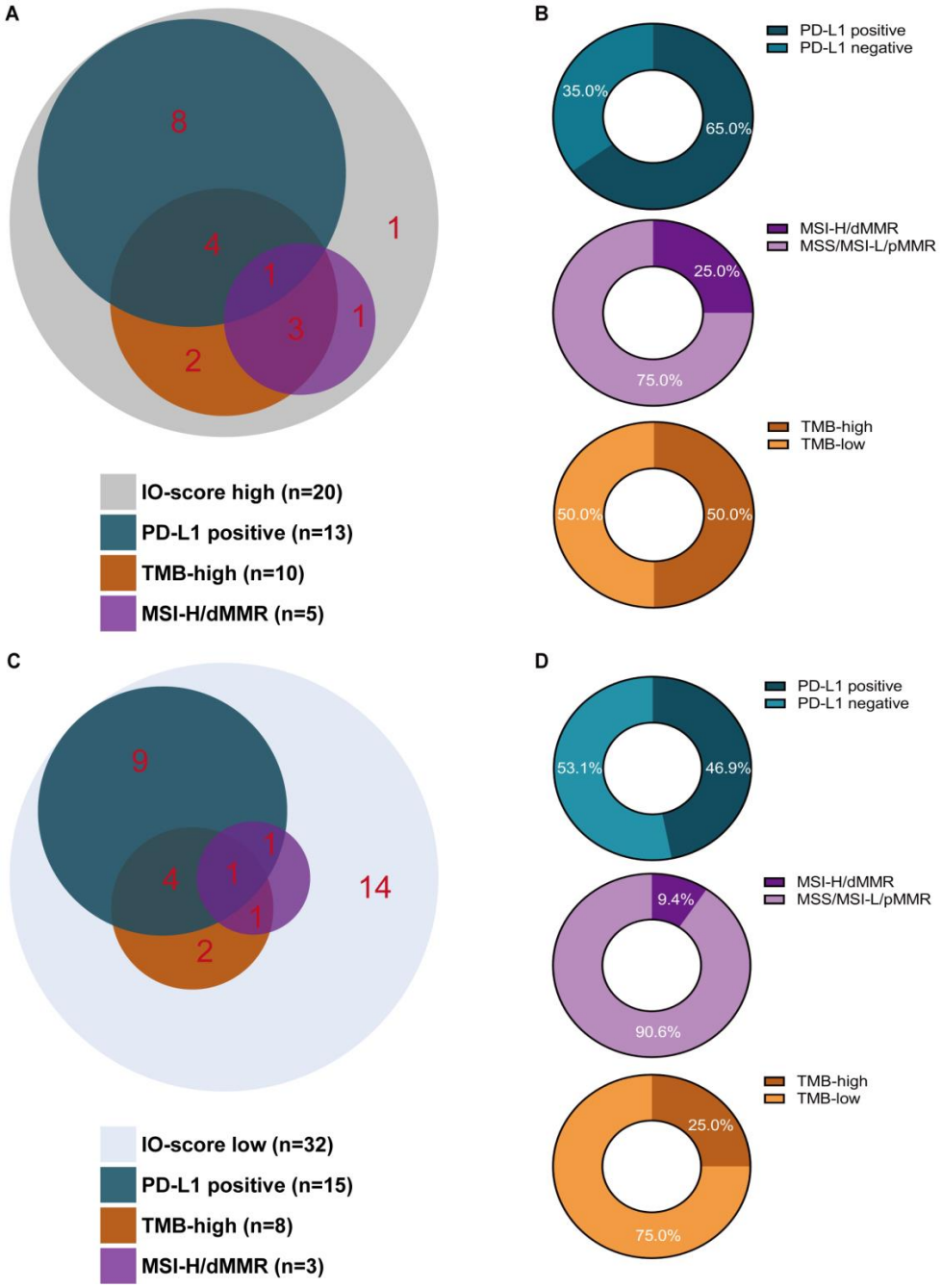
A
8
4
1
1
3
1
2
IO-score high (n=20)
PD-L1 positive (n=13)
TMB-high (n=10)
MSI-H/dMMR (n=5)
B
PD-L1 positive
PD-L1 negative
35.0%
65.0%
MSI-H/dMMR
MSS/MSI-L/pMMR
25.0%
75.0%
TMB-high
TMB-low
50.0%
50.0%
C
9
4
1
1
1
14
2
IO-score low (n=32)
PD-L1 positive (n=15)
TMB-high (n=8)
MSI-H/dMMR (n=3)
D
PD-L1 positive
PD-L1 negative
53.1%
46.9%
MSI-H/dMMR
MSS/MSI-L/pMMR
9.4%
90.6%
TMB-high
TMB-low
25.0%
75.0%

Figure S6. The IO-score gene expression signature highlights the complex tumor and immune microenvironment.