

Development and Validation of Artificial Intelligence-Based Model for Bladder Cancer Immunophenotyping Using Whole Slide Images

Supplementary Information

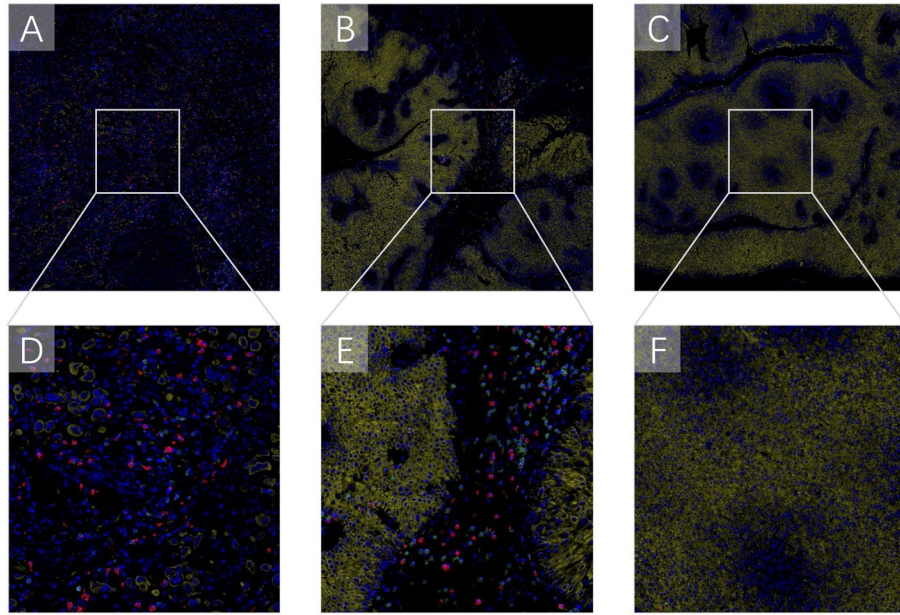


Figure S1. Immunofluorescence staining. Panels A–C show the desert, excluded, and inflammatory immune phenotypes, respectively. Panels D–F present the corresponding magnified views. CD8 is shown in red, CD3 in green, PanCK in yellow, and DAPI in blue.

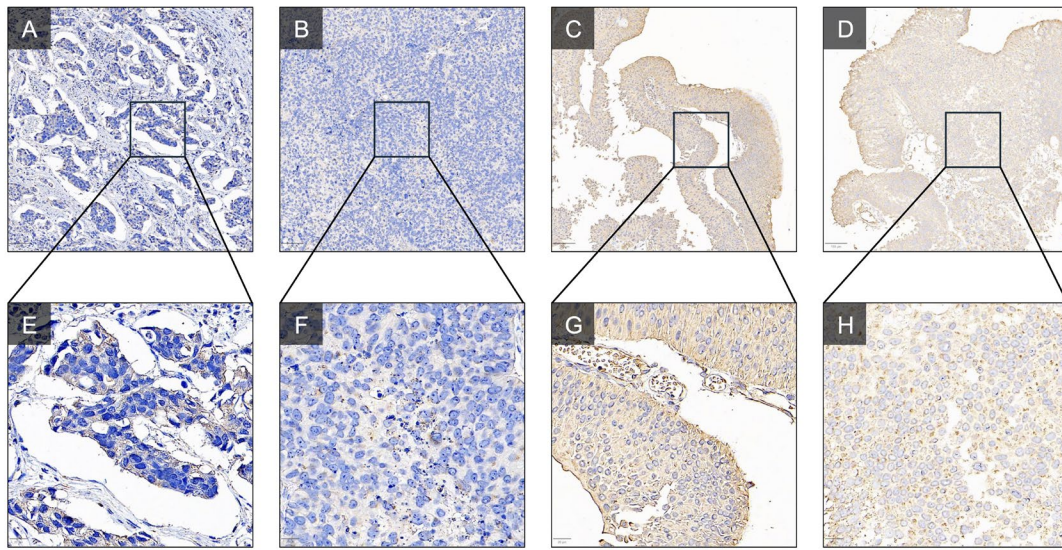


Figure S2. Representative PD-L1 immunohistochemistry staining in selected patients. Panels A and B show PD-L1–negative tumors (TPS < 1%), while panels C and D show PD-L1–positive tumors (TPS ≥ 1%). Panels E–H present the corresponding magnified views of these regions.

	RHWU training cohort	RHWU validation cohort	PHHC validation cohort	Additional validation cohort
Immunophenotype				
Desert	310 (31.90%)	78 (32.10%)	48 (24.00%)	39 (32.50%)
Excluded	331 (34.05%)	82 (33.74%)	88 (44.00%)	39 (32.50%)
Inflamed	331 (34.05%)	83 (34.16%)	64 (32.00%)	42 (35.00%)

Table S1. Distribution of immunephenotypes.

Phenotyping	num	Response	Non-response	Effective Rate
Inflamed	40	29	11	72.5%
Excluded	27	7	20	25.9
Desert	21	4	17	19.0%
Total	88	40	48	45.5%

Table S2 (a). Contingency Table of Immunophenotype versus Immunotherapy Response in the RHWU Cohort. $\chi^2=21.86027$ $p<0.005$.

Phenotyping	num	Control	Progress	Control Rate
Inflamed	40	34	6	85.0%
Excluded	27	9	18	33.3%
Desert	21	5	16	23.8%
Total	88	48	40	54.5%

Table S2 (b). Contingency Table of Immunophenotype versus Immunotherapy Disease Control Rate in the RHWU Cohort. $\chi^2=27.86492$ $p<0.005$.

Phenotyping	num	Response	Non-response	Effective Rate
Inflamed	10	8	2	80.0%
Excluded	8	2	6	25.5%
Desert	2	0	2	0.0%
Total	20	10	10	50.0%

Table S3 (a). Contingency Table of Immunophenotype versus Immunotherapy Response in the PHHC Cohort. $\chi^2=7.6$ $p<0.05$.

Phenotyping	num	Control	Progress	Control Rate
Inflamed	10	9	1	90.0%
Excluded	8	3	5	37.5%
Desert	2	0	2	0.0%
Total	20	12	8	60.0%

Table S3 (b). Contingency Table of Immunophenotype versus Immunotherapy Disease Control Rate in the PHHC Cohort. $\chi^2=8.4$ $p<0.05$.

Model	Variable	Estimate	Std. Error	z value	p value	OR	95% CI for OR	AIC
MIPDAS	(Intercept)	0.9694	0.3541	2.738	0.006	2.64	1.36 – 5.52	104.41
	Inflamed							
	Excluded	-2.0192	0.5641	-3.579	<0.001	0.13	0.04 – 0.38	
	Desert	-2.4163	0.6590	-3.667	<0.001	0.09	0.02 – 0.30	
TPS	TPS	0.0978	0.0708	1.382	0.167	1.10	0.35 – 4.96	123.32
	TPS.result	-1.5675	0.5291	-2.962	0.003	0.21	0.02 – 0.67	115.20

Table S4 (a). Logistic Regression Analysis of Immunophenotype-Based Prediction of ICI Treatment Response in the RHWU Cohort.

Model	Variable	Estimate	Std. Error	z value	p value	OR	95% CI for OR	AIC
MIPDAS	(Intercept)	1.7346	0.4428	3.9173	<0.001	5.6667	2.56 – 15.00	97.24
	Inflamed							
	Excluded	-2.4277	0.6023	-4.0309	<0.001	0.0882	0.03 – 0.27	
	Desert	-2.8978	0.6772	-4.2791	<0.001	0.0551	0.01 – 0.19	
TPS	TPS	-0.0094	0.0696	-0.1350	0.893	0.99	0.86 – 1.74	125.25
	TPS.result	-0.4796	0.4606	-1.041	0.298	0.62	0.25 – 1.52	124.18

Table S4 (b). Logistic Regression Analysis of Immunophenotype-Based Prediction of Disease Control Rate in the RHWU Cohort.

Model	Variable	Estimate	Std. Error	z value	p value	OR	95% CI for OR	AIC
MIPDAS	(Intercept)	1.3863	0.7906	1.754	0.0795	4.00	1.00 – 26.5	25.005
	Inflamed							
	Excluded	-2.4849	1.1365	-2.186	0.0288	0.08	0.01 – 0.65	
	Desert	-18.9524	2797.4420	-0.007	0.9946	~0.00	NA –2.65e+181	
TPS	TPS	1.1854	0.5177	2.290	0.0220	3.27	1.49 – 11.95	18.424
	TPS.result	-3.5835	1.3176	-2.720	0.0065	0.03	0.00 – 0.26	20.710

Table S5 (a). Logistic Regression Analysis of Immunophenotype-Based Prediction of ICI Treatment Response in the PHHC Cohort.

Model	Variable	Estimate	Std. Error	z value	p value	OR	95% CI for OR	AIC
MIPDAS	(Intercept)	2.1972	1.0541	2.0845	<0.001	9.0000	1.69 – 165.98	23.09
	Inflamed							
	Excluded	-2.7081	1.2824	-2.1118	<0.001	0.0667	0.00 – 0.62	
	Desert	-19.7633	2797.4421	-0.0071	<0.001	0.0000	NA –5.97e+180	
TPS	TPS	0.6719	0.3598	1.867	0.0618	1.96	1.16 – 29.52	24.19
	TPS.result	-2.1972	1.0541	-2.084	0.0371	0.11	0.01 – 0.77	25.89

Table S5 (b). Logistic Regression Analysis of Immunophenotype-Based Prediction of Disease Control Rate in the PHHC Cohort.

RHWU Cohort

For the Immunophenotype-based evaluation system, the Inflamed phenotype was used as the reference group. The Excluded phenotype demonstrated an odds ratio (OR) of 0.13 (95% CI: 0.04–0.38), indicating that patients with the Excluded phenotype had approximately an 87% lower likelihood of responding to ICI therapy compared to those with the Inflamed phenotype. Similarly, the Desert phenotype had an OR of 0.09 (95% CI: 0.02–0.30), suggesting a 91% reduction in the probability of treatment response. The 95% confidence intervals for both phenotypes did not include 1, confirming the robustness of these associations. The model achieved an AIC value of 104.41 and converged after four Fisher scoring iterations, indicating good overall model fit.

Furthermore, when evaluating disease control rate as the outcome, the MIPDAS immunophenotype demonstrated even stronger predictive performance. Compared to the Inflamed phenotype (control rate: 85.0%), the Excluded phenotype (control rate: 33.3%) showed an OR of 0.088 (95% CI: 0.03–0.27, $p < 0.001$), indicating an approximately 91.2% reduction in the likelihood of achieving disease control. The Desert phenotype (control rate: 23.8%) exhibited an OR of 0.055 (95% CI: 0.01–0.19, $p < 0.001$), corresponding to a 94.5% reduction in disease control probability. The model achieved a superior AIC value of 97.24, suggesting better fit for predicting disease control compared to response prediction.

These findings demonstrate that the MIPDAS immunophenotyping classification serves as a strong predictive factor for ICI therapeutic response. Compared with the Inflamed phenotype, patients with Excluded and Desert phenotypes showed significantly reduced response probabilities (by approximately 87% and 91%, respectively), with highly significant p-values. This classification system substantially enhances the predictive performance of the logistic regression model and provides valuable guidance for clinical decision-making in immunotherapy.

For the PD-L1 TPS-based evaluation system, higher TPS values were associated with an increasing trend in ICI treatment response probability (Estimate > 0). The TPS coefficient was positive (Estimate = 0.0978), suggesting that each one-unit increase in TPS corresponded to approximately a 10% increase in the odds of response (OR = 1.103). However, this association did not reach statistical significance ($p = 0.167$), and the model fit was moderate (AIC = 123.32), indicating that continuous TPS alone is a weak predictor of immunotherapy response.

In contrast, when TPS was analyzed as a categorical variable (TPS.result, where 1 = TPS $< 1\%$), the coefficient was -1.57 , implying that patients with low TPS had a markedly lower response probability than those with high TPS (OR ≈ 0.21). This association was statistically significant, indicating that the categorical TPS classification provides greater explanatory power for predicting ICI response compared to the continuous TPS measure.

For disease control rate prediction, both continuous TPS and categorical TPS.result showed no significant associations. Continuous TPS yielded an OR of 0.99 (95% CI: 0.86–1.74, $p = 0.893$), indicating no predictive value for disease control. The categorical TPS.result model showed a non-significant OR of 0.62 (95% CI: 0.25–1.52, $p = 0.298$). The AIC values for these models (125.25 and 124.18, respectively) were substantially higher than that of the MIPDAS model (97.24), confirming the superior predictive capability of immunophenotyping for disease control outcomes.

PHHC Cohort

For the Immunophenotype model, the Excluded phenotype had a regression coefficient of -2.48 (OR = 0.083, $p = 0.0288$), indicating that patients with an Excluded immune phenotype were

approximately 91.7% less likely to respond to ICI treatment compared with the reference (Inflamed) group, a statistically significant finding. The Desert phenotype showed an extremely large coefficient and standard error, likely due to the limited sample size or a complete separation issue, rendering the result unreliable ($p = 0.9946$).

When analyzing disease control rate in the PHHC cohort, the Inflamed phenotype demonstrated a high control rate of 90.0% (OR = 9.00, 95% CI: 1.69–165.98). The Excluded phenotype (control rate: 37.5%) showed a significant reduction in disease control probability (OR = 0.067, 95% CI: 0.00–0.62, $p < 0.001$). Similar to the response analysis, the Desert phenotype estimates remained unstable due to limited sample size (control rate: 0.0%). The model AIC was 23.09, indicating good fit.

For the TPS-based system, the regression coefficient for continuous TPS was 1.1854 ($p = 0.022$), demonstrating a significant positive association between TPS and treatment response. Each one-unit increase in TPS was associated with an approximately 3.27-fold increase in the odds of achieving an ICI response (OR = 3.27). When TPS was treated as a categorical variable, the coefficient was -3.5835 , indicating that the low-TPS group had a substantially lower probability of response compared to the high-TPS group (OR ≈ 0.03). However, given the relatively small cohort size, these results should be interpreted with caution as potential sampling bias may exist.

For disease control rate prediction, continuous TPS showed a borderline significant association (OR = 1.96, 95% CI: 1.16–29.52, $p = 0.0618$), suggesting a trend toward improved disease control with higher TPS values. The categorical TPS.result model demonstrated a significant association (OR = 0.11, 95% CI: 0.01–0.77, $p = 0.0371$), indicating that patients with TPS < 1% had an 89% lower probability of achieving disease control compared to those with TPS $\geq 1\%$. The AIC values were comparable across models (MIPDAS: 23.09, continuous TPS: 24.19, categorical TPS: 25.89).

In summary, the MIPDAS immunophenotyping system consistently demonstrated superior predictive performance for both treatment response and disease control rate across both independent cohorts. The Inflamed phenotype was associated with the highest response and control rates (72.5%/85.0% in RHWU; 80.0%/90.0% in PHHC), while Excluded and Desert phenotypes showed significantly reduced probabilities. Notably, the immunophenotype-based models achieved lower AIC values compared to TPS-based models, indicating better model fit. The TPS system showed variable predictive value: continuous TPS significantly predicted response in the PHHC cohort but not in RHWU, while categorical TPS.result demonstrated more consistent associations. These findings support the clinical utility of immunophenotyping for patient stratification in immunotherapy, particularly for identifying patients most likely to achieve disease control.

Feature	Gain	Cover	Frequency
I-I	0.2245869	0.18719904	0.05097841
I/T	0.1384624	0.14471914	0.13290761
Graph_T-I_Nsubgraph	0.13453324	0.06555532	0.10167625
Graph_I-I_Coreness	0.1203844	0.1378774	0.05922843
T-S	0.11029768	0.07605825	0.19465273
Graph_T-I_AuthorityScore	0.10735274	0.09450872	0.14586714
Graph_T-I_Eigenvector	0.08273674	0.15054778	0.13084347
Graph_I-I_Betweenness	0.0532794	0.12205879	0.1424971
T-T	0.0283665	0.02147555	0.04134886

Table S6 XGBoost feature importance ranking. The XGBoost model was trained on LASSO-selected features using five-fold cross-validation and grid search to optimize hyperparameters. The final configuration was: n_estimators = 300, max_depth = 5, learning_rate = 0.05, subsample = 0.8, colsample_bytree = 0.8, gamma = 0.1, min_child_weight = 3.

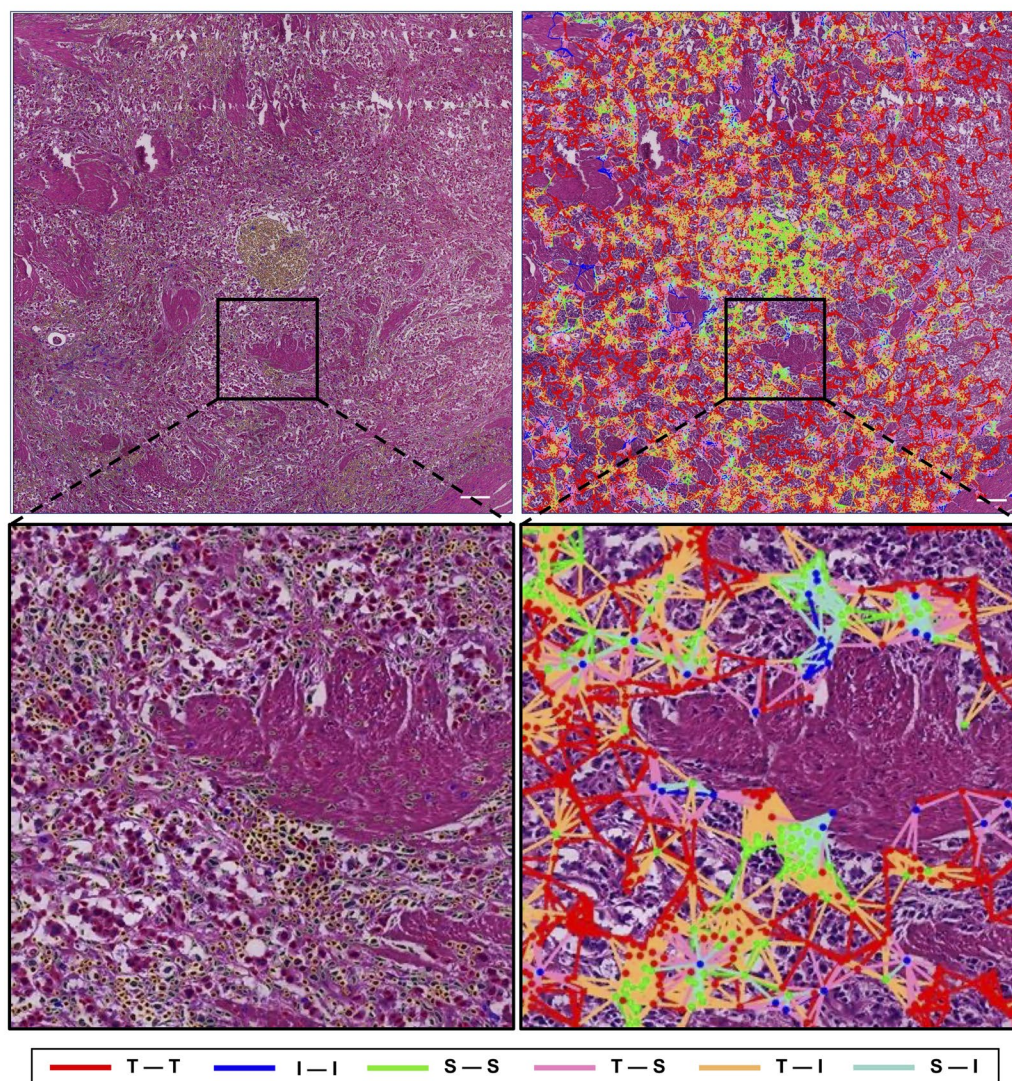


Figure S3. Enlarged results of cell graph networks construction. (Left) Hover-Net segmentation and classification results. (Right) cell graph networks construction.

We evaluated the cell classification accuracy of HoVer-Net across 25 randomly selected regions based on MxIF and paired H&E staining. The model demonstrated high performance with an overall accuracy of 0.90. Class-specific accuracy rates were 0.89 for tumor cells, 0.93 for immune cells, and 0.92 for stromal cells, confirming robust cell-type discrimination capability.

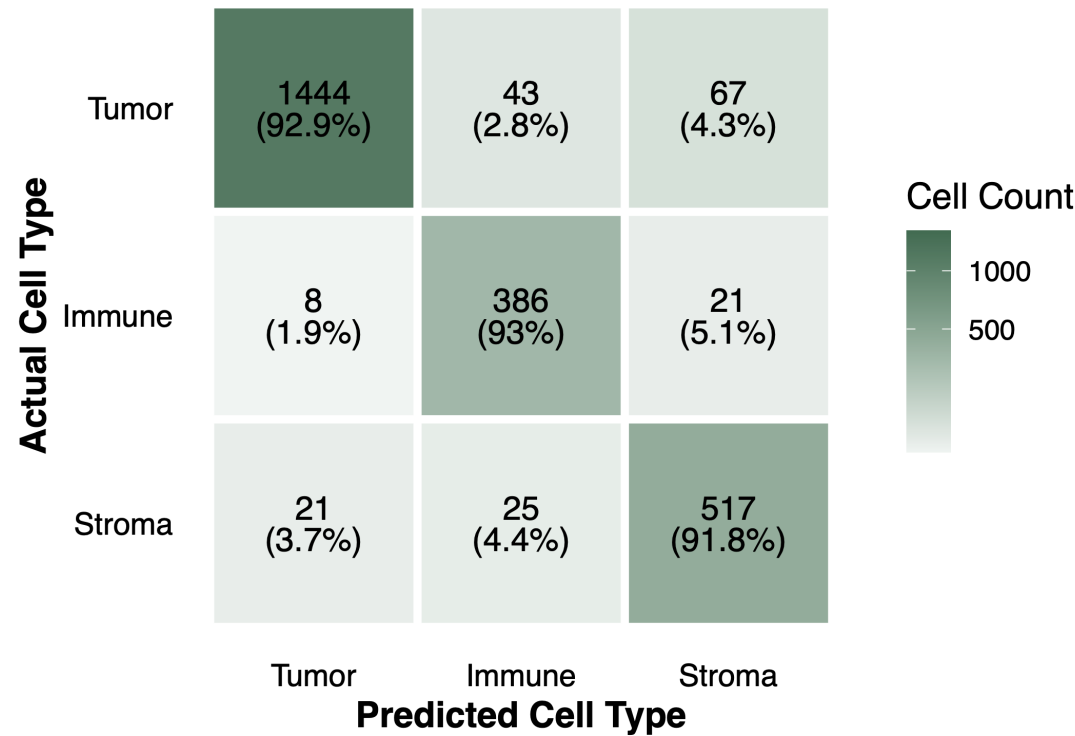


Figure S4. Cell recognition accuracy confusion matrix of 25 selected regions.

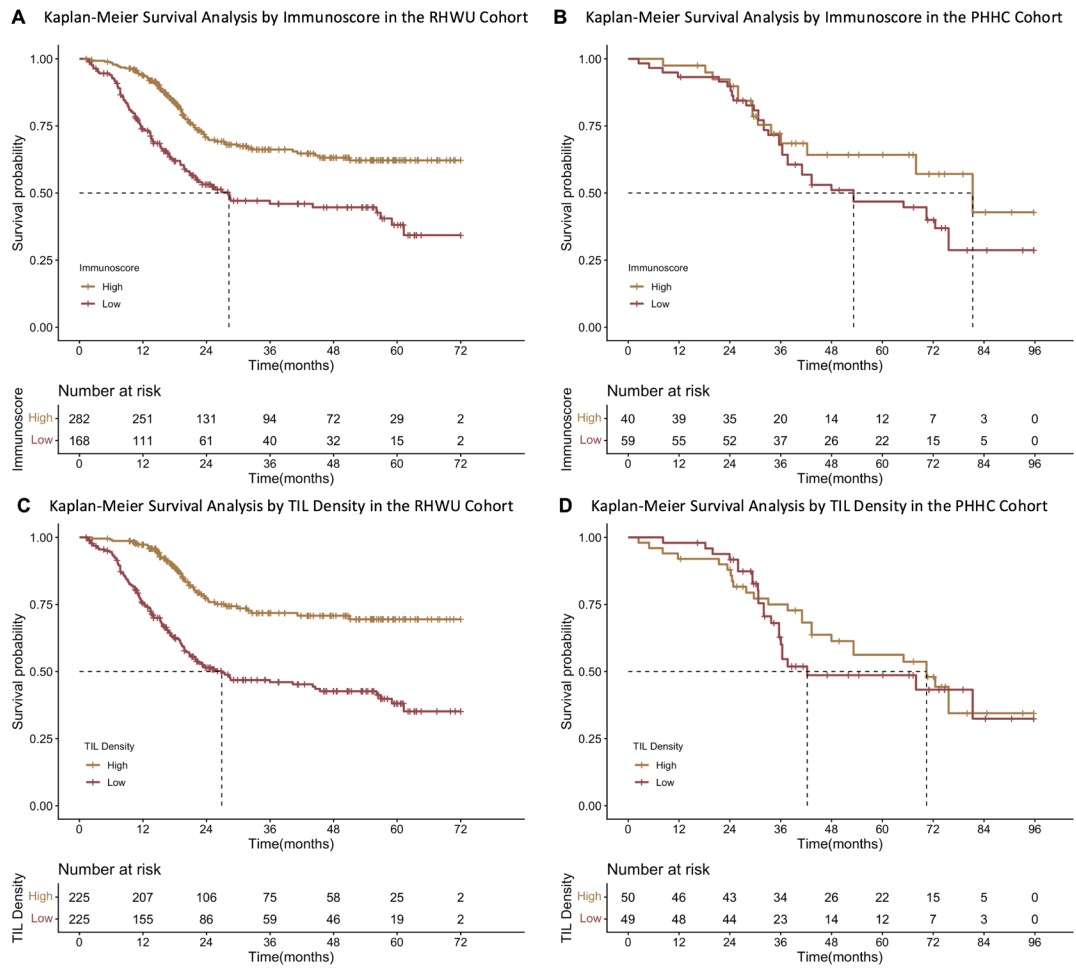


Figure S5. Prognostic Stratification by the Immunoscore and TIL Density Systems in the RHWU and PHHC Cohorts.

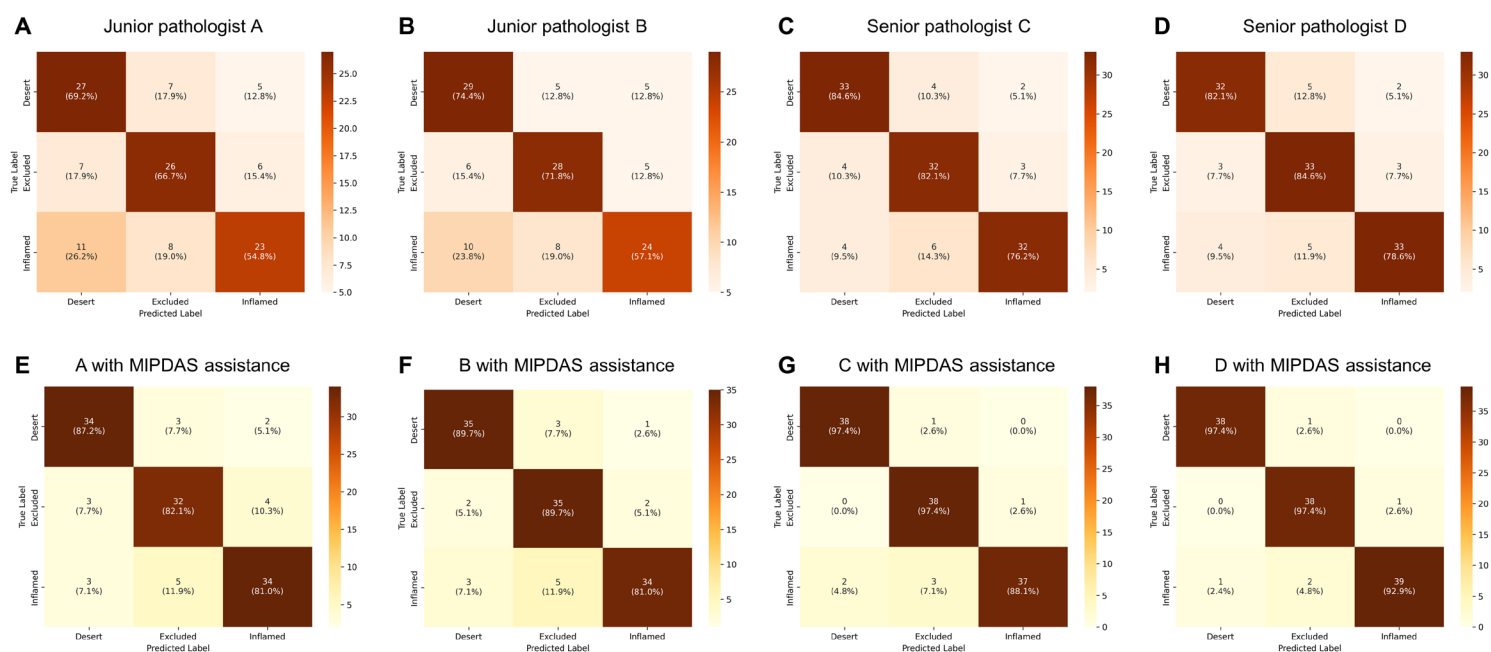


Figure S6. Confusion matrix for each pathologist with and without AI assistance. (A-D) Pathologists without MIPDAS assistance. (E-H) Pathologists with MIPDAS assistance.

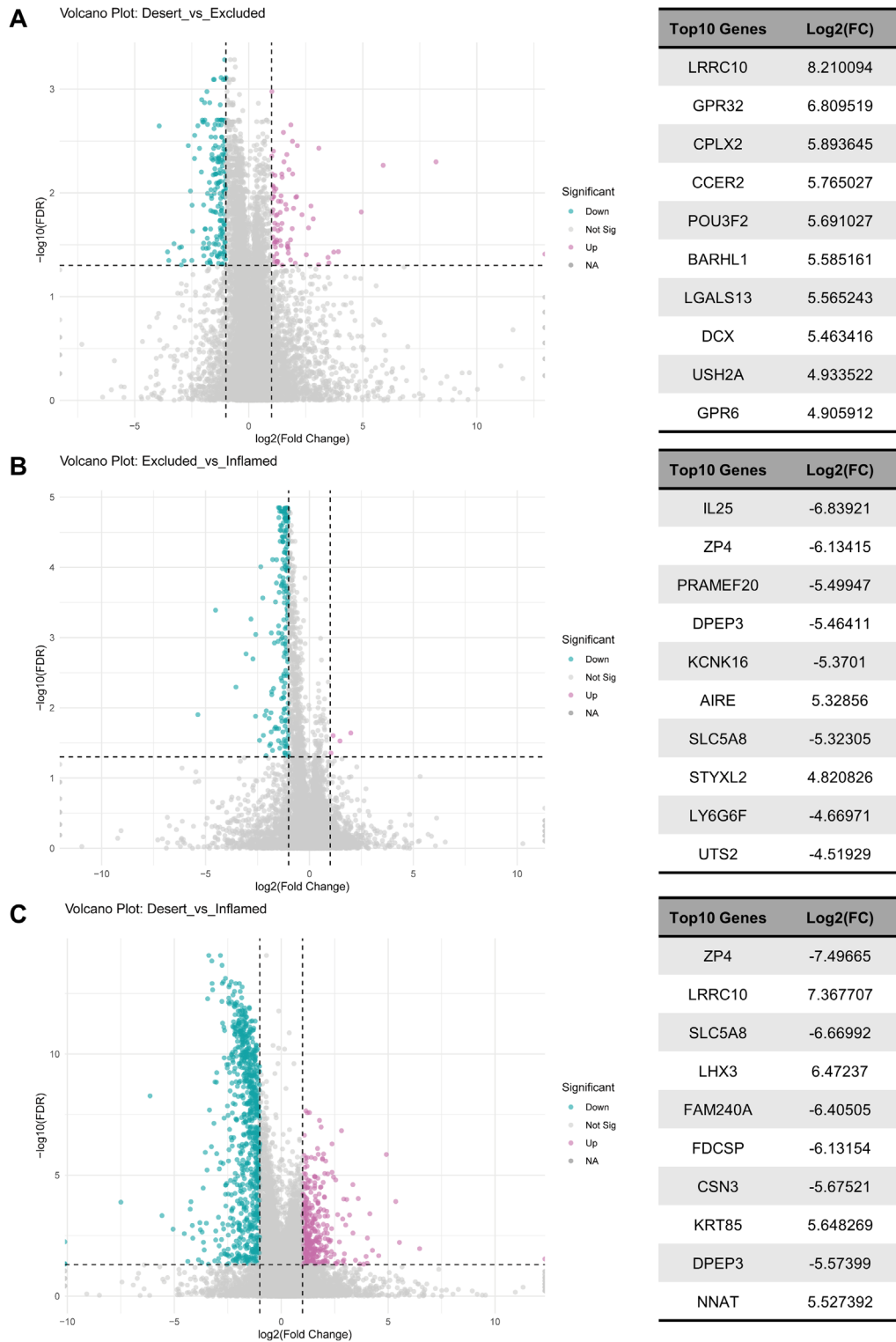


Figure S7. Volcano plots for MIPDAS-predicted Immunophenotypes. (A) Desert vs Excluded. (B) Excluded vs Inflamed. (C) Desert vs Inflamed.

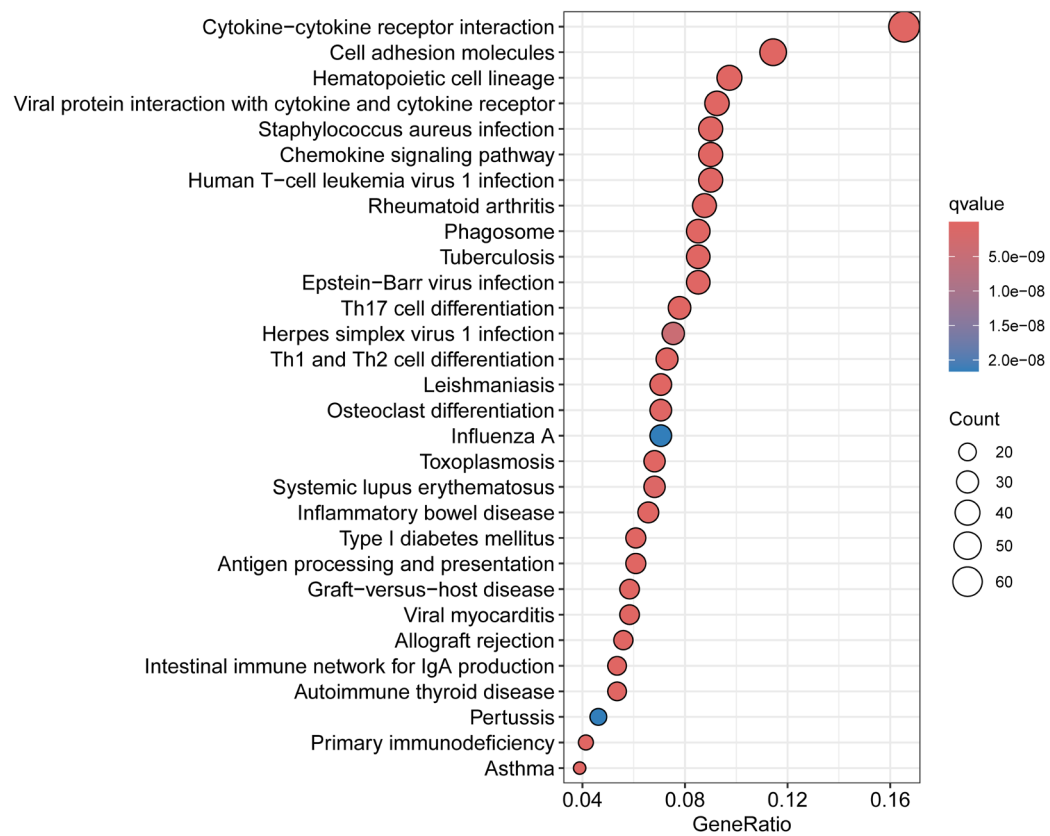


Figure S8. Kyoto Encyclopedia of Genes and Genomes analysis for Desert vs Inflamed type.

Primary antibody	Source organism	Manufacturer	Catalogue number	Dilution ratio	Corresponding color
CD3	Rabbit	Abcam	Ab16669	1:1000	Purple
CD20	Rabbit	Abcam	Ab64088	1:200	Yellow
aSMA	Rabbit	Proteintech	14395-1-AP	1:2000	Blue
Pan-CK	Mouse	Thermo	53-9003-82	1:200	Green

Table S7. Antibodies used for Multiplex immunofluorescence.

Primary antibody	Source organism	Manufacturer	Catalogue number	Dilution ratio	Corresponding color
CD3	Mouse	Servicebio	GB12014	1:2000	Green
CD8	Mouse	Servicebio	GB54196	1:5000	Red
Pan-CK	Mouse	Servicebio	GB122053	1:5000	Yellow

Table S8. Antibodies used for Immunofluorescence.

ONTOLOGY	ID	Description	Count
Biological Process	GO:0002443	leukocyte mediated immunity	144
Biological Process	GO:0002449	lymphocyte mediated immunity	128
Biological Process	GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	126
Biological Process	GO:0019724	B cell mediated immunity	85
Biological Process	GO:0016064	immunoglobulin mediated immune response	84
Biological Process	GO:0002377	immunoglobulin production	74
Cellular Component	GO:0019814	immunoglobulin complex	119
Cellular Component	GO:0042101	T cell receptor complex	59
Cellular Component	GO:0098802	plasma membrane signaling receptor complex	72
Cellular Component	GO:0009897	external side of plasma membrane	77
Cellular Component	GO:0042611	MHC protein complex	18
Cellular Component	GO:0072562	blood microparticle	37
Molecular Function	GO:0003823	antigen binding	112
Molecular Function	GO:0140375	immune receptor activity	48
Molecular Function	GO:0042605	peptide antigen binding	27
Molecular Function	GO:0005125	cytokine activity	49
Molecular Function	GO:0008009	chemokine activity	23
Molecular Function	GO:0042379	chemokine receptor binding	26

Table S9. Enrichment results of GO pathway.

Feature name	Category	Description	Algorithm
Nsubgraph	Topological	The number of nuclei in the subgraph where the nucleus is located.	<code>igraph.Graph().vcount()</code>
Degrees	Topological	The number of edges connected with the nucleus.	<code>igraph.Graph().degree()</code>
Coreness	Topological	The coreness of a nucleus is k if it is a member of the k-core but not a member of the (k+1)-core.	<code>igraph.Graph().coreness()</code>
Eccentricity	Topological	The maximum of the shortest distance from this nucleus to all other nuclei in the graph.	<code>igraph.Graph().eccentricity()</code>
Eccentricity_normed	Topological	The normalized value of the eccentricity with the number of the nuclei in the subgraph.	<code>igraph.Graph().eccentricity()/n</code>
Harmonic Centrality	Topological	The harmonic centrality of a nucleus measures how easily other nuclei can be reached from it. It is computed by the mean inverse distance to all other nuclei.	<code>igraph.Graph().harmonic_centrality()</code>
Closeness	Topological	The closeness centrality of a nucleus measures how easily other nuclei can be reached from it. It is computed by the number of nuclei minus one divided by the sum of the lengths of all geodesics from/to the given nucleus.	<code>igraph.Graph().closeness()</code>
Betweenness	Topological	The betweenness is how important role the nucleus plays in the graph. It is computed by the ratio of the number of the shortest paths that pass the nucleus over the number of paths that pass the nucleus in the subgraph.	<code>igraph.Graph().betweenness()</code>
Betweenness_normed	Topological	The normalized value of the betweenness with the number of the nuclei in the subgraph.	<code>igraph.Graph().betweenness()/((n-1)*(n-2)/2)</code>
Clustering Coefficient	Topological	The clustering coefficient is the probability that two neighbors of a nucleus are connected and form a loop.	<code>igraph.Graph().transitivity_local_undirected()</code>
MinEdgeLength	Topological	The minimum length of the edges of the nucleus. If there is no edge connected with the nucleus, the value is set as 100 pixels, which is the upper limit of distance between two connected nuclei.	Calculated according to the description
MeanEdgeLength	Topological	The mean length of the edges of the nucleus. If there is no edge connected with the nucleus, the value is set as 100 pixels, which is the upper limit of distance between two connected nuclei.	Calculated according to the description
StromaBlocker	Topological	For each inflammatory nucleus and its closest tumor nucleus, calculate the number of stroma nucleus whose distance to these two nuclei are both smaller than the distance between this inflammatory nucleus and its closest tumor nucleus.	Implemented according to https://github.com/popellab/MIBC-Predictive-models

Table S10. The definition of extracted topological features.

Name
T-T
I-I
T-S
S-S
T-I
I-S
S_number
T_number
I_number
I/T
I/S
Graph_I-I_Eigenvector
Graph_I-I_Degrees
Graph_I-I_AuthorityScore
Graph_I-I_Coreness
Graph_I-I_Diversity
Graph_I-I_HubScore
Graph_I-I_NeighborhoodSize
Graph_I-I_Strength
Graph_I-I_ClusteringCoefficient
Graph_I-I_Eccentricity
Graph_I-I_Eccentricity_normed
Graph_I-I_HarmonicCentrality
Graph_I-I_Closeness
Graph_I-I_Betweenness
Graph_I-I_Betweenness_normed
Graph_T-I_Nsubgraph
Graph_T-I_Eigenvector
Graph_T-I_Degrees
Graph_T-I_AuthorityScore
Graph_T-I_Coreness
Graph_T-I_Diversity
Graph_T-I_HubScore
Graph_T-I_NeighborhoodSize
Graph_T-I_Strength
Graph_T-I_Eccentricity
Graph_T-I_Eccentricity_normed
Graph_T-I_HarmonicCentrality
Graph_T-I_Closeness
Graph_T-I_Betweenness
Graph_T-I_Betweenness_normed

Graph_I-S_Nsubgraph
 Graph_I-S_Eigenvector
 Graph_I-S_Degrees
 Graph_I-S_AuthorityScore
 Graph_I-S_Coreness
 Graph_I-S_Diversity
 Graph_I-S_HubScore
 Graph_I-S_NeighborhoodSize
 Graph_I-S_Strength
 Graph_I-S_Eccentricity
 Graph_I-S_Eccentricity_normed
 Graph_I-S_HarmonicCentrality
 Graph_I-S_Closeness
 Graph_I-S_Betweenness
 Graph_I-S_Betweenness_normed
 stromaBlocker

Table S11. The details of 57 topological spatial features.