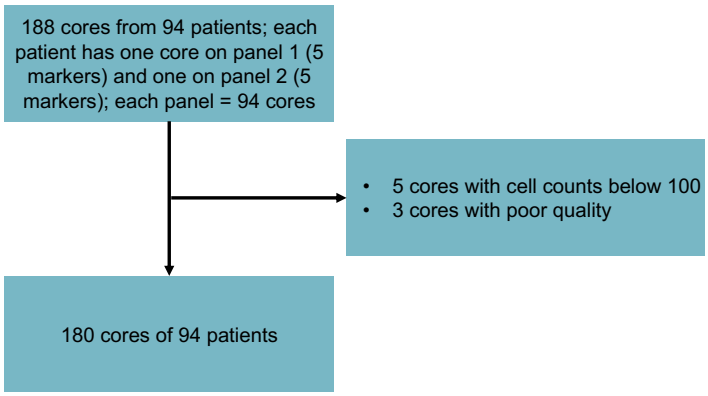
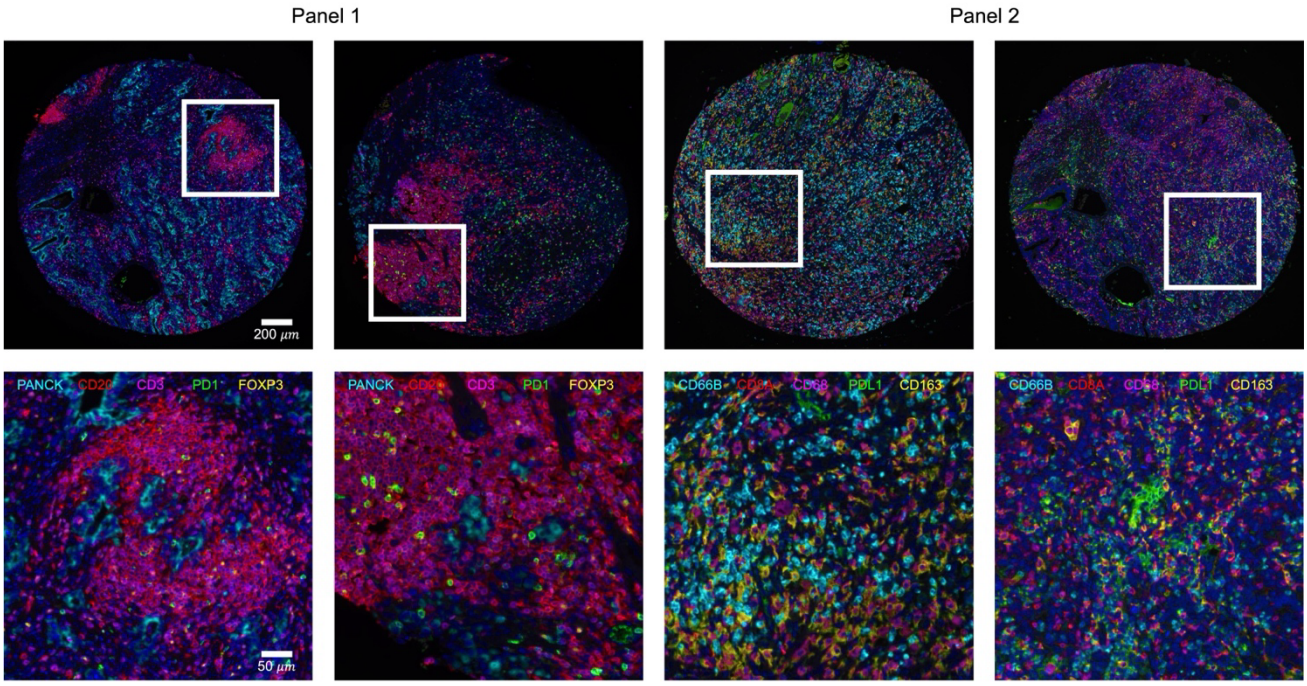
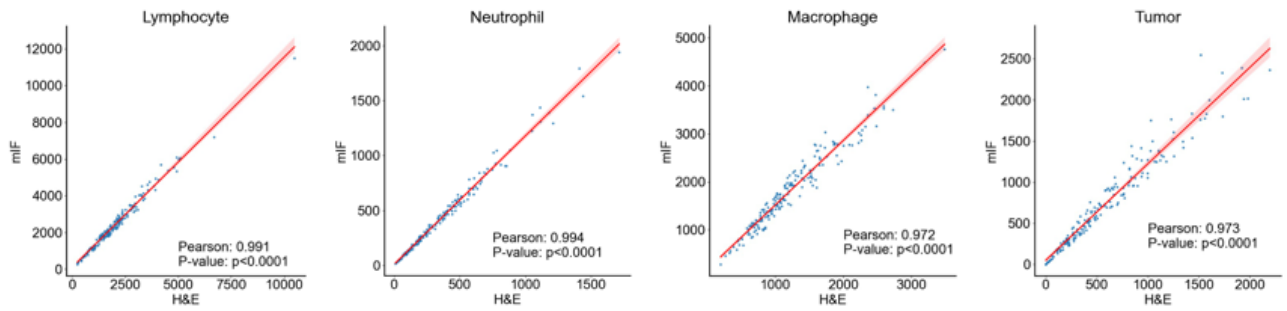
SUPPLEMENTARY FIGURES



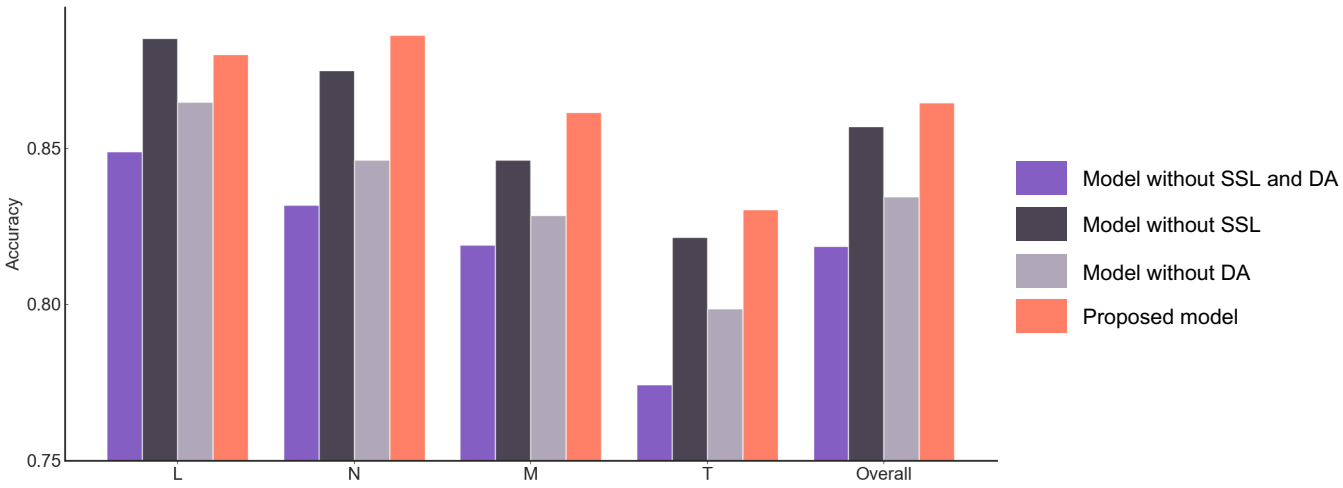
Supplementary Fig. 1 | Flow chart of sample inclusion and exclusion on 1st TMA.



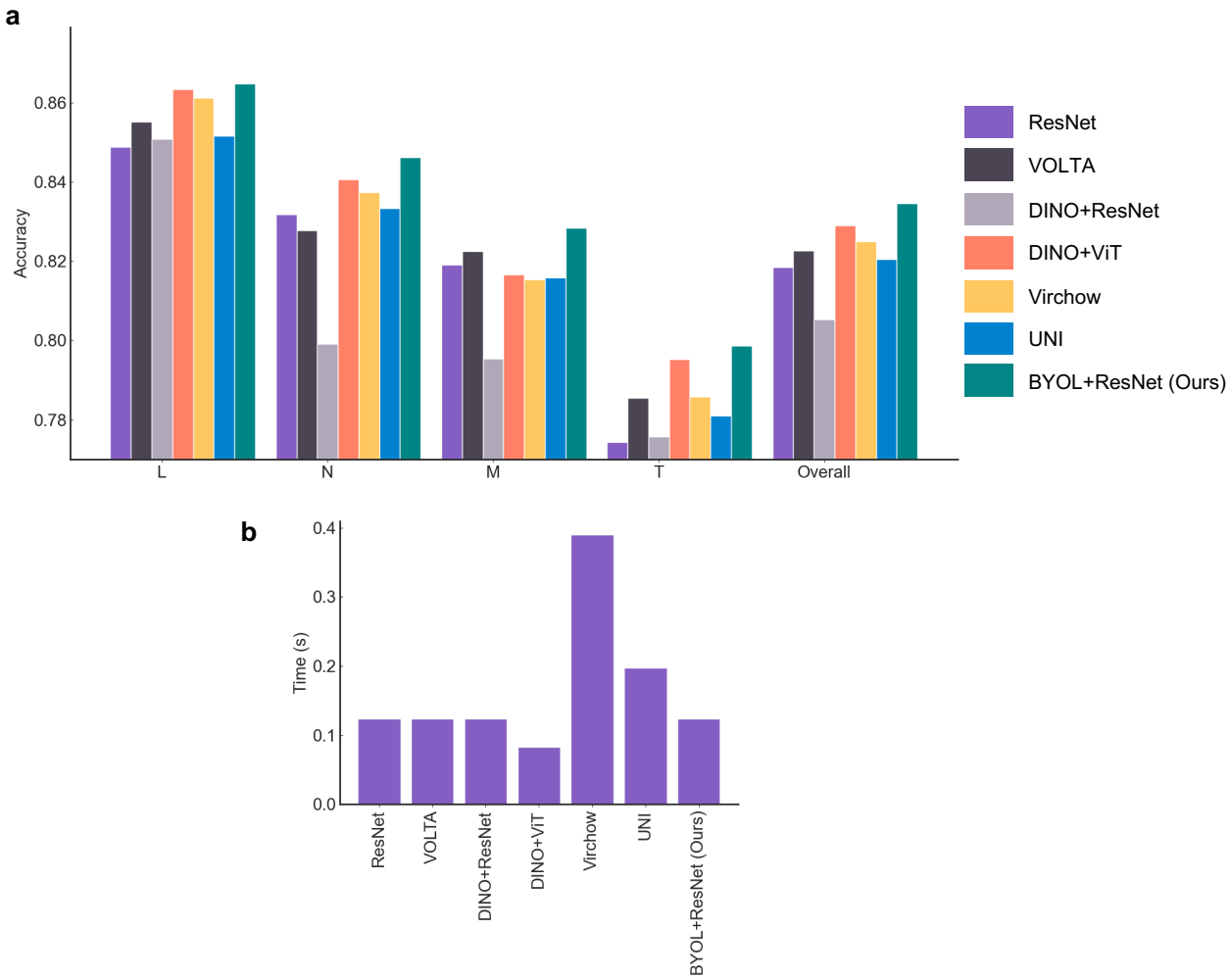
Supplementary Fig. 2 | Example mIF images of TMA cores. Panel 1 includes five protein markers: CD3, CD20, pan-CK, PD1, and Foxp3. Panel 2 includes another five markers: CD66b, CD68, CD8a, PD-L1, and CD163.



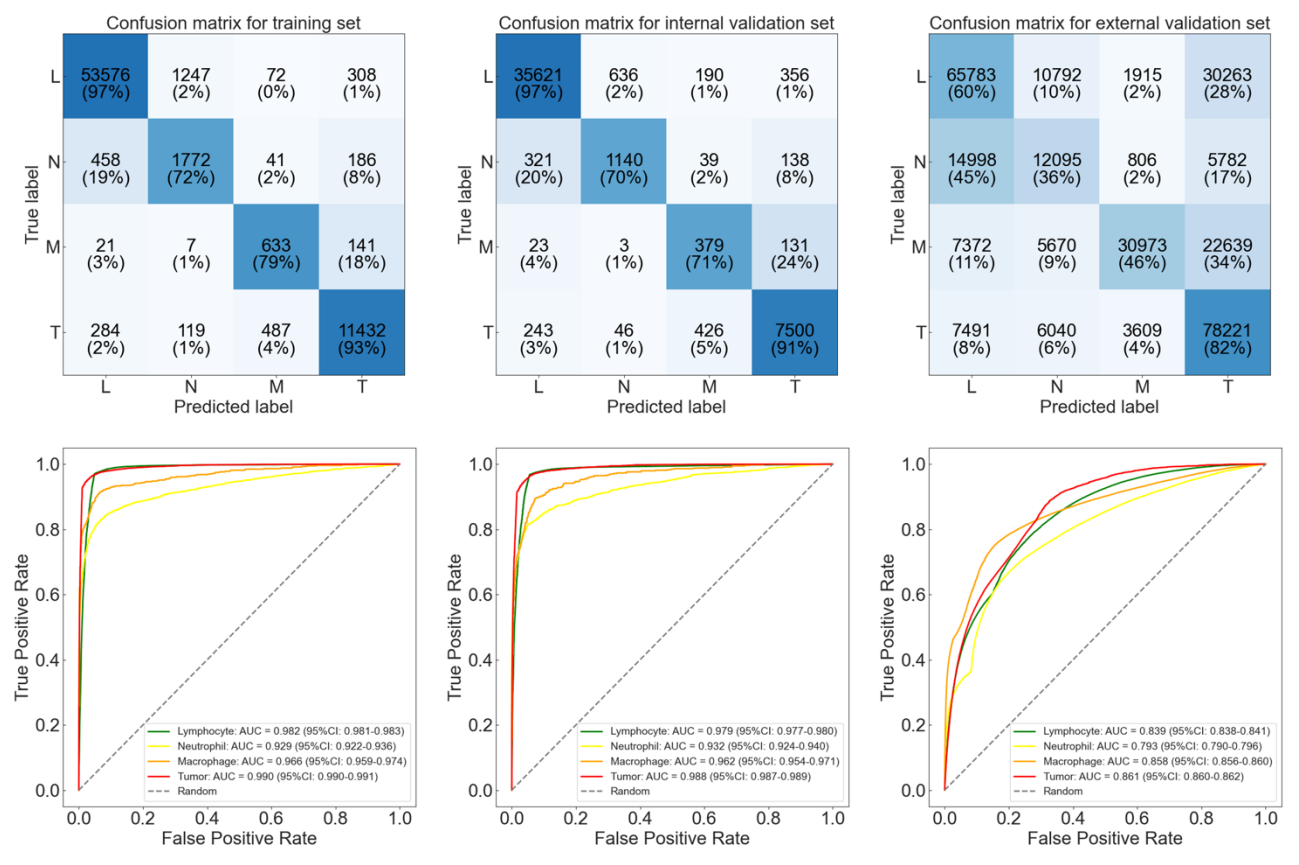
Supplementary Fig. 3 | Scatter plot of the cell counts between the model's output based on H&E images and the ground truth obtained from mIF images across four cell types for each TMA core.



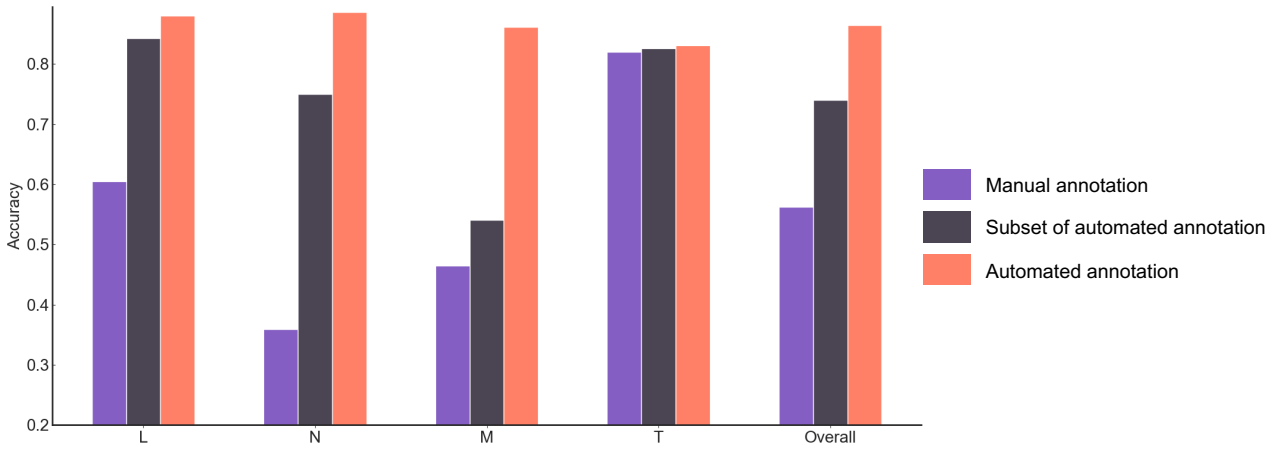
Supplementary Fig. 4 | Performance of cell classification models with and without self-supervised learning (SSL) or domain adaptation (DA) on H&E images. Accuracy of cell classification model in the external validation set. L, lymphocyte; N, neutrophil; M, macrophage; T, tumor cell. Source data are provided as a Source Data file.



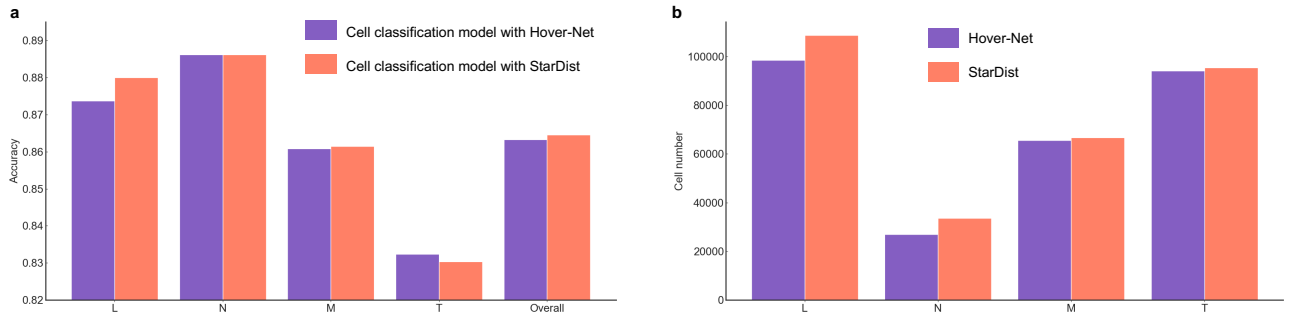
Supplementary Fig. 5 | Performance comparison of cell classification models using the proposed method and state-of-the-art methods. a, Accuracy of cell classification model in the external validation set (2nd TMA). All models were trained without domain adaptation. L, lymphocyte; N, neutrophil; M, macrophage; T, tumor cell. **b**, Average inference time per cell classification in the external validation set. Source data are provided as a Source Data file.



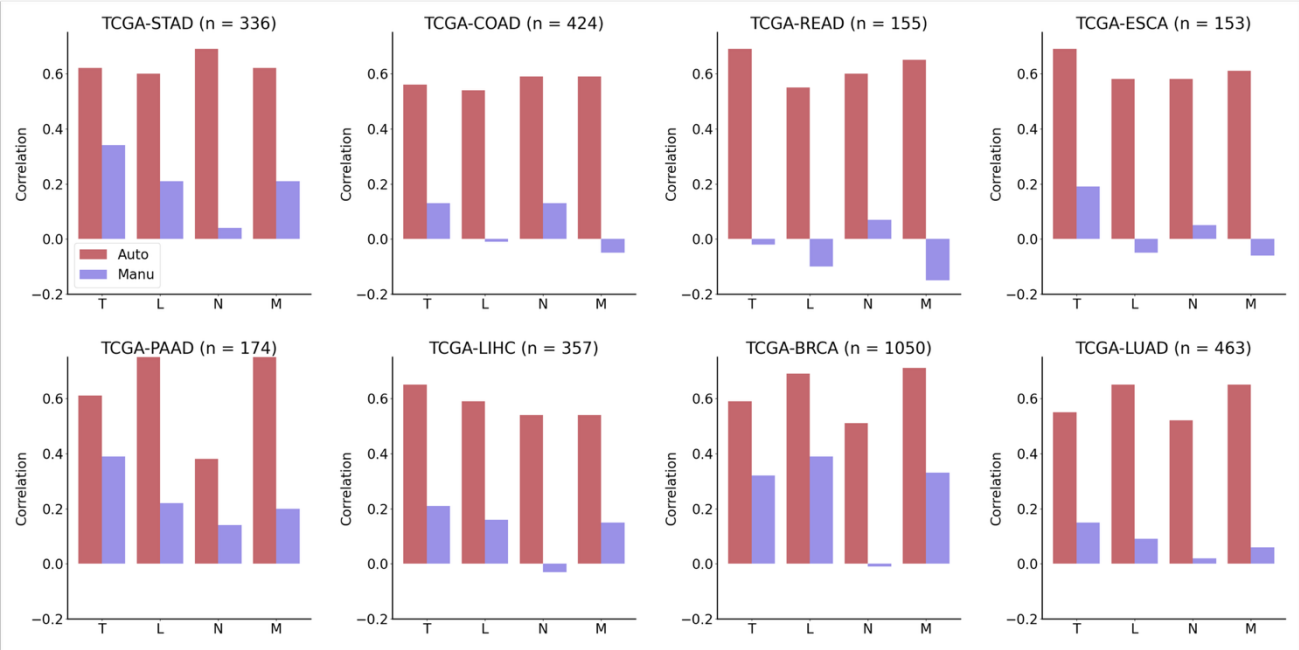
Supplementary Fig. 6 | Performance of cell classification model using traditional manual annotation approach on H&E images. a, Confusion matrix of cell classification model in the training, internal validation, and external validation sets. L, lymphocyte; N, neutrophil; M, macrophage; T, tumor cell. **b,** The ROC curves and AUCs of cell classification model in the training, internal validation, and external validation sets. Source data are provided as a Source Data file.



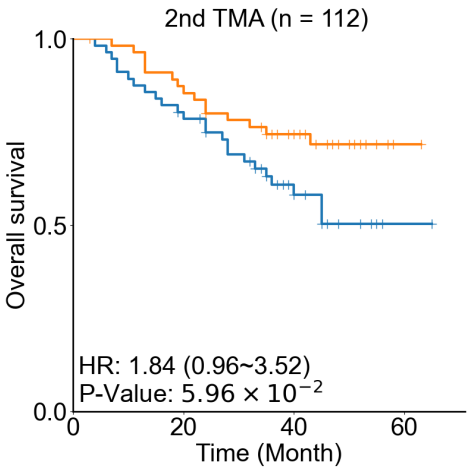
Supplementary Fig. 7 | Performance comparison of cell classification models training by automated or traditional manual cell annotation dataset. Accuracy of cell classification model in the external validation set (2nd TMA). L, lymphocyte; N, neutrophil; M, macrophage; T, tumor cell. Source data are provided as a Source Data file.



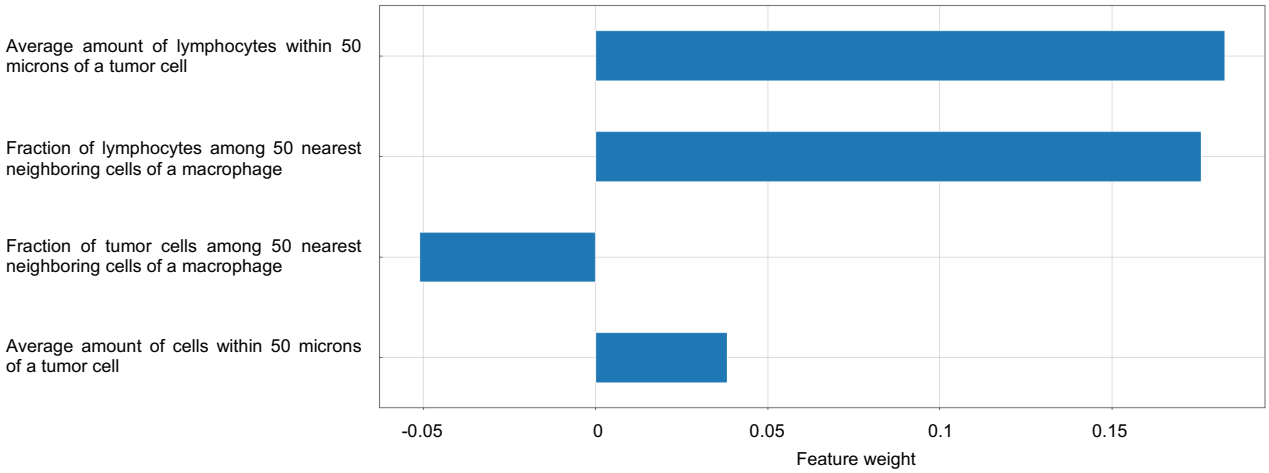
Supplementary Fig. 8 | Performance comparison of cell classification models using different cell segmentation methods. a, Accuracy of cell classification model in the external validation set (2nd TMA). L, lymphocyte; N, neutrophil; M, macrophage; T, tumor cell. **b**, Number of cells by applying cell segmentation method to the external validation set. Source data are provided as a Source Data file.



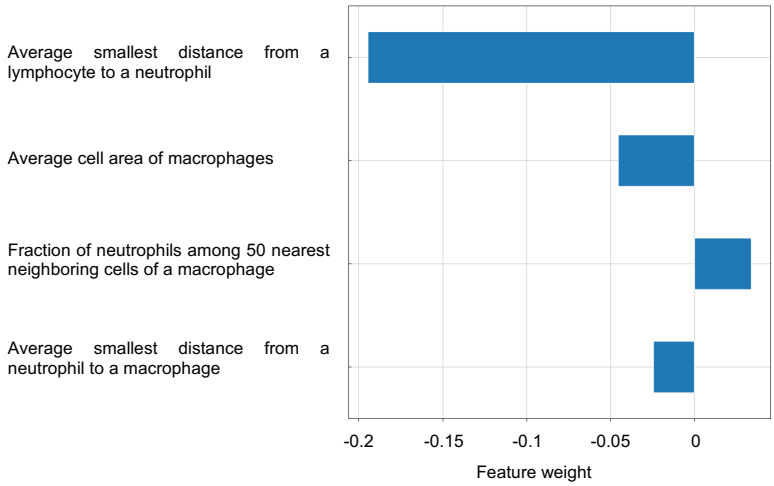
Supplementary Fig. 9 | Performance of cell classification model on WSIs across 8 common tumor types including lung, breast, esophageal, stomach, colon, rectum, liver, and pancreas cancers in TCGA. Correlation are the pair-wise Pearson correlation of image-based cell classification results and gene expression-based estimates. L, lymphocyte; N, neutrophil; M, macrophage; T, tumor cell. Source data are provided as a Source Data file.



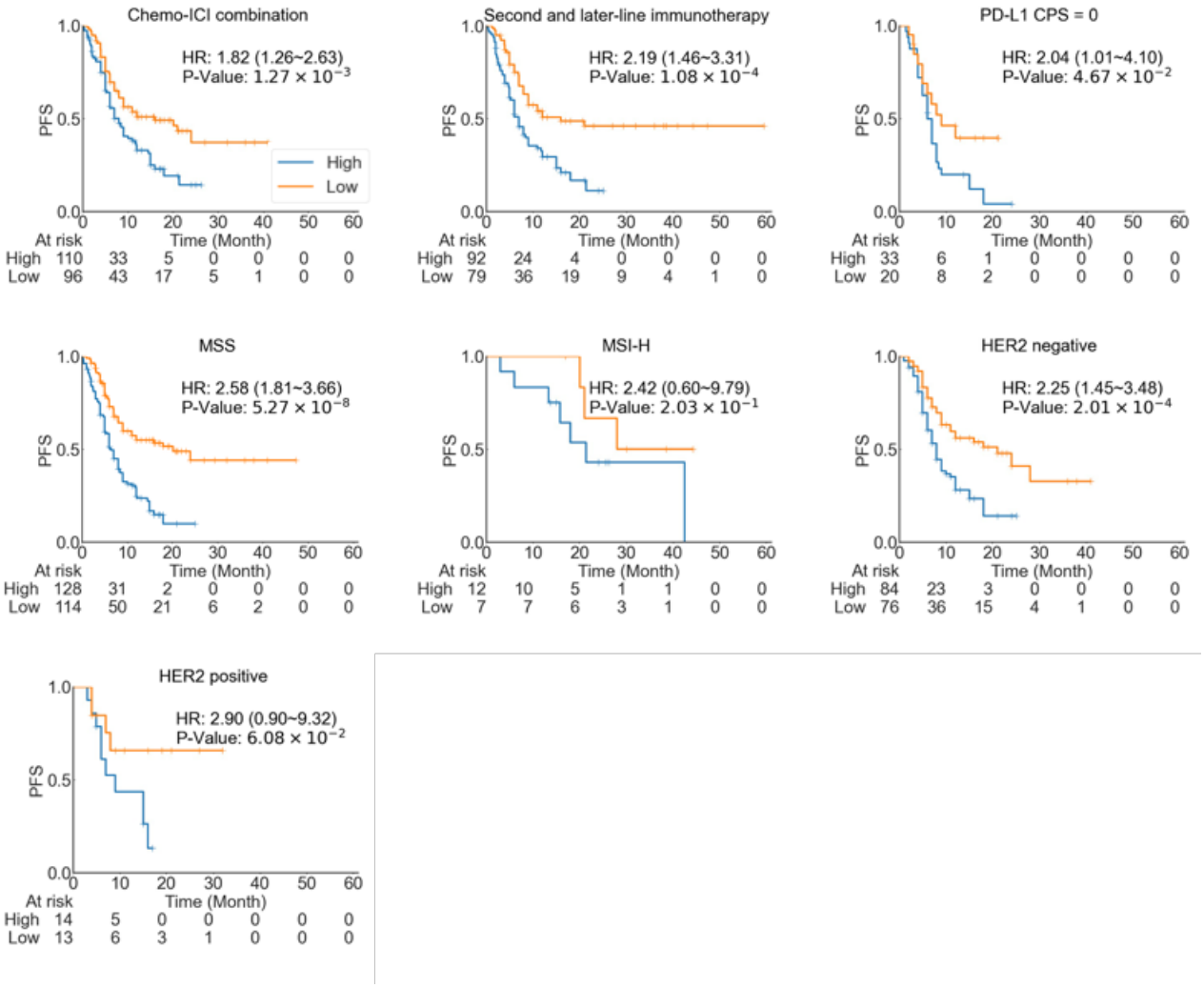
Supplementary Fig. 10 | Kaplan-Meier curves of PFS for patients with high vs low spatial biomarker value when using DINO+ViT in the 2nd TMA. P values were determined by two-sided log-rank test. Hazard ratios (HR) with 95% confidence interval were computed using the Cox proportional hazards model.



Supplementary Fig. 11 | Feature weights in the multivariate logistic regression model for predicting objective response. Source data are provided as a Source Data file.

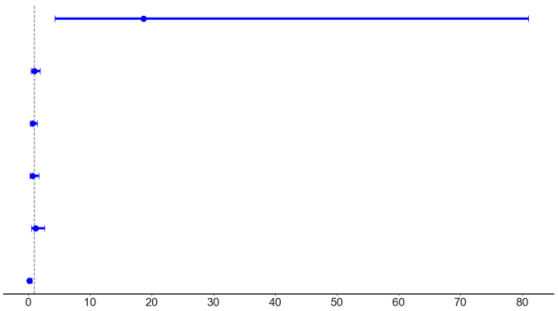


Supplementary Fig. 12 | Feature weights in the multivariate Cox regression model for PFS. Source data are provided as a Source Data file.

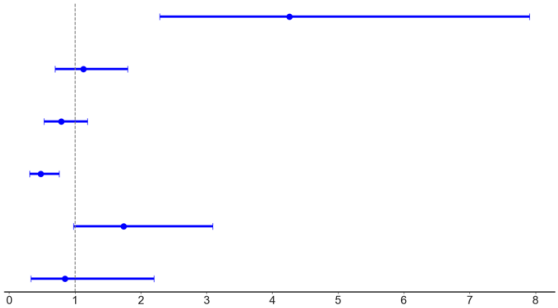


Supplementary Fig. 13 | Kaplan-Meier curves of PFS for patients with high vs low spatial biomarker value across different subgroups. P values were determined by two-sided log-rank test. Hazard ratios (HR) with 95% confidence interval were computed using the Cox proportional hazards model.

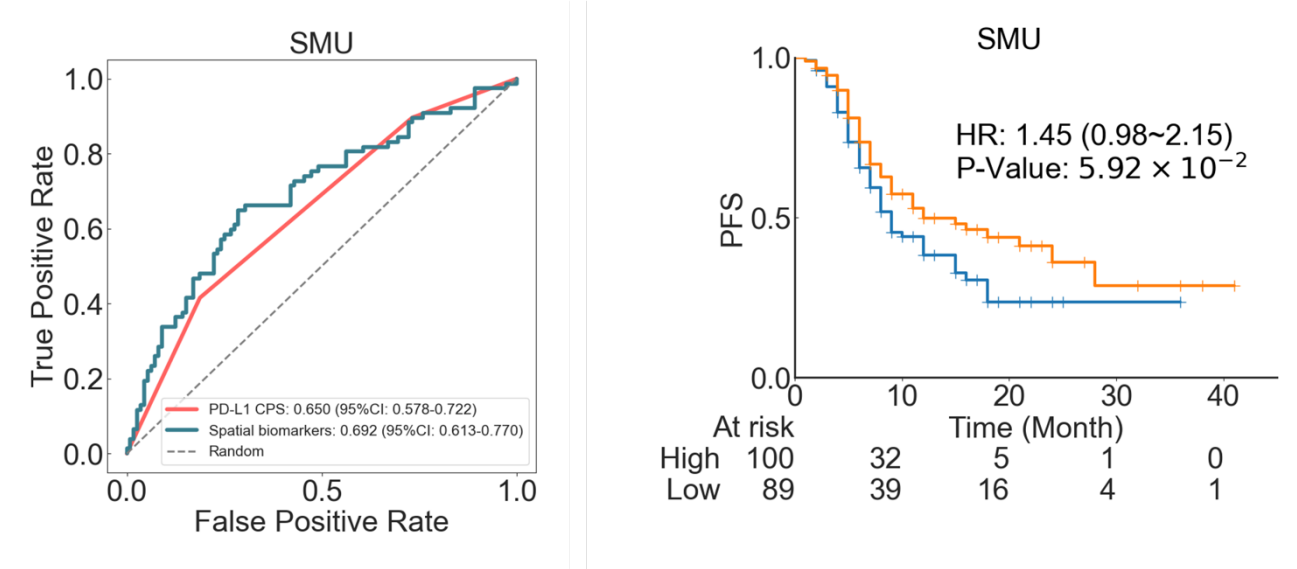
Stanford	Hazard Ratio	P-value
Spatial biomarker	18.69 (4.31,81.01)	<0.001
Age [≥65 v.s. <65]	0.97 (0.48,1.95)	0.939
Gender [Male v.s. Female]	0.70 (0.34,1.45)	0.342
PD-L1 CPS [≥1 v.s. 0]	0.69 (0.27,1.75)	0.436
Grade [3 v.s. 1/2]	1.16 (0.52,2.62)	0.716
MSI [MSI-H v.s. MSS]	0.16 (0.05,0.56)	0.004



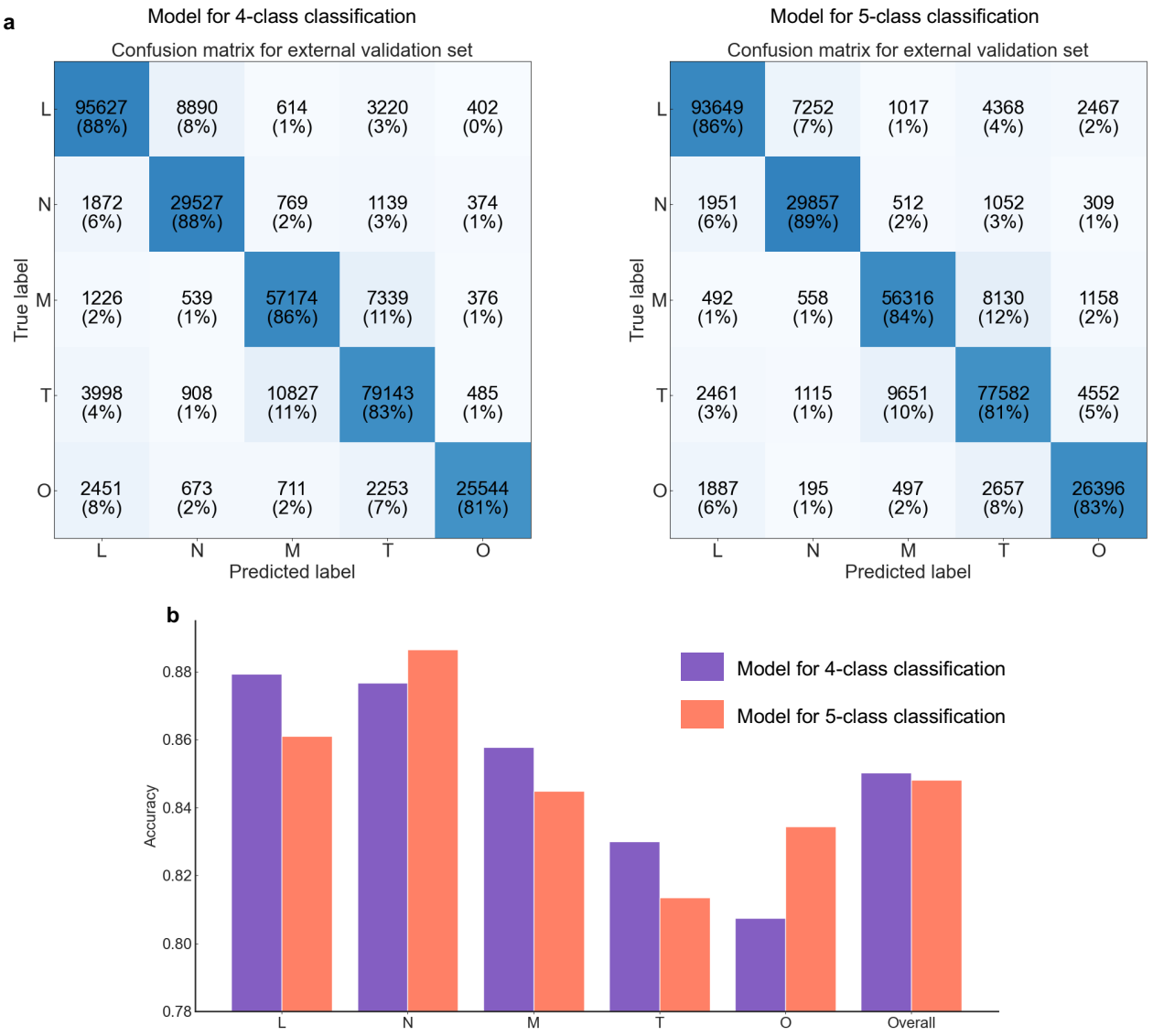
SMU	Hazard Ratio	P-value
Spatial biomarker	4.26 (2.29,7.91)	<0.001
Age [≥65 v.s. <65]	1.13 (0.70,1.80)	0.621
Gender [Male v.s. Female]	0.79 (0.53,1.19)	0.257
PD-L1 CPS [≥1 v.s. 0]	0.48 (0.31,0.76)	0.002
Grade [3 v.s. 1/2]	1.74 (0.98,3.09)	0.061
MSI [MSI-H v.s. MSS]	0.85 (0.33,2.20)	0.744



Supplementary Fig. 14 | Multivariate Cox regression analysis of PFS using the spatial biomarker and clinicopathologic variables in the Stanford and SMU cohort, respectively. P-values were computed using the two-sided Wald test. HRs are presented with 95% CI.



Supplementary Fig. 15 | Performance of traditional manual annotation approach for immunotherapy response and outcome prediction. a, The ROC curves and AUCs of the spatial biomarker vs. PD-L1 CPS for immunotherapy response prediction in the external validation (SMU) cohorts. **b,** Kaplan-Meier curves of PFS for patients with high vs low spatial biomarker value in the external validation cohorts. The cutoff point was determined based on the optimal P value in the discovery cohort and the same cutoff was used in the external validation cohort. P values were determined by two-sided log-rank test. Hazard ratios (HR) with 95% confidence interval were computed using the Cox proportional hazards model.



Supplementary Fig. 16 | Performance comparison of cell classification models training by 4 or 5-class dataset. Accuracy of cell classification model in the external validation sets. L, lymphocyte; N, neutrophil; M, macrophage; T, tumor cell; O, others. Source data are provided as a Source Data file.

SUPPLEMENTARY TABLES

Supplementary Table 1 | Concordance between automated and manual cell annotation.

Cell type	Count	Cohen's kappa (95% CI)
Tumor cells	4691	0.891 (0.884, 0.897)
Lymphocytes	20442	0.904 (0.900, 0.907)
Others	86785	0.920 (0.917,0.923)

Supplementary Table 2 | The number of cells for each cell type in the training, internal validation, and external validation sets.

Dataset	Lymphocyte	Neutrophil	Macrophage	Tumor cell	Total
Training	225,619	44,775	123,101	67,458	460,953
Internal validation	161,110	24,745	123,091	52,904	361,850
External validation	108,753	33,681	66,654	95,361	304,449
Total	495,482	103,201	312,846	215,723	1,127,252

Supplementary Table 3 | Performance of cell classification models with and without self-supervised learning (SSL) or domain adaptation (DA) on H&E images.

SSL	DA	Lymphocyte	Neutrophil	Macrophage	Tumor cell	Overall
×	×	0.85	0.83	0.82	0.77	0.82
×	✓	0.89	0.87	0.85	0.82	0.86
✓	×	0.86	0.85	0.83	0.80	0.83
✓	✓	0.88	0.89	0.86	0.83	0.86

Supplementary Table 4 | Performance comparison of cell classification models using the proposed method and state-of-the-art methods. All models were trained without domain adaptation.

Method	Lymphocyte	Neutrophil	Macrophage	Tumor cell	Overall
ResNet	0.85	0.83	0.82	0.77	0.82
VOLTA	0.86	0.83	0.82	0.79	0.82
DINO+ResNet	0.85	0.80	0.80	0.78	0.81
DINO+ViT	0.86	0.84	0.82	0.80	0.83
Virchow	0.86	0.84	0.82	0.79	0.82
UNI	0.85	0.83	0.82	0.78	0.82
BYOL+ResNet	0.86	0.85	0.83	0.80	0.83

Supplementary Table 5-24: Refer to separate spreadsheet files.

Supplementary Table 25 | Clinical characteristics of the 1st TMA cohort.

	1st TMA
Sex	
Female	35 (37%)
Male	59 (63%)
Age (median [IQR])	70 (59-77)
Tumor size (cm)	
<4	25 (27%)
≥4	73 (73%)
Tumor location	
Cardia	18 (19%)
Body	22 (23%)
Antrum	49 (52%)
Whole	5 (5%)
Depth of invasion	
pT1	2 (2%)
pT2	13 (14%)
pT3	58 (62%)
pT4a	16 (17%)
pT4b	5 (5%)
Lymph node metastasis	
pN0	22 (23%)
pN1	11 (12%)
pN2	23 (25%)
pN3a	27 (29%)
pN3b	11 (12%)
Metastasis	
pM0	93 (99%)
pM1	1 (1%)
Cancer Stage	
I	8 (9%)
II	27 (29%)
III	58(62%)
IV	1 (1%)
Grade	
II	21 (22%)
III	71 (76%)
IV	2 (2%)

Supplementary Table 26 | Clinical characteristics of the 2nd TMA cohort.

	2nd TMA
Sex	
Female	32 (29%)
Male	80 (71%)
Age (median [IQR])	61 (51-67)
Tumor size (cm)	
<4	35 (31%)
≥4	77 (69%)
Tumor location	
Cardia	10 (9%)
Body	38 (34%)
Antrum	59 (53%)
Whole	5 (4%)
Depth of invasion	
pT1	8 (7%)
pT2	13 (12%)
pT3	26 (23%)
pT4a	59 (53%)
pT4b	6 (5%)
Lymph node metastasis	
pN0	26 (23%)
pN1	13 (12%)
pN2	18 (16%)
pN3a	28 (25%)
pN3b	27 (24%)
Metastasis	
pM0	102 (91%)
pM1	10 (9%)
Cancer Stage	
I	17 (15%)
II	16 (14%)
III	69 (62%)
IV	10 (9%)
Grade	
I	5 (4%)
II	25 (22%)
III	82 (73%)

Supplementary Table 27 | Clinical characteristics of the Stanford immunotherapy cohort.

	Stanford immunotherapy
Sex	
Female	66 (67%)
Male	32 (33%)
Age (median [IQR])	64 (53-73)
Primary tumor site	
Esophagus	21 (21%)
GE Junction	18 (18%)
Stomach	59 (60%)
MSI status	
MSS	64 (65%)
MSI-H	11 (11%)
Unknown	23 (23%)
PD-L1 CPS	
0%	15 (15%)
1-9%	34 (35%)
≥ 10%	38 (39%)
Unknown	11 (11%)
Line of therapy	
1L	52 (53%)
2L	27 (28%)
3L+	19 (19%)
Concurrent chemotherapy	
Yes	52 (53%)
No	46 (47%)
Grade	
I	1 (1%)
II	22 (22%)
III	53 (54%)
Unknown	22 (22%)

Supplementary Table 28 | Clinical characteristics of the SMU immunotherapy cohort.

	SMU immunotherapy
Sex	
Female	81 (43%)
Male	108 (57%)
Age (median [IQR])	56 (48-65)
Tumor location	
Cardia	51 (27%)
Body	42 (22%)
Antrum	81 (43%)
Whole	15 (8%)
MSI status	
MSS	178 (94%)
MSI-H	8 (4%)
Unknown	3 (2%)
HER2 status	
Negative	160 (85%)
Positive	27 (14%)
Unknown	2 (1%)
PD-L1 CPS	
0%	38 (20%)
1-9%	98 (52%)
≥ 10%	53 (28%)
Line of therapy	
1L	64 (34%)
2L	71 (38%)
3L+	54 (29%)
Concurrent chemotherapy	
Yes	154 (81%)
No	35 (19%)
Grade	
I	4 (2%)
II	26 (14%)
III	159 (84%)

Supplementary Table 29 | Antibody sources and staining conditions.

Reagent	Source	Identifier	Dilution
CD3	Cell Signaling Technology	Cat#85061S	1:600
CD8	Cell Signaling Technology	Cat#85336S	1:500
CD20	Abcam	Cat#ab78237	1:3000
PD1	Cell Signaling Technology	Cat#43248	1:50
PDL1	Cell Signaling Technology	Cat#13684	1:100
Pan-CK	Cell Signaling Technology	CST4545	1:200
CD66b	Gene Tex	GTX19779	1:300
CD68	ZSGB-BIO	ZM0060	1:1
CD163	Cell Signaling Technology	Cat#93498	1:1000
FoxP3	Abcam	Cat#ab20034	1:500

Supplementary Table 30 | Proportion of cells that received probabilities below 0.5 in the internal and external validation sets.

Cell type	Internal validation	External validation
Lymphocytes	0.002	0.004
Neutrophils	0.012	0.011
Macrophages	0.006	0.006
Tumor cells	0.007	0.005