

Prognostic value of a modified Immunoscore in patients with stage I–III resectable colon cancer

Ke Zhao^{1,2*}, Xiaomei Wu^{3*}, Zhenhui Li^{4*}, Yingyi Wang^{5*}, Zeyan Xu¹, Yajun Li¹, Lin Wu⁶, Su Yao⁷, Yanqi Huang^{1,8}, Changhong Liang¹, Zaiyi Liu¹

¹Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China; ²School of Medicine, South China University of Technology, Guangzhou 510006, China; ³Department of Radiology, the Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou 510655, China; ⁴Department of Radiology, the Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Yunnan Cancer Center, Kunming 650118, China; ⁵Department of Radiology, Zhuhai People's Hospital, Zhuhai Hospital Affiliated with Jinan University, Zhuhai 519000, China; ⁶Department of Pathology, the Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Yunnan Cancer Center, Kunming 650118, China; ⁷Department of Pathology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China; ⁸the Second School of Clinical Medicine, Southern Medical University, Guangzhou 510080, China

*These authors contributed equally to this work.

Correspondence to: Zaiyi Liu. Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, No. 106 Zhongshan Er Road, Guangzhou 510080, China. Email: liuzaiyi@gdph.org.cn; Changhong Liang. Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, No. 106 Zhongshan Er Road, Guangzhou 510080, China. Email: liangchanghong@gdph.org.cn; Yanqi Huang. Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, No. 106 Zhongshan Er Road, Guangzhou 510080, China. Email: yikiann@126.com; Su Yao. Department of Pathology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, No. 106 Zhongshan Er Road, Guangzhou 510080, China. Email: ysjulien@126.com.

Abstract

Objective: The Immunoscore method has proved fruitful for predicting prognosis in patients with colon cancer. However, there is still room for improvement in this scoring method to achieve further advances in its clinical translation. This study aimed to develop and validate a modified Immunoscore (IS-mod) system for predicting overall survival (OS) in patients with stage I–III colon cancer.

Methods: The IS-mod was proposed by counting CD3+ and CD8+ immune cells in regions of the tumor core and its invasive margin by drawing two lines of interest. A discovery cohort (N=212) and validation cohort (N=103) from two centers were used to evaluate the prognostic value of the IS-mod.

Results: In the discovery cohort, 5-year survival rates were 88.6% in the high IS-mod group and 60.7% in the low IS-mod group. Multivariate analysis confirmed that the IS-mod was an independent prognostic factor for OS [adjusted hazard ratio (HR)=0.36, 95% confidence interval (95% CI): 0.20–0.63]. With less annotation and computation cost, the IS-mod achieved performance comparable to that of the Immunoscore-like (IS-like) system (C-index, 0.676 vs. 0.661, P=0.231). The 2-category IS-mod using 47.5% as the threshold had a better prognostic value than that using a fixed threshold of 25% (C-index, 0.653 vs. 0.573, P=0.004). Similar results were confirmed in the validation cohort.

Conclusions: Our method simplifies the annotation and accelerates the calculation of Immunoscore method, thus making it easier for clinical implementation. The IS-mod achieved comparable prognostic performance when compared to the IS-like system in both cohorts. Besides, we further found that even with a small reference set (N≥120), the IS-mod still demonstrated a stable prognostic value. This finding may inspire other institutions to develop a local reference set of an IS-mod system for more accurate risk stratification of colon cancer.

Keywords: Immunoscore; colon cancer; whole-slide image; overall survival; digital pathology

Submitted Jan 09, 2021. Accepted for publication Mar 26, 2021.

doi: 10.21147/j.issn.1000-9604.2021.03.09

View this article at: <https://doi.org/10.21147/j.issn.1000-9604.2021.03.09>

Introduction

Colon cancer is one of the leading causes of death in many countries (1). Although the tumor-node-metastasis (TNM) classification of the Union for International Cancer Control has set the foundation for the prognosis of patients with resectable colon cancer (2), the clinical prognosis can vary substantially across patients within the same stage. The recent demonstration of the prognostic performance of Immunoscore, a composite scoring system summarizing the density of CD3+ and CD8+ T-cell infiltration within the core of tumor (CT) and its invasive margin (IM), favours its implementation as a new component to the TNM staging to achieve better patient stratification (3-5).

However, before translating the Immunoscore system into routine clinical practice, a few issues remain to be addressed. Firstly, the Immunoscore calculation process requires pathologists to delimitate the entire tumor region on whole-slide images (WSIs), which is time-consuming. Moreover, given the very high resolution of WSIs (6,7), calculating the immune cell density for the entire tumor region is a computationally intensive task. Finally, the fixed thresholds (25% and 70%) of the Immunoscore used for patient stratification were determined on a study population of patients with certain heterogeneity (8-10). As the colon cancer incidence varies across countries, owing to variations in Human Development Index scores (1), the absolute thresholds for Immunoscore categorization may not be applicable to other populations. Therefore, there is still considerable room for the fine-tuning of the Immunoscore regarding its generation and categorization to achieve further advances in its clinical translation.

Therefore, we performed this study to propose a modified Immunoscore (IS-mod) system to facilitate rapid annotation and generation and determine whether the IS-mod system could achieve accurate survival prediction in patients with resectable colon cancer. We further investigated how many samples would be needed as a reference set to establish an effective IS-mod system.

Materials and methods

Patients

Two cohorts of patients with stage I-III colon cancer were

enrolled (*Figure 1A*). Patients in the discovery cohort were enrolled from Guangdong Provincial People's Hospital (GDPH, Mar 2009 to Dec 2014). Patients from Yunnan Cancer Hospital (YCH, Jan 2009 to Dec 2014) served as the validation cohort. Exclusion criteria were neoadjuvant therapy (radiotherapy, chemotherapy) and death within 30 d of surgery. This retrospective study was approved by the Institutional Review Board at both institutions. Clinicopathological information, including age at diagnosis, sex, TNM classification, and anatomic site, was obtained from the medical records archives. The primary endpoint was overall survival (OS), which was defined as the date of the surgery until death. Patients who were still alive at the last followed date were labelled as censored data.

Immunohistochemical staining

The most invasive part of the primary tumor was chosen from the Paraffin tissue block from each patient by a pathologist who was blinded to patient clinical information and outcome (8,11,12). Two 4- μ m sections, containing both tumor region and invasive margin, were cut from the block for immunohistochemistry. Several steps were taken: deparaffinage, antigen retrieval solution (using 10 \times concentrate solution, Novocastra, Leica) and primary [human anti-CD3 (Gene Tech, Cat# GT200229) and anti-CD8 (Cat# GT211229) rabbit monoclonal antibodies] and secondary (rabbit-anti-mouse IgG, Bond Refine Detection Kit, Leica) antibodies, according to the manufacturer's recommendations in a Ventana BenchMark automated staining system. Finally, sections were subsequently incubated with 3,3-Diaminobenzidine, counterstained with hematoxylin, and mounted with special glue. Internal positive control was used for quality assurance. The stained tissue sections were imaged using a digital Whole Slide Scanning system (Leica, Aperio-AT2, USA) at 40 \times magnification.

IS-mod generation

We established the IS-mod system based on the method previously described (8), with a certain fine-tuning steps as hereby. The ImageScope software (Version 12.4.3; Leica Biosystems, Wetzlar, Germany) was used for image

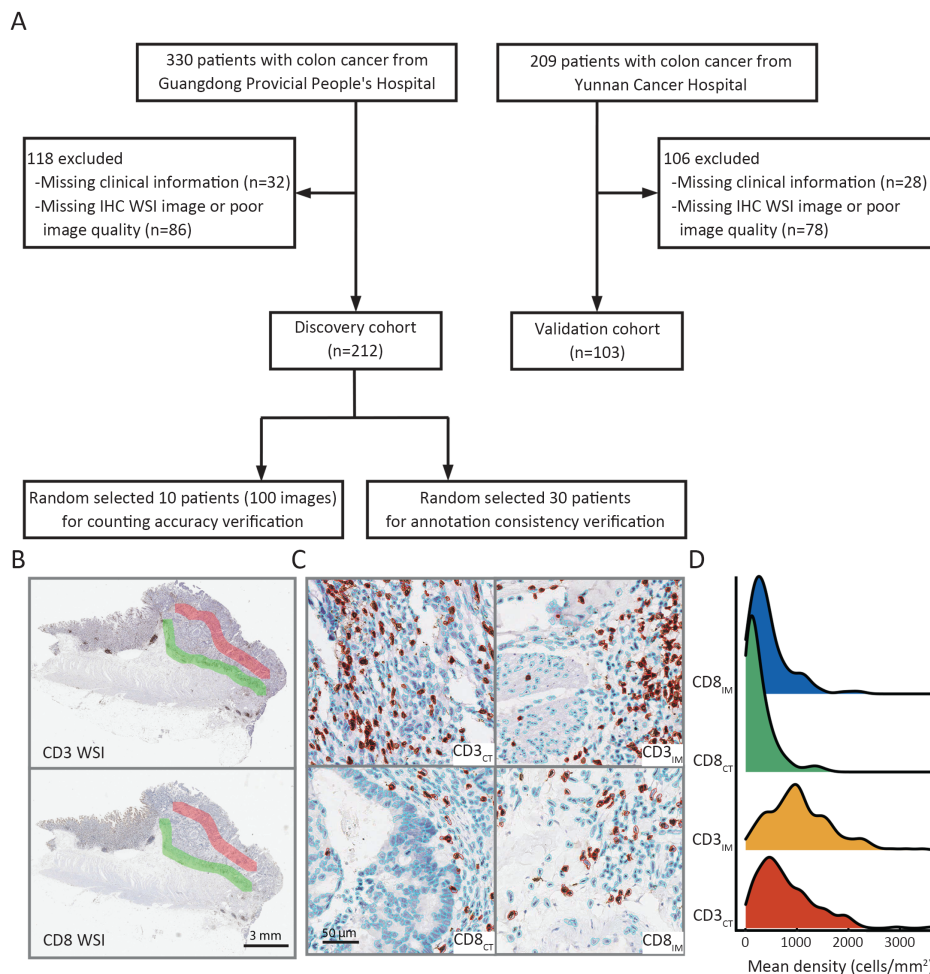


Figure 1 Study profile and modified Immunoscore calculation. (A) Study profile; (B) CD3 and CD8 WSIs with annotation. Red region for CT and green region for IM; (C) Examples of cell segmentation results. Immune cells were marked in red, and other types were marked in blue; (D) Mean density distributions of four cells counting results. CT, core of the tumor; IM, invasive margin; WSI, whole-slide image.

annotation. The annotation procedure was blinded to patient clinical information and outcome. On each CD3-stained WSI, one line was drawn at the tumor invasive margin (Line_{IM}), and another line with a similar length was drawn at the core of tumor (Line_{CT}). Subsequently, two regions of interest (ROIs) were automatically generated by expanding 2,400 pixels (at 40× magnification) outwards with both lines as the center line in each case (Figure 1B). Next, in both ROIs, a self-developed software (based on MATLAB R2018a, MathWorks, USA) was used to count the CD3-positive cells (Figure 1C). Next, the mean densities of positive cells in the CT and IM regions were calculated. The same procedure was performed for each CD8 image.

The distribution of immune cell density in the reference set (the discovery cohort) and the quantile of each patient

were calculated (Figure 1D). Hence, four quantiles (CD3_{IM}, CD3_{CT}, CD8_{IM}, and CD8_{CT}) were obtained, and the average of these four quantiles was calculated to obtain the IS-mod (percentile). According to the IS-mod (percentile) distribution in the reference set, patients were divided into three categories (high/intermediate/low) to obtain a 3-category IS-mod. Similarly, a 2-category IS-mod (high/low) was obtained according to a two-level threshold. These thresholds were determined in the reference set to balance the proportion of patients in each category (1:1:1 for high/intermediate/low, and 1:1 for high/low).

Immune cell counting and annotation consistency assessment

We randomly selected 100 image blocks from 20 WSIs in

the discovery cohort to assess immune cells counting accuracy. These images were annotated by a pathologist (S.Y.) using ImageJ software (Version 1.51; NIH, USA) to get the ground truth. The interclass correlation coefficient (ICC) was calculated for the purpose of evaluating the consistency of the positive cell count (13). The Bland-Altman method was used to measure the agreement between the software and pathologist annotation (14).

Also, we evaluated the consistency between different annotators. Thirty cases (60 WSIs) were randomly selected from the discovery cohort. Annotator 1 (S.Y.) and Annotator 2 (L.W.) independently annotated these images' CT and IM regions. The ICC was also used to evaluate the agreement between different annotators, and an ICC>0.85 was considered to indicate a high level of agreement.

Compared with the calculation process associated with the Immunoscore, we selected only a band-shaped ROI (Figure 1B) in the tumor region to calculate the immune cell density. To investigate whether the ROIs reflected the immune status of the tumor region, we randomly selected 100 patients from the discovery cohort. The whole tumor region was segmented for CD3- and CD8-stained WSIs. The density of positive cells in the entire tumor region was calculated using the same cell-counting software. A correlation analysis [Pearson correlation coefficient (r)] was performed to obtain the immune cell densities in the band-shaped ROI and the whole tumor region.

Prognostic value of IS-mod

For the 3- and 2-category IS-mod, the Kaplan-Meier method was used to analyze the survival curves. Log-rank statistics were used to test the differences in survival distributions. The Cox model was used to compute hazard ratio (HR) of IS-mod and other clinicopathological factors (age, sex, TNM stage, and tumor site) for OS in the two cohorts. A reference model was constructed based on independent clinicopathological factors in multivariate survival analyses. A full model was developed by combining IS-mod with independent clinicopathological factors to investigate the IS-mod's incremental value for outcome prediction.

The performance of the IS-mod and the reference/full models was assessed using Harrell's C-index, the time-dependent area under the receiver operating characteristic curve (tAUC), and the integrated area under the receiver operating characteristic curve (iAUC) (15-17). The two-tailed Student's *t* test for dependent samples was used to

compare the two C-indices (and two tAUCs), and the iAUCs were compared using a Wilcoxon rank-sum test for dependent samples (18).

Comparison of IS-mod system and Immunoscore-like (IS-like) system

As we used a modified annotation process, especially for the CT region, it was worth exploring whether the prognostic value of the IS-mod system is comparable to that of the Immunoscore system. After calculating the CD3+ and CD8+ cells densities in the segmented whole tumor region (the invasive margin densities were already calculated), an IS-like system was established that was very similar to the Immunoscore calculation process (8). Performance evaluation indexes, such as C-index and iAUC, were calculated for the IS-like system.

Sample size of reference set

To address how many samples of the reference set were needed to constitute an effective IS-mod system, we arranged patients in the discovery cohort according to time (date of surgery). Next, a series of reference sets were established by extracting patients at 10-patient intervals from the discovery cohort. These sets serve as references for immune cell density distributions. The IS-mod of each case in the two cohorts was calculated according to the reference set. C-indexes were calculated to reflect the trend in IS-mod prediction performance changes with the increase in sample size.

Statistical analysis

All statistical analyses were performed with R software (Version 3.6.1; R Foundation for Statistical Computing, Vienna, Austria) using the following packages: BlandAltmanLeh, irr, MASS, prodlim, rms, survcomp, survival, survminer, and timeROC. A P-value of <0.05 was considered statistically significant.

Results

Patients

In total, 315 consecutive cases with histologically confirmed colon cancer were included in this study. Among them, 212 patients from GDPH served as the discovery cohort, and 103 patients from YCH served as the validation cohort. The baseline clinicopathological characteristics are

shown in Table 1, including age, sex, T category, N category, TNM stage, and tumor site. Significant differences were found between the two cohorts regarding age, T category, TNM stage, and tumor site ($P < 0.05$, Table 1). As of the last follow-up, 63 (29.7%) patients died in the discovery cohort and 22 (21.4%) in the validation cohort. The median follow-up time was 6.4 [interquartile range (IQR), 5.7–8.3] years in the discovery cohort and 4.2 (IQR, 3.7–5.0) years in the validation cohort. The median time for OS had not been reached.

Calculation and assessment of IS-mod

Compared with the previous Immunoscore annotation

Table 1 Distributions of demographic and clinicopathologic characteristics of colon cancer patients in two cohorts

Variables	n (%)		P*
	Discovery cohort (N=212)	Validation cohort (N=103)	
Age (year)			<0.001
≤60	76 (35.8)	59 (57.3)	
>60	136 (64.2)	44 (42.7)	
Sex			0.739
Male	91 (42.9)	47 (45.6)	
Female	121 (57.1)	56 (54.4)	
T category			<0.001
T1	5 (2.4)	0 (0)	
T2	24 (11.3)	0 (0)	
T3	161 (75.9)	67 (65.0)	
T4	22 (10.4)	36 (35.0)	
N category			0.679
N0	112 (52.8)	49 (47.6)	
N1	66 (31.1)	36 (35.0)	
N2	34 (16.0)	18 (17.5)	
TNM stage			0.002
I	24 (11.3)	0 (0)	
II	88 (41.5)	49 (47.6)	
III	100 (47.2)	54 (52.4)	
Tumor site			<0.001
Cecum ascending colon	94 (44.3)	27 (26.2)	
Descending colon	25 (11.8)	14 (13.6)	
Sigmoid colon	83 (39.2)	36 (35.0)	
Transverse colon	10 (4.7)	26 (25.2)	

*, P value was performed by χ^2 test.

process (8), we used a faster annotation method. Instead of outlining the entire tumor region, we delineated two similar-length lines at the IM and CT regions on the WSIs. After counting the immune cells in the above two areas, we obtained mean density distributions of $CD3_{IM}$, $CD3_{CT}$, $CD8_{IM}$ and $CD8_{CT}$. The 3-category IS-mod was calculated using 38.5% and 60.8% as thresholds. In the same way, the 2-category IS-mod was established based on a cutoff value of 47.5%.

Figure 2A shows examples of the software-based segmentation and pathologist’s manual annotation results. Perfect concordance of the immune cell counting was observed between the software segmentation and the pathologist annotation [ICC=0.985, 95% confidence interval (95% CI): 0.973–0.991 for CD3; ICC=0.932, 95% CI: 0.884–0.961 for CD8; Figure 2B]. Bland-Altman plots show good agreements between the software and the pathologist’s annotation (Figure 2C,D). After performing a consistency analysis for different annotators, we found that the interrater reliability indexes to be high (ICCs, 0.864–0.951; Figure 3A).

We also performed a correlation analysis to investigate whether a band-shaped subset region could reflect the whole tumor region’s immune status. For CD3, a strong correlation was observed ($r=0.81$, $P < 0.001$; Figure 3B). A similar result was also found for the CD8-stained WSIs ($r=0.89$, $P < 0.001$; Figure 3C).

Prognostic value of IS-mod

Patients with a higher IS-mod had much longer survival. The 5-year survival rates for the low, intermediate, and high IS-mod groups were 58.3%, 77.5%, and 88.4%, respectively (the unadjusted HR for high vs. low=0.28, 95% CI: 0.15–0.54, $P < 0.001$; Figure 4A). For the 2-category IS-mod, 107 (50.5%) patients had a low IS-mod, and 105 (49.5%) patients had a high IS-mod. The 5-year survival rates for the low and high IS-mod groups were 60.7% and 88.6%, respectively (the unadjusted HR for high vs. low=0.28, 95% CI: 0.16–0.49, $P < 0.001$; Figure 5A).

Similar results were confirmed in the validation cohort. Survival rates at five years were recorded for 70.6% of patients with a low IS-mod, 81.5% of patients with an intermediate IS-mod, and 96.0% of patients with the high IS-mod (the unadjusted HR for high vs. low=0.09, 95% CI: 0.01–0.70, $P=0.021$; Figure 4B). For the IS-mod with 2-level, 62 patients (60.2%) were classified as the low IS-mod and 41 (39.8%) as the high IS-mod. The 5-year survival

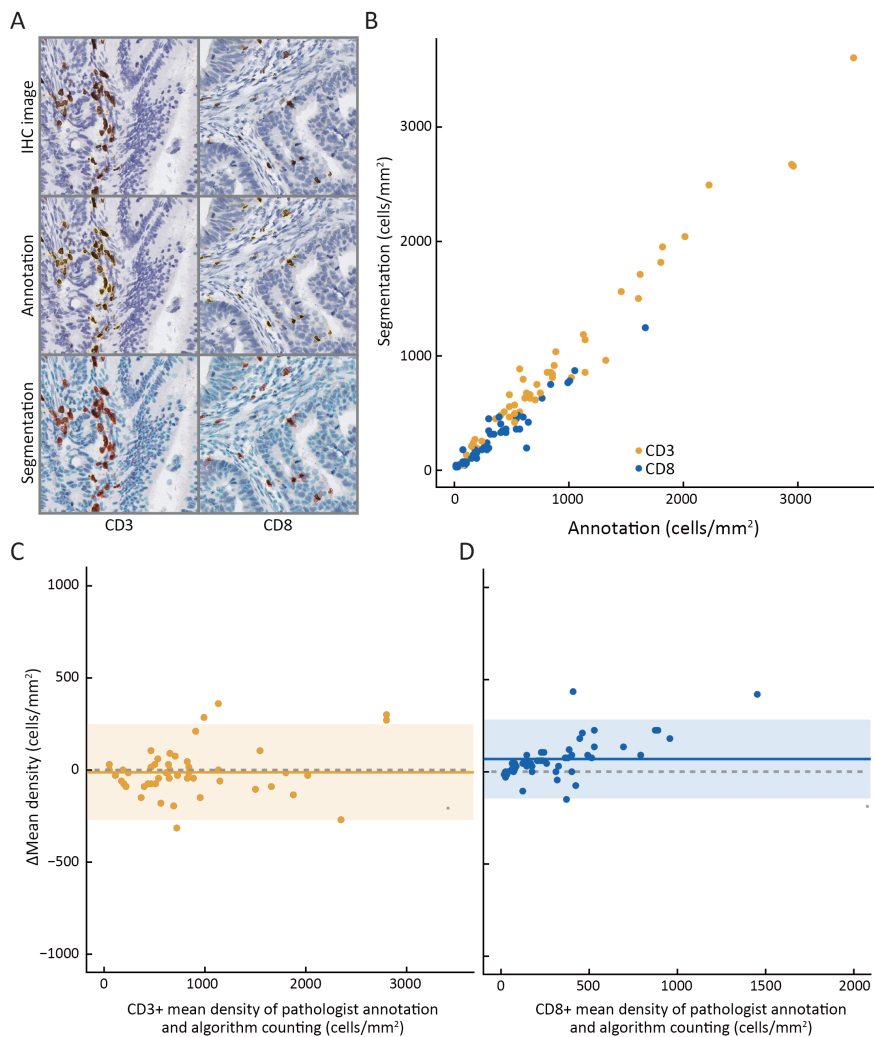


Figure 2 Immune cell counting consistency analyses. (A) Examples of IHC (CD3 and CD8) image block ($1,024 \times 1,024$ pixel² at $40\times$ magnification), pathologist annotation, and segmentation result; (B) Concordance of immune cells counting between automatic segmentation and pathologist annotation (CD3: ICC=0.985, 95% CI, 0.973–0.991; CD8: ICC=0.932, 95% CI, 0.884–0.961); (C) Bland-Altman plot for CD3+ cells counting estimation; (D) Bland-Altman plot for CD8+ cell counting estimation. The solid horizontal line is mean, the dashed line is zero, and the shaded regions are 95% CIs; IHC, immunohistochemical; ICC, interclass correlation coefficient; 95% CI, 95% confidence interval.

rates for the low and high IS-mod groups were 69.4% and 95.1%, respectively (the unadjusted HR for high vs. low=0.11, 95% CI: 0.02–0.46, $P=0.003$; Figure 5B).

Performance comparison of IS-mod and IS-like

Table 2 lists performance of IS-mod and IS-like for predicting OS in the two cohorts. To avoid the influence of the cutoff setting, the continuous IS-mod and IS-like were used to compare performance. The C-indexes, AUCs at 5-year, and iAUCs were calculated. For both cohorts, the IS-

mod performed slightly better than the IS-like (0.676 vs. 0.661 in the discovery cohort; 0.671 vs. 0.657 in the validation cohort), but the differences were not statistically significant (all $P>0.05$). Table 2 also summarizes the predictive performance of the IS-mod, using different cutoff values. Compared with the IS-mod based on 25% and 70% as thresholds recommended in the previous study (8), the IS-mod based on the threshold determined by the reference set had better prognostic performance, especially for the 2-category IS-mod ($P<0.05$, Supplementary Table S1).

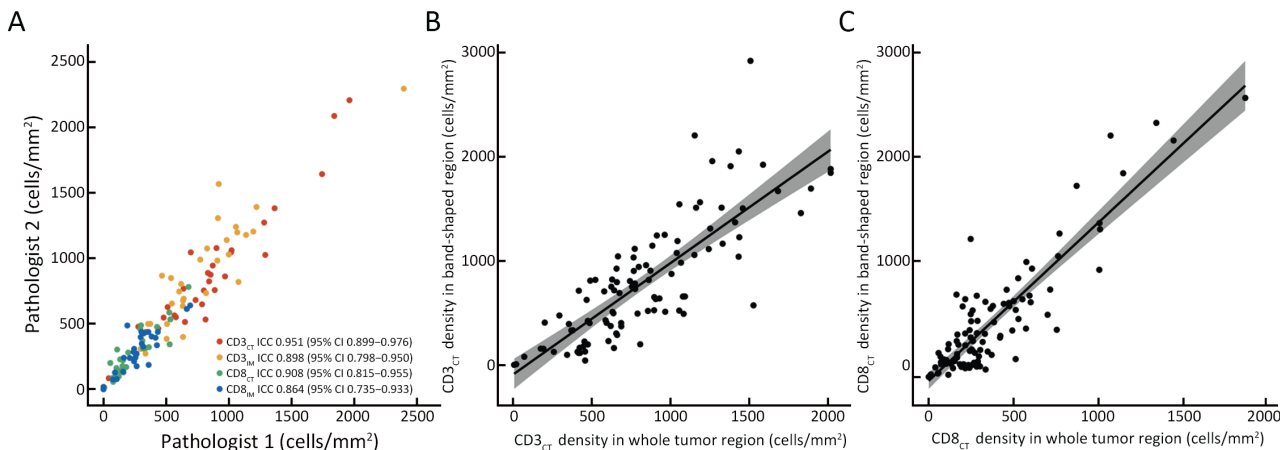


Figure 3 Consistency and correlation analyses. (A) CT and IM annotation consistency analysis (CD3_{CT}: ICC=0.951, 95% CI, 0.899–0.976; CD3_{IM}: ICC=0.898, 95% CI, 0.798–0.950; CD8_{CT}: ICC=0.908, 95% CI, 0.815–0.955; CD8_{IM}: ICC=0.864, 95% CI, 0.735–0.933); (B,C) Correlation analyses for immune cells CD3 (r=0.81, P<0.001) (B) and CD8 (r=0.89, P<0.001) (C) density in the whole tumor region and the band-shaped region. CT, core of the tumor; IM, invasive margin; ICC, interclass correlation coefficient; 95% CI, 95% confidence interval.

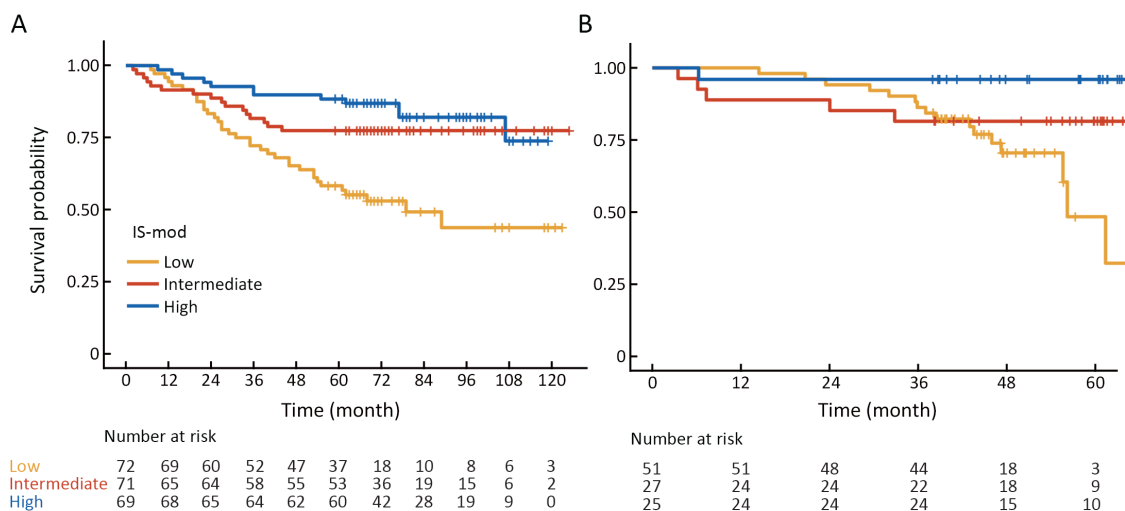


Figure 4 Kaplan-Meier curves analysis for 3-category IS-mod. (A) The discovery cohort (HR for high vs. low=0.28, 95% CI, 0.15–0.54, P<0.001); (B) The validation cohort (HR for high vs. low=0.09, 95% CI, 0.01–0.70, P=0.021). IS-mod, modified Immunoscore; HR, hazard ratio; 95% CI, 95% confidence interval.

Incremental value of IS-mod

The IS-mod is an independent prognostic factor that can provide incremental value for predicting OS in patients with stage I–III colon cancer. After univariate and multivariate analyses (2-category IS-mod used only), we identified IS-mod, age, and TNM stage as independent predictors for OS (the adjusted HR=0.36, 95% CI: 0.20–0.63 in the discovery cohort; the adjusted HR=0.15, 95% CI: 0.03–0.67 in the validation cohort; Table 3). A reference model (TNM stage plus age) and a full model

(reference model plus IS-mod) were developed. The full model showed better discrimination performance than the reference model (iAUC, 0.775 vs. 0.746 in the discovery cohort; 0.788 vs. 0.767 in the validation cohort; Supplementary Table S2; Figure 5C,D).

Sample size of reference set

The procedure for constructing the reference sets is shown in Figure 6A. We found that when the reference set’s sample size was small (N<70), the IS-mod’s predictive

Table 2 Prediction performance of IS-mod (percentile), IS-like (percentile), 3-category, and 2-category IS-mod in two cohorts

Variables	Discovery cohort			Validation cohort		
	C-index (95% CI)	AUC at 5-year (95% CI)	iAUC	C-index (95% CI)	AUC at 5-year (95% CI)	iAUC
IS-mod (percentile)	0.676 (0.608–0.743)	0.706 (0.623–0.789)	0.657	0.671 (0.557–0.786)	0.705 (0.593–0.818)	0.704
IS-like (percentile)	0.661 (0.594–0.729)	0.682 (0.601–0.763)	0.654	0.657 (0.553–0.760)	0.710 (0.599–0.821)	0.664
3-cat IS-mod	0.643 (0.581–0.706)	0.676 (0.600–0.752)	0.614	0.623 (0.518–0.727)	0.666 (0.565–0.767)	0.513
2-cat IS-mod	0.653 (0.597–0.710)	0.683 (0.616–0.751)	0.638	0.666 (0.580–0.752)	0.690 (0.607–0.773)	0.643
3-cat IS-mod _{25%,70%}	0.602 (0.539–0.665)	0.611 (0.535–0.687)	0.591	0.600 (0.500–0.700)	0.609 (0.500–0.719)	0.526
2-cat IS-mod _{25%}	0.573 (0.521–0.625)	0.579 (0.516–0.642)	0.553	0.568 (0.461–0.674)	0.574 (0.458–0.691)	0.487

IS, Immunoscore; cat, category; IS-mod, modified Immunoscore; IS-like, Immunoscore-like; IS-mod_{25%,70%}, 3-category IS-mod using 25% and 70% as thresholds; IS-mod_{25%}, 2-category IS using 25% as threshold; 95% CI, 95% confidence interval; AUC, area under the receiver operating characteristic curve; iAUC, integrated AUC.

Table 3 Univariate and multivariate analyses in two cohorts

Variables	Discovery cohort				Validation cohort			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Age	1.03 (1.01–1.06)	0.004	1.03 (1.01–1.05)	0.021	1.01 (0.98–1.04)	0.631		
Sex								
Male	1				1			
Female	1.02 (0.62–1.68)	0.947			1.37 (0.59–3.17)	0.468		
Stage								
I	1		1					
II	3.96 (0.52–30.1)	0.184	3.87 (0.51–29.5)	0.191	1		1	
III	14.8 (2.04–107)	0.008	11.5 (1.59–83.8)	0.016	4.85 (1.64–14.4)	0.004	3.30 (1.01–9.17)	0.049
Location								
Cecum ascending	1				1			
Descending	0.43 (0.15–1.22)	0.113			0 (0–Inf)	0.997		
Sigmoid	0.81 (0.48–1.37)	0.429			0.69 (0.26–1.85)	0.466		
Transverse	0.54 (0.13–2.25)	0.379			0.68 (0.23–1.97)	0.477		
IS-mod								
Low	1		1		1		1	
High	0.28 (0.16–0.49)	<0.001	0.36 (0.20–0.63)	<0.001	0.11 (0.02–0.46)	0.003	0.15 (0.03–0.67)	0.013

IS-mod, modified Immunoscore; HR, hazard ratio; 95% CI, 95% confidence interval.

performance was unstable. However, with a sample size of more than 70 cases, the performance improved steadily in both cohorts. However, when the sample size increased to more than 120 cases, the performance tended to be stable (Figure 6B).

Discussion

In this study, to pave the way for the clinical translation of the Immunoscore, we proposed an IS-mod system that could facilitate rapid annotation and calculation. The IS-mod achieved comparable prognostic performance as the

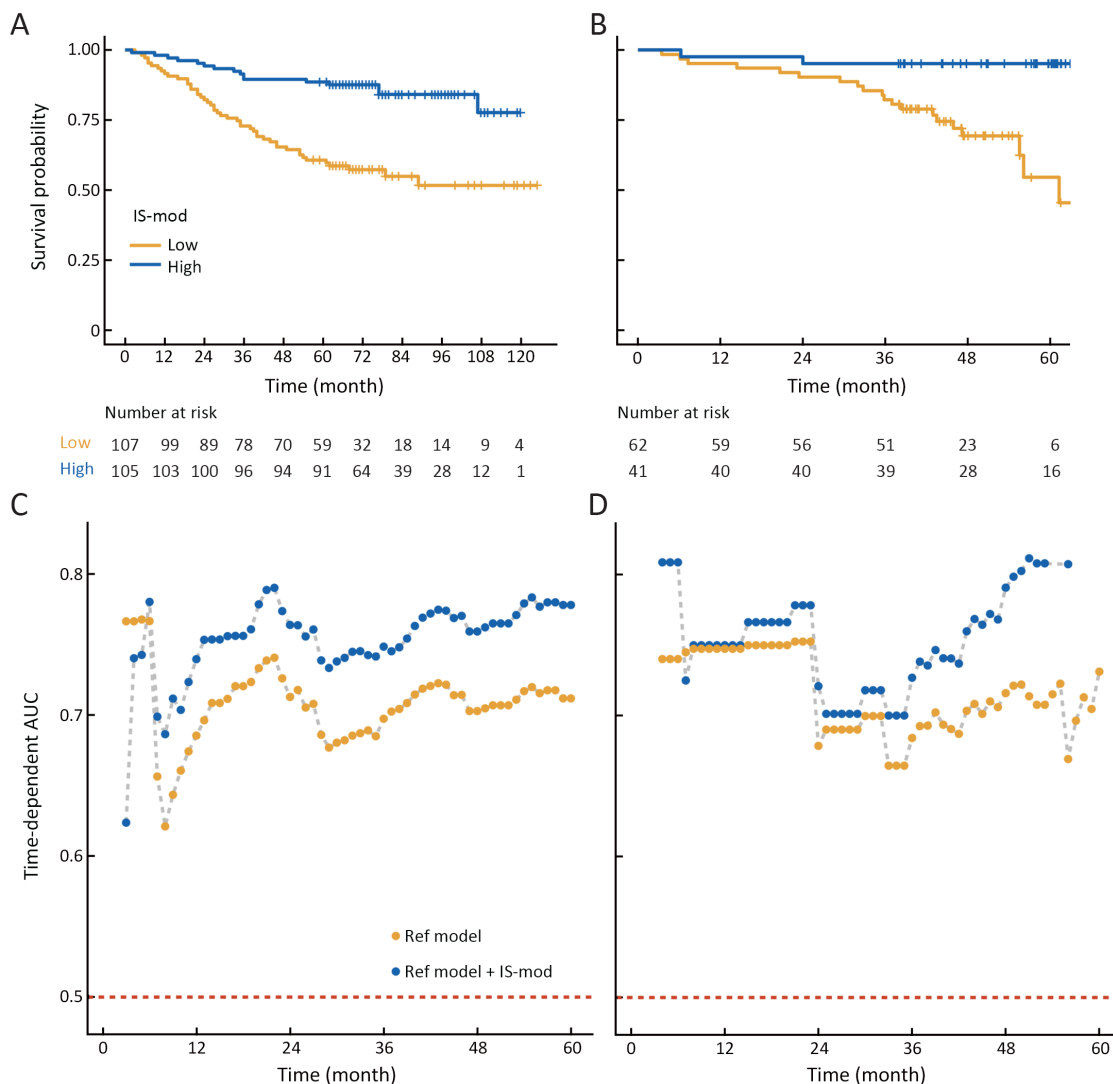


Figure 5 Kaplan-Meier curves analysis and performance of models. (A,B) Kaplan-Meier survival curves of overall survival of high vs. low IS-mod [(A) HR=0.28, 95% CI, 0.16–0.49; P<0.001; C-index=0.653, 95% CI, 0.597–0.710; (B) HR=0.11, 95% CI, 0.02–0.46; P=0.003; C-index=0.666, 95% CI, 0.580–0.752]; (C,D) Time-dependent AUC was measured from one month to five years at 1-month intervals, reflecting the performance of predicting overall survival at different timepoints. A reference model was developed with only TNM stage and age, and a full model (reference model + IS-mod) is shown compared with the reference model. Left panel: discovery cohort; Right panel: validation cohort. AUC, area under the receiver operating characteristic curve; TNM, tumor-node-metastasis; IS-mod, modified Immunoscore.

IS-like system in both cohorts. The scoring threshold we identified, based on the distribution of immune cell density in the discovery cohort, provided improved patient stratification for stage I–III colon cancer compared to the fixed, previously recommended threshold. We further demonstrated the validity of the IS-mod as an independent prognostic factor for patients with colon cancer from two centers.

With an improved understanding of the tumor-immune

system interaction, ongoing studies are responsible for revolution in the field of oncology by revealing advanced immunological biomarkers that have proved fruitful for prognostication (3,19–21). A potential prognostic marker, the “Immunoscore”, has been established to quantify the *in situ* immune infiltrate by quantifying CD3- and CD8-positive T cells at both the CT and IM regions. Regarding the prognostic and predictive power, the Immunoscore has shown complementary and superior performance to the

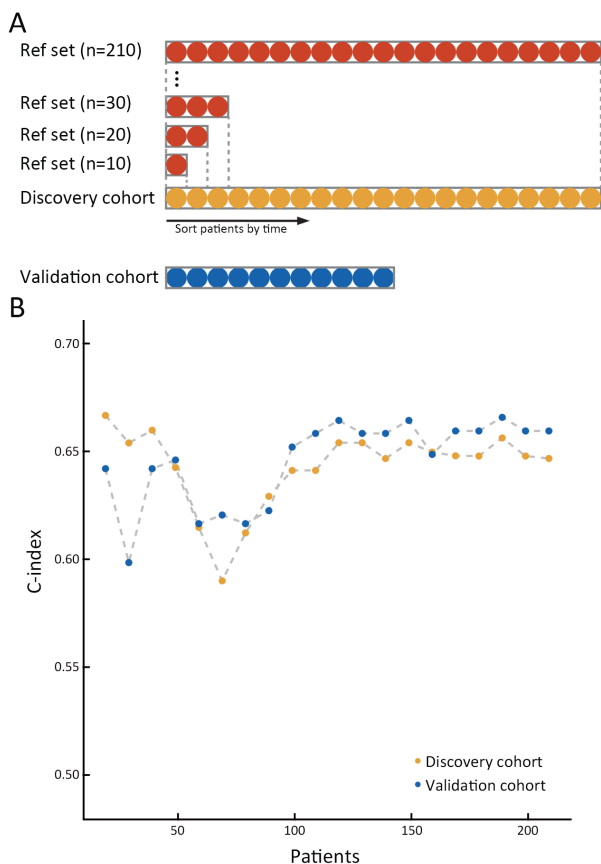


Figure 6 Sample size of reference set. (A) Patients in discovery cohort were arranged by time, and a series of reference sets were established by extracting patients at 10-patient intervals. The reference set serves as the immune cell density distributions reference. Patients in discovery cohort and validation cohort used the reference set to get the modified Immunoscore; (B) C-index values change with the increase of reference set sample size. Ref set, reference set.

gold-standard TNM classification for the prediction of survival outcomes in colon cancer. Despite these encouraging results, additional work is required before the clinical implementation of the Immunoscore. Firstly, the 48% of discordance found when comparing the Immunoscore to hematoxylin and eosin-based optical tumor-infiltrating lymphocytes evaluation calls for automated evaluation to ensure reproducibility. Compared with the Immunoscore that previous research mentioned (8,22), the IS-mod we proposed could enhance such biomarker's robustness and reproducibility. Firstly, instead of including the whole tumor region in WSIs, our annotation method merely requires the delimitation of two lines with a similar length. This method was expected to

ensure a more balanced density estimation in the two regions (CT and IM), with a high consistency between pathologists observed in our ICC analysis. In this way, our method simplified the annotation process and accelerated the calculation time, paving the way to enhancing clinical implementation. Secondly, we conducted consistency analysis on immune cells counting between that estimated by the automatic segmentation and that manually annotated by the pathologist, which was necessary but omitted in previous studies (3,8,19,22-24). The interclass consistency analysis and Bland-Altman plots demonstrated good consistency in the immune cell counting between the automatic segmentation and pathologist's annotation.

It is noteworthy that we used a simpler way of annotation, especially for the CT regions, where we drew a Line_{CT} to calculate the immune cell density, rather than taking the entire tumor region into account, as with the Immunoscore. According to consistency analysis, the band-shaped ROI can strongly reflect the tumor's overall immune status. Also, we compared the prognostic performance of the IS-mod and IS-like. We found that the IS-mod achieved comparable prognostic performance as the IS-like in both the discovery and validation cohorts. Simultaneously, our method only used a part of the whole tumor region to compute the immune cells density, thus speeding up the possibility of clinical implementation.

The IS-mod was established based on the distribution of immune cell density in this study's discovery cohort. Different thresholds were used to obtain a 3-category (37.8% and 61.8%) and 2-category (47.5%) IS-mod. When stratified into three categories using the 25% and 70% as thresholds reported in a previous study (8), the IS-mod yielded a barely satisfactory OS prediction (*Table 2, Supplementary Table S1*). When considered as a 2-category biomarker, the IS-mod showed better prognostic value using the 47.5% threshold compared with using the reported 25% threshold (C-index, 0.653 vs. 0.573, $P=0.004$).

Interestingly, our results also showed that the reference set of an IS-mod system might not require thousands of samples. Using a small data set ($N \geq 120$) as the reference set, as shown in our results, the IS-mod was demonstrated to be useful in predicting survival. The result suggests that an IS-mod system could be established based on comparably limited reference data sets in each institute, ensuring its feasibility in real clinical practice.

This study's limitations included the relatively small sample size in both cohorts and the retrospective nature of

the data collection. Besides, to validate our proposed approach, a publicly available CD3- and CD8-stained WSI data set for colon cancer are currently lacking. Moreover, it is necessary to develop deep learning-based automated analysis software, reducing pathologists' workload one step further.

Conclusions

We supplied evidence of a scoring system's prognostic potential that relies on quantifying lymphocyte populations to guide patient stratification in colon cancer. We established a modified Immunoscore system that will enhance reproducibility and eased it in clinical practice through rapid annotation and calculation. Given the prognostic validity demonstrated by the proposed IS-mod, observation of our study also suggests that the reference set of a prognostic IS-mod system may not require thousands of samples. Our findings may encourage more tailored investigation using a reference dataset of patients from a more homogeneous population regarding ethnic background.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (No. 2017YFC1309102), National Natural Science Foundation of China (No. 81771912, No. 82001986, No. 82071892), National Science Fund for Distinguished Young Scholars (No. 81925023), and High-level Hospital Construction Project (No. DFJH201805 and No. DFJH201914).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
2. Brierley JD, Gospodarowicz MK, Wittekind C. *TNM classification of malignant tumours*. 8th edition. Hoboken: Wiley, 2017.
3. Pagès F, Kirilovsky A, Mlecnik B, et al. *In situ* cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *J Clin Oncol* 2009;27:5944-51.
4. Galon J, Pagès F, Marincola FM, et al. The immune score as a new possible approach for the classification of cancer. *J Transl Med* 2012;10:1.
5. Wei ZW, Wu J, Huang WB, et al. Immune-infiltration based signature as a novel prognostic biomarker in gastrointestinal stromal tumour. *EBioMedicine* 2020;57:102850.
6. Chen CL, Chen CC, Yu WH, et al. An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning. *Nat Commun* 2021;12:1193.
7. Zhao K, Li Z, Yao S, et al. Artificial intelligence quantified tumour-stroma ratio is an independent predictor for overall survival in resectable colorectal cancer. *EBioMedicine* 2020;61:103054.
8. Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet* 2018;391:2128-39.
9. Reichling C, Taieb J, Derangere V, et al. Artificial intelligence-guided tissue analysis combined with immune infiltrate assessment predicts stage III colon cancer outcomes in PETACC08 study. *Gut* 2020;69:681-90.
10. Sun G, Dong X, Tang X, et al. The prognostic value of immunoscore in patients with colorectal cancer: A systematic review and meta-analysis. *Cancer Med* 2019;8:182-9.
11. Huijbers A, Tollenaar RA, v Pelt GW, et al. The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the VICTOR trial. *Ann Oncol* 2013;24:179-85.
12. National Health Commission of the People's Republic of China. National guidelines for diagnosis and treatment of colorectal cancer 2020 in China (English version). *Chin J Cancer Res* 2020;32:415-45.
13. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966;19:3-11.
14. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.
15. Harrell FE. *Regression Modeling Strategies: With*

- Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Cham: Springer International Publishing, 2015. Available online: <http://link.springer.com/10.1007/978-3-319-19425-7>
16. Uno H, Cai T, Lu T, et al. Evaluating prediction rules for t-year survivors with censored regression models. *J Am Stat Assoc* 2007;102:527-37.
 17. Allison DB, Paultre F, Maggio C, et al. The use of areas under curves in diabetes research. *Diabetes Care* 1995;18:245-50.
 18. Haibe-Kains B, Desmedt C, Sotiriou C, et al. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics* 2008; 24:2200-8.
 19. Laghi L, Bianchi P, Miranda E, et al. CD3+ cells at the invasive margin of deeply invading (pT3–T4) colorectal cancer and risk of post-surgical metastasis: a longitudinal study. *Lancet Oncol* 2009;10:877-84.
 20. Kather JN, Suarez-Carmona M, Charoentong P, et al. Topography of cancer-associated immune cells in human solid tumors. *ELife* 2018;7:e36967.
 21. Wang M, Yu L, Wei X, et al. Role of tumor gene mutations in treatment response to immune checkpoint blockades. *Precis Clin Med* 2019;2:100-9.
 22. Trabelsi M, Farah F, Zouari B, et al. An Immunoscore system based on CD3+ and CD8+ infiltrating lymphocytes densities to predict the outcome of patients with colorectal adenocarcinoma. *Onco Targets Ther* 2019;12:8663-73.
 23. Mlecnik B, Tosolini M, Kirilovsky A, et al. Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction. *J Clin Oncol* 2011;29:610-8.
 24. Lee H, Sha D, Foster NR, et al. Analysis of tumor microenvironmental features to refine prognosis by T, N risk group in patients with stage III colon cancer (NCCTG N0147) (Alliance). *Ann Oncol* 2020; 31:487-94.

Cite this article as: Zhao K, Wu X, Li Z, Wang Y, Xu Z, Li Y, Wu L, Yao S, Huang Y, Liang C, Liu Z. Prognostic value of a modified Immunoscore in patients with stage I–III resectable colon cancer. *Chin J Cancer Res* 2021;33(3):379-390. doi: 10.21147/j.issn.1000-9604.2021.03.09

Table S1 P values for comparison of IS-mod (3- and 2-category) and IS-mod using 25% and 70% as thresholds

Variables	Discovery cohort			Validation cohort		
	C-index	AUC at 5-year	iAUC	C-index	AUC at 5-year	iAUC
IS-mod (3-category) vs. IS-mod _{25%,70%}	0.029	0.007	<0.001	0.283	0.080	0.970
IS-mod (2-category) vs. IS-mod _{25%}	0.004	0.002	<0.001	0.021	0.012	<0.002

IS, Immunoscore; IS-mod, modified Immunoscore; IS-mod_{25%,70%}, 3-category IS-mod using 25% and 70% as thresholds; IS-mod_{25%}, 2-category IS-mod using 25% as threshold; AUC, area under the receiver operating characteristic curve; iAUC, integrated AUC.

Table S2 Discrimination ability of the reference and full (reference model + IS-mod) models in two cohorts

Variables	Discovery cohort		Validation cohort	
	Reference model	Full model	Reference model	Full model
C-index	0.709 (0.652–0.761)	0.752 (0.700–0.804)	0.705 (0.615–0.795)	0.759 (0.672–0.845)
AUC at 5-year	0.745 (0.675–0.814)	0.797 (0.735–0.860)	0.742 (0.644–0.840)	0.806 (0.717–0.894)
iAUC	0.746	0.775	0.767	0.788

AUC, area under the receiver operating characteristic curve; iAUC, integrated AUC.