Original Article
Assessing the clinical utility of tumor invasion proportion of lymph nodes for enhanced risk stratification in N1 colorectal cancer

Ru-Jie Chen[1*], Dong Xu[3*], Xiao-Yan Fan[2*], Yi-Huan Qiao[1], Xun-Jiang Jiang[1], Jun Hao[2], Yong-Tao Du[1], Xi-Hao Chen[1], Yuan Guo[4], Jun Zhu[1,5], Ji-Peng Li[1,2]

[1]Department of Digestive Surgery, Xijing Hospital of Digestive Diseases, Fourth Military Medical University, Xi'an, Shaanxi, China; [2]Department of Experiment Surgery, Xijing Hospital, Fourth Military Medical University, Xi'an, Shaanxi, China; [3]Division of Digestive Surgery, Xi'an International Medical Center Hospital of Digestive Diseases, Xi'an, Shaanxi, China; [4]Department of Breast, Shaanxi Provincial Cancer Hospital, Xi'an, Shaanxi, China; [5]Department of General Surgery, The Southern Theater Air Force Hospital, Guangzhou, Guangdong, China. *Equal contributors.

Abstract: N staging systems are paramount clinical features for colorectal cancer (CRC). In N1 stage (N1) CRC, patients present with a limited number of metastatic lymph nodes, yet their prognoses vary widely. The tumor invasion proportion of lymph nodes (TIPLN) has gained attention, but its prognostic value in N1 CRC remains unclear. We retrospectively analyzed 416 N1 CRC patients who underwent radical surgery from January 2014 to December 2018, reviewing 713 hematoxylin and eosin (H&E)-stained slides to assess TIPLN. Overall survival was the primary outcome in our study. Using restricted cubic splines, we explored the relationship between TIPLN and prognosis, with Cox regression and subgroup analyses adjusting for potential confounders. We found that increased TIPLN was associated with an unfavorable prognosis. At a cut-off value of 50%, patients with high-TIPLN exhibiting poorer outcomes than their low-TIPLN counterparts (hazard ratio: 3.77, $P < 0.001$). Furthermore, high-TIPLN was confirmed as an independent risk factor for overall survival (hazard ratio: 3.12, $P < 0.001$) after adjusting for clinical confounders. Notably, TIPLN could also enhance risk stratification within the T and N stages, where patients with low-risk (T1-3 stage) and high-TIPLN demonstrated poorer overall survival compared to those with high-risk (T4 stage) and low-TIPLN (hazard ratio: 2.54, $P = 0.021$). In conclusion, TIPLN is a promising prognostic indicator for N1 CRC patients that complements traditional N staging system for a more comprehensive evaluation. Integrating TIPLN into the TNM staging system could enhance risk stratification and guide treatment decisions.

Keywords: Colorectal cancer, N1 stage, prognostic implications, restricted cubic splines, tumor invasion proportion of lymph nodes

Introduction

Colorectal cancer (CRC) is the third most common malignant cancer and the third-leading cause of cancer-related mortality worldwide [1]. The prognosis of patients with CRC largely depends on the disease stage according to the American Joint Committee on Cancer (AJCC)/ Union for International Cancer Control (UICC) Tumor-Node-Metastasis (TNM) staging system [2, 3]. Lymph node (LN) status plays a pivotal role within the TNM staging system [4, 5] and is considered a pivotal factor for prognosis

evaluation and clinical decision-making [6, 7]. However, within the same N stage, patients with similar conditions may still exhibit substantial differences in their clinical outcomes [8-12].

The current N staging system primarily focuses on the number of metastatic lymph nodes (MLNs) while ignoring the extent of tumor cells invasion among MLNs. Simply labeling a LN as positive when tumor cells are present in a tumor-draining LN could result in significant loss of microscopic information. Wang et al. employed a distinctive method to extract prog-

nostic information from the dissected LNs of patients with gastric cancer [13]. They developed a novel prognostic indicator by calculating the ratio of tumor cell area over LN area. The results demonstrated that the tumor cell area ratio was increased with disease progression, and that a high-level area ratio was found to be significantly associated with an unfavorable prognosis.

However, there is still uncertainty about whether tumor invasion proportion of LNs (TIPLN) is related to prognosis in CRC patients. In CRC, any T stage accompanied by the presence of tumor cells in LNs is defined as late-stage disease, classified as stage III or IV [14]. Given the central importance of LN status in the staging of CRC, delving into the impact of TIPLN is of utmost importance. Hence, the assessment of TIPLN in CRC could serve as a valuable complement to LN status.

In CRC pathogenesis, current evidence posits a sequential model of metastatic spread [15, 16]. It has been proposed that LNs serve as a reservoir for tumor cells, which can then spread to distant organs [17]. Recently, Li et al. have reported a method for spatiotemporal quantification of metastatic CRC cell proliferation and distribution in LNs using ultimate Three-dimensional (3D) imaging of solvent-cleared organs and 3D rapid immunostaining [18]. Experimental research in animal models has shown that tumor cells proliferate and expand over time within LNs before further metastasis occurs. Exploring the distribution area ratio of metastatic tumor cells in LNs could provide a theoretical basis for further understanding the biological characteristics of CRC.

Thus, the aim of the present study was to assess the prognostic value of TIPLN in patients with CRC. The results could potentially offer novel evidence for the new category method of integrating TIPLN into the current N staging system and guide treatment decisions after curative resection.

Materials and methods

Study design and patients

According to the 8th edition of the AJCC TNM staging system [14], N1a (1 MLN) and N1b (2-3 MLNs) stage CRC are classified as the same TNM stage. To ensure cohort homogeneity and

reduce confounding, especially given the challenges of defining TIPLN values in patients with multiple lymph nodes, we focused on N1 CRC patients as our primary study population. The analysis workflow based on pathological slides is outlined in **Figure 1A**. First, hematoxylin and eosin (H&E)-stained slides of LNs were scanned into whole-slide images (WSIs). Next, we manually labeled WSIs and calculated the TIPLN for each patient. Finally, survival analysis was performed to investigate the relationship between TIPLN and the prognosis of the patients. The study was conducted and reported in accordance with the Strengthening the Reporting of Observational studies in Epidemiology statement [19].

In this retrospective study, 714 consecutive CRC patients with 1105 pathological slides were selected from patients who underwent radical resection between January 2014 and December 2018 in the Xijing Hospital. The inclusion criteria were as follows: (a) pathologically diagnosed with N1 CRC; (b) treated with curative surgery; (c) age older than 18 years; (d) complete clinical and pathological data. The exclusion criteria were as follows: (a) received neoadjuvant therapy; (b) absence of positive LNs; (c) presence of distant metastasis; (d) pathological slides exhibiting bubbles, tissue folds, or faded staining. The flow of participant selection and criteria is illustrated in **Figure 1B**. Finally, 416 CRC patients with 713 pathological slides were included. This study was approved by the Medical Ethics Committee of the First Affiliated Hospital of the Air Force Medical University (KY20212211-C-1).

Data collection and follow-up

The clinicopathological information was extracted from the electronic medical records, including age, sex, pathological characteristics of specimens (tumor size, T stage, N stage, and the number of LNs examined), serum albumin level, and tumor biomarkers [carcinoembryonic antigen (CEA), carbohydrate antigen 19-9 (CA19-9), and carbohydrate antigen 125 (CA125)]. CEA above 5 ng/ml indicates elevated levels, CA125 above 35 U/mL indicates elevated levels, and CA19-9 above 37 U/mL indicates elevated levels.

The primary endpoint of the study was overall survival (OS), calculated from the time of surgery until death from any cause or last follow-
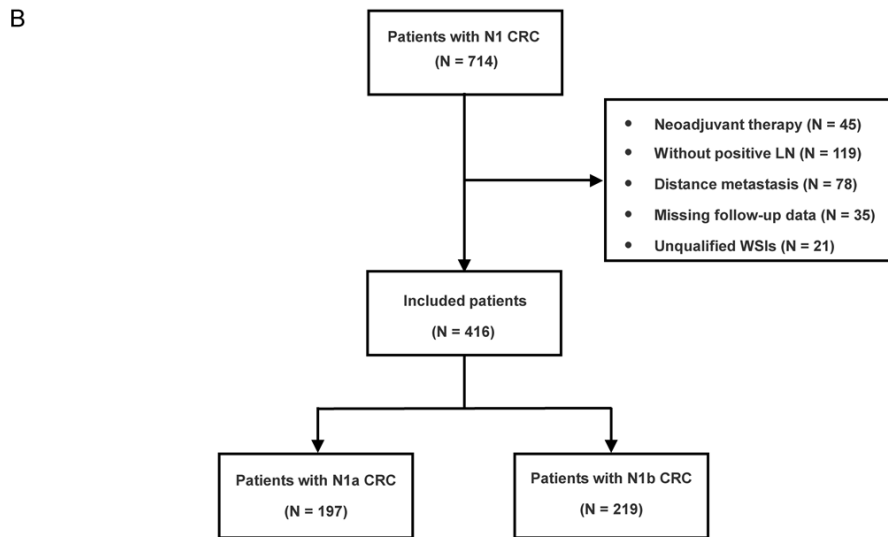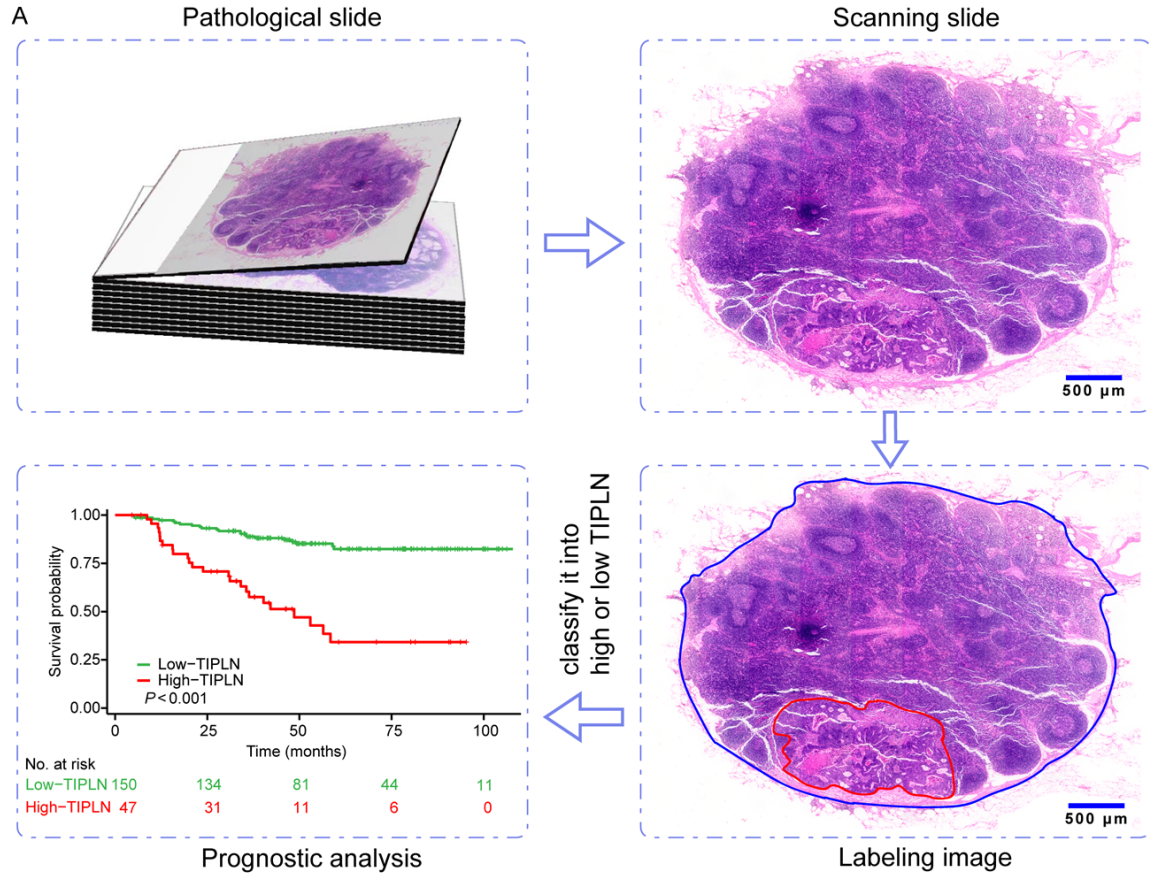
Figure 1. Flowchart and selection process of the study. A. H&E-stained pathological slides were first scanned to obtain WSIs. The WSIs were then labeled for calculating TIPLN. Finally, the TIPLNs were used to analyze the patient's prognosis. B. The flowchart of our study. H&E: hematoxylin and eosin; WSIs: whole-slide images; TIPLN: tumor invasion proportion of lymph nodes.

up. Follow-up was conducted every 3 months during the first year, and then every 6 months after surgery. The study was censored on 18 April 2023.

Assessment of TIPLN

The LN H&E-stained pathological slides were retrieved from the pathology archive. All slides

were first scanned for WSIs using the Olympus VS200 slide scanner at 20× magnification. Then, WSIs were analyzed by two experienced pathologists who were blinded to patient clinical information. The tumor and LN regions were meticulously annotated in different colors using OLYMPUS OlyVIA 3.3 software.

If the area annotations by the two senior pathologists differed by less than 10%, their average was used. For cases with discrepancies exceeding 10%, the WSIs were reviewed by a third senior pathologist, and a consensus result was reached through collaborative discussion.

The TIPLN for each MLN was calculated as the ratio of the tumor area to the total area of the MLN. The patient-level TIPLN was determined as the maximum TIPLN value among all MLNs for each patient.

Statistical analysis

Descriptive statistics were shown as frequencies and percentages (n, %) for categorical variables and median (interquartile range, IQR) for continuous variables. Comparisons were made using the Mann-Whitney U test and χ^2 tests when applicable. Survival curves and probabilities were estimated with the Kaplan-Meier (K-M) method and compared by log-rank test. Univariable and multivariable Cox regression proportional hazards models were applied to evaluate prognostic factors. The results were presented as hazard ratio (HR) with corresponding 95% confidence interval (CI). Harrell's concordance index (C-index) was used to evaluate the discriminative abilities of the prognostic models. The C-index value ranges from 0.5 to 1.0, with 0.5 indicating random predictions and 1.0 indicating perfect concordance. All analyses were implemented using R software (version 4.2.1), and a two-sided *P* value < 0.05 was considered statistically significant.

Results

Patient characteristics

A total of 714 patients diagnosed with N1 CRC participated in our study between January 2014 and December 2018, while 416 patients remained eligible for analysis, comprising 197 N1a patients and 219 N1b patients (**Figure 1B**). The main baseline clinical characteristics of the included patients are presented in Supplementary Table 1. The median age was 62 years (IQR, 53-69); 235 (56%) were male. Eight (1.9%) patients were T1 stage, 58 (14%) patients were T2 stage, 294 (71%) patients were T3 stage, and 56 (13%) patients were T4 stage. The median number of LNs retrieved was 16 (IQR, 13-19). The median follow-up time was 61 months (95% CI: 54-67). Estimated overall survival rate were 81.6% (95% CI: 77.9%-85.5%) at 3 years, and 70.6% (95% CI: 65.8%-75.7%) at 5 years.

Increased TIPLN was significantly associated with poor prognosis

A representative image of the annotated H&E-stained pathological slide is shown in **Figure 2A**. The LN region is outlined in blue, and the tumor region is outlined in red. The distribution of TIPLN values is plotted in **Figure 2B**. The median TIPLN were 18% (IQR, 4%-46%) and 39% (IQR, 23%-64%) in N1a and N1b patients, respectively.

Subsequently, we performed survival analysis based on TIPLN. A strongly negative correlation was found between the TIPLN and OS in patients with N1 CRC. In the multivariate Cox analysis, TIPLN showed an independent prognostic significance after adjusting for prognostic factors significant in univariable analyses, including T stage, albumin, CEA, CA19-9, and CA125 (HR: 1.02, *P* < 0.001; Supplementary Table 2).

A restrictive cubic spline (RCS) function was applied to present linear or nonlinear prognostic profiles of TIPLN. The RCS showed that the TIPLN presented a linear profile (nonlinearity P = 0.503, > 0.05) for the prognosis of CRC (**Figure 2C**), and we could further divide patients into high or low TIPLN groups.

Identification of the optimal cut-off value for TIPLN

To facilitate clinical application and evaluate the prognostic ability of TIPLN on survival, we stratified patients into high-TIPLN and low-TIPLN groups. The optimal cut-off point that best differentiated the survival outcome was identified using maximally selected log-rank statistics [20]. For the N1a, N1b, and overall cohorts, the optimal cut-off values for TIPLN
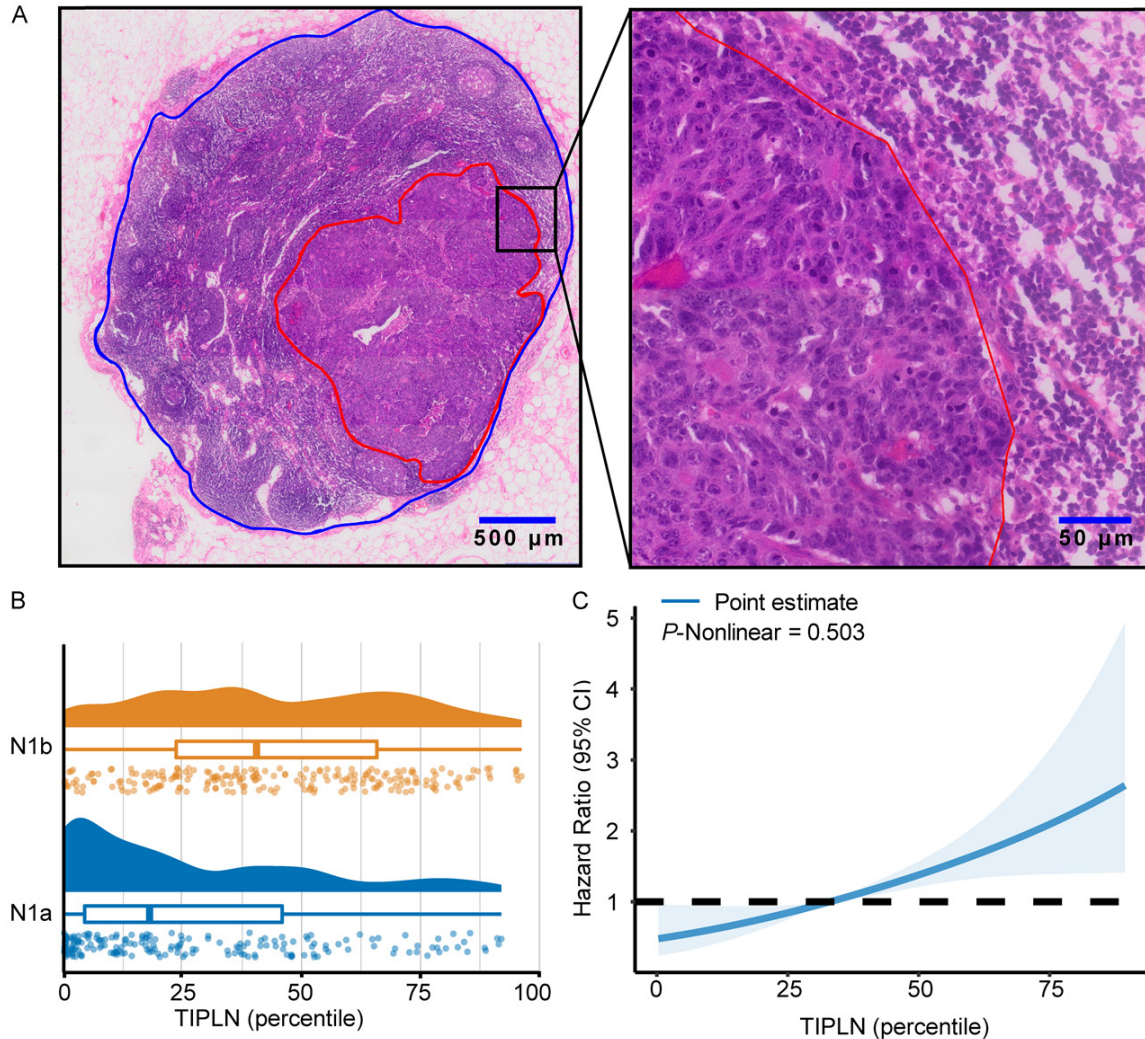
Figure 2. Association of TIPLN and overall risk of death in N1 patients. A. Representative image of the annotated WSIs for lymph node quantification (red for tumor tissue, and blue for lymph node tissue). B. Distribution of TIPLN in N1a and N1b patients. The central box signifies the interquartile range, and the line within the box indicates the median. C. Association of TIPLN with HR of over survival by restricted cubic splines. Knots for restricted cubic splines were placed at the 10th, 50th, and 90th percentiles of TIPLN. TIPLN: tumor invasion proportion of lymph nodes; HR: hazard ratio.

were set at 50%, 52%, and 50%, respectively (**Figure 3A-C**). Ultimately, we chose 50% as the uniform cut-off value for all patients.

Among the 416 patients with N1 CRC, 271 (65.1%) were classified into low-TIPLN group, while 145 (34.9%) were classified into high-TIPLN group. Their basic clinical characteristics are listed in **Table 1**. Patients in the low-TIPLN group exhibited higher albumin ($P = 0.037$) and lower levels of CEA ($P = 0.007$), CA19-9 ($P = 0.022$), and CA125 ($P = 0.009$) compared to those in the high-TIPLN. Additionally, no notable differences in age, sex, T stage, number of LNs

retrieved or tumor size were observed between the two groups.

The K-M survival analysis indicated that patients with high-TIPLN had a markedly inferior survival in both N1a (for high-TIPLN vs low-TIPLN, HR: 5.11, 95% CI: 2.85-9.16, $P < 0.001$; **Figure 3D**) and N1b stage (for high-TIPLN vs low-TIPLN, HR: 3.19, 95% CI: 1.86-5.47, $P < 0.001$; **Figure 3E**). However, when considering the entire N1 cohort, no significant disparity in OS was observed between patients in N1a and N1b (for N1b vs N1a, HR: 1.19, 95% CI: 0.81 to 1.75, $P = 0.364$; **Figure 3F**). Importantly, TIPLN
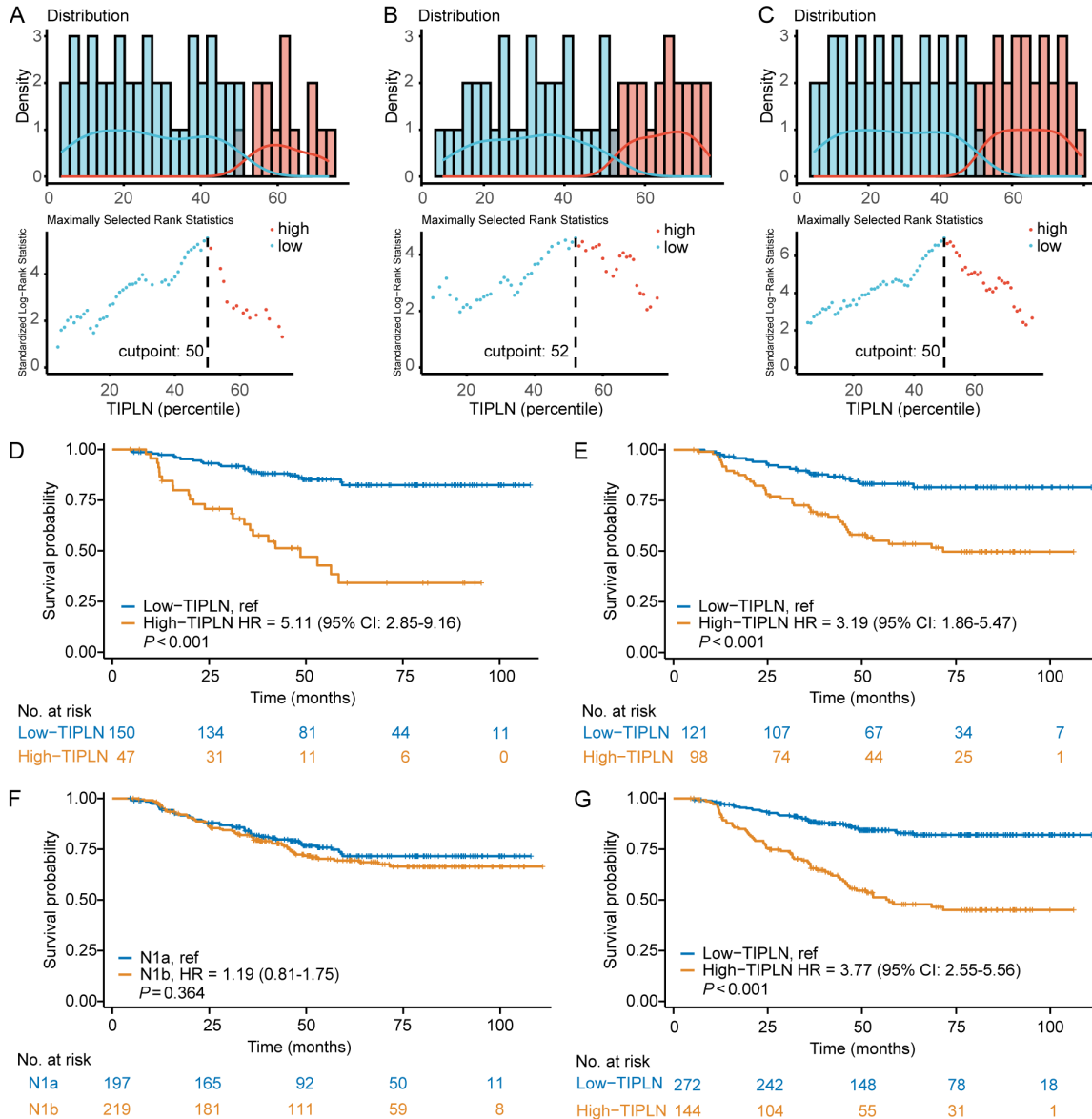
Figure 3. Assessment of prognostic performance of TIPLN. A-C. Determination of optimal cut-off values for TIPLN in N1a, N1b, and overall patient cohorts. D. K-M curve according to TIPLN in patients with N1a. E. K-M curve according to TIPLN in patients with N1b. F. K-M curve according to N stage in all cohorts of patients. G. K-M curve according to TIPLN in all cohorts of patients. TIPLN: tumor invasion proportion of lymph nodes; K-M: Kaplan-Meier.

remained a strong predictor of OS in N1 patients (for high-TIPLN vs low-TIPLN, HR: 3.77, 95% CI: 2.55-5.56, $P < 0.001$; **Figure 3G**). These results demonstrated that TIPLN is robust for N1 CRC prognostic stratification and superior to conventional N staging system.

TIPLN was an independent prognostic factor for N1 CRC patients

To validate the prognostic ability of TIPLN, we further performed Cox regression analyses. In

the univariate Cox analysis, T stage (HR: 2.09, $P = 0.021$), TIPLN (HR: 3.77, $P < 0.001$), albumin levels (HR: 2.26, $P < 0.001$), CEA levels (HR: 3.39, $P < 0.001$), CA19-9 levels (HR: 3.10, $P < 0.001$), and CA125 levels (HR: 2.42, $P < 0.001$) were found to have significantly associated with OS. After adjusting for clinicopathological variables that were significant in univariable analysis, the multivariate Cox analysis demonstrated that TIPLN continued to be an independent prognostic factor for OS (HR: 3.12, $P < 0.001$; **Table 2**). For internal validation, we

Table 1. Baseline characteristics in low- and high-TIPLN groups

Characteristic	Low-TIPLN (N = 271)	High-TIPLN (N = 145)	P value
Age, years	62.0 (53.0-69.0)	61.0 (54.0-67.0)	0.526
Sex			0.261
Male	159 (58.7%)	76 (52.4%)	
Female	112 (41.3%)	69 (47.6%)	
T stage			0.306
T1	3 (1.1%)	5 (3.4%)	
T2	40 (14.8%)	18 (12.4%)	
T3	194 (71.6%)	100 (69.0%)	
T4	34 (12.5%)	22 (15.2%)	
N stage			< 0.001
N1a	150 (55.4%)	47 (32.4%)	
N1b	121 (44.6%)	98 (67.6%)	
TNM stage			0.266
IIIA	43 (15.9%)	23 (15.9%)	
IIIB	227 (83.7%)	119 (82.0%)	
IIIC	1 (0.4%)	3 (2.1%)	
Size, cm			0.573
≤ 5	190 (70.1%)	97 (66.9%)	
> 5	81 (29.9%)	48 (33.1%)	
No. of lymph nodes retrieved	16.0 (13.0-19.0)	15.0 (12.0-18.0)	0.144
Albumin, g/L	44.4 (42.2-46.5)	43.7 (40.9-45.8)	0.037
CEA, ng/mL	2.3 (1.5-3.6)	2.8 (1.7-10.8)	0.007
CA19-9, U/mL	11.3 (7.1-20.1)	12.9 (7.9-30.9)	0.022
CA125, U/mL	10.6 (8.2-14.6)	12.0 (8.6-16.8)	0.009

Data are median (IQR) or n (%). TIPLN: tumor invasion proportion of lymph nodes; CEA: carcinoembryonic antigen; CA19-9: carbohydrate antigen 19-9; CA125: carbohydrate antigen 125.

employed 1000 bootstrap repetitions. The bias-corrected C-index was 0.732 (95% CI: 0.663-0.799).

Notably, in the subgroup analysis of OS, the HR between patients with high-TIPLN and low-TIPLN was generally consistent across all baseline characteristic subgroups, indicating that TIPLN provided a decent prediction performance for patients with N1 CRC (**Figure 4**).

Facilitate prognostic risk stratification for N1 CRC patients by TIPLN

In accordance with the National Comprehensive Cancer Network (NCCN) guideline, patients with stage III CRC are stratified into low-risk (T1-3, N1) and high-risk (T4 or N2) groups to determine the duration of adjuvant chemotherapy [21]. Recognizing the importance of this stratification in clinical practice, our study adopted a similar strategy, focusing on subgroup analyses

based on the risk categories. Interestingly, TIPLN proved to be a significant predictor in low-risk (T1-3 stage) group (for high-TIPLN vs low-TIPLN, HR: 3.50, 95% CI: 2.27-5.39, *P* < 0.001; **Figure 5A**) and high-risk (T4 stage) group (for high-TIPLN vs low-TIPLN, HR: 5.19, 95% CI: 2.06-13.09, *P* < 0.001; **Figure 5B**). Contrastingly, the OS of the patients with low-risk but high-TIPLN was even worse than the patients with high-risk but low-TIPLN (HR: 2.54, 95% CI: 1.15-5.60, *P* = 0.021). Additionally, no statistically significant difference in OS was observed between the two risk groups among patients with low-TIPLN (HR: 1.39, 95% CI: 0.62-3.14, *P* = 0.425; **Figure 5C**).

Relationship between TIPLN and other lymph node indexes

Finally, we investigated the relationship of TIPLN with other proposed prognostic indicators like lymph node ratio (LNR) and log odds of

Table 2. Univariable and multivariable Cox analysis of categorized TIPLN

Variable	Univariate COX analysis		Multivariate COX analysis[a]	
	HR (95% CI)	P value	HR (95% CI)	P value
Age (> 60 vs ≤ 60 years)	1.10 (0.75-1.61)	0.625	-	-
Sex (female vs male)	1.13 (0.77-1.65)	0.529	-	-
T stage (T3-4 vs T1-2)	2.09 (1.12-3.89)	0.021	1.90 (1.00-3.60)	0.049
N stage (N1b vs N1a)	1.19 (0.81-1.75)	0.364	-	-
Size (> 5 vs ≤ 5 cm)	1.28 (0.87-1.91)	0.214	-	-
TIPLN (high vs low)	3.77 (2.55-5.56)	< 0.001	3.00 (1.99-4.50)	< 0.001
Albumin (< 40 vs ≥ 40 g/L)	2.26 (1.55-3.30)	< 0.001	1.89 (1.28-2.78)	0.001
CEA (> 5 vs ≤ 5 ng/mL)	3.39 (2.27-5.07)	< 0.001	2.15 (1.35-3.42)	0.001
CA19-9 (> 37 vs ≤ 37 U/mL)	3.10 (2.05-4.69)	< 0.001	1.42 (0.87-2.33)	0.162
CA125 (> 35 vs ≤ 35 U/mL)	2.42 (1.35-4.32)	0.003	1.20 (0.66-2.20)	0.545

[a]Only significant prognostic factors in the univariate analysis were included in the multivariate model. TIPLN: tumor invasion proportion of lymph nodes; HR: hazard ratio; CI: confidence interval; CEA: carcinoembryonic antigen; CA19-9: carbohydrate antigen 19-9; CA125: carbohydrate antigen 125.
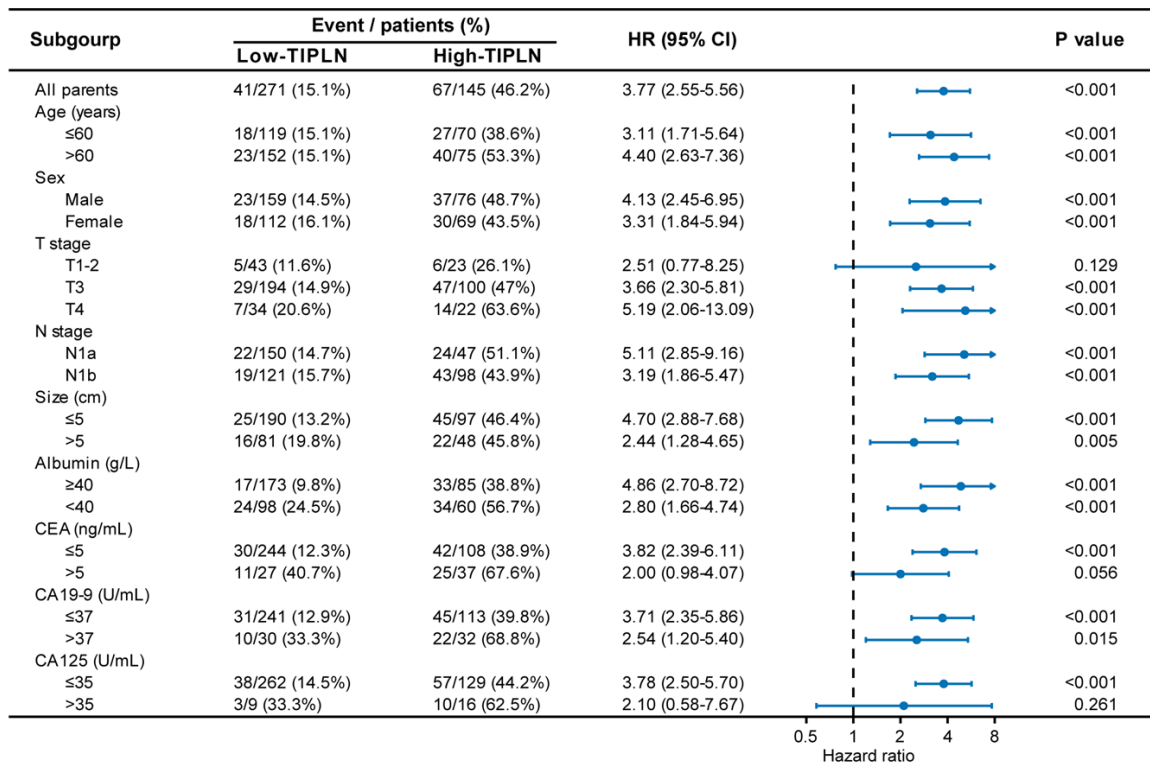
Subgourp	Event / patients (%)		HR (95% CI)		P value
	Low-TIPLN	High-TIPLN			
All parents	41/271 (15.1%)	67/145 (46.2%)	3.77 (2.55-5.56)		<0.001
Age (years)					
≤60	18/119 (15.1%)	27/70 (38.6%)	3.11 (1.71-5.64)		<0.001
>60	23/152 (15.1%)	40/75 (53.3%)	4.40 (2.63-7.36)		<0.001
Sex					
Male	23/159 (14.5%)	37/76 (48.7%)	4.13 (2.45-6.95)		<0.001
Female	18/112 (16.1%)	30/69 (43.5%)	3.31 (1.84-5.94)		<0.001
T stage					
T1-2	5/43 (11.6%)	6/23 (26.1%)	2.51 (0.77-8.25)		0.129
T3	29/194 (14.9%)	47/100 (47%)	3.66 (2.30-5.81)		<0.001
T4	7/34 (20.6%)	14/22 (63.6%)	5.19 (2.06-13.09)		<0.001
N stage					
N1a	22/150 (14.7%)	24/47 (51.1%)	5.11 (2.85-9.16)		<0.001
N1b	19/121 (15.7%)	43/98 (43.9%)	3.19 (1.86-5.47)		<0.001
Size (cm)					
≤5	25/190 (13.2%)	45/97 (46.4%)	4.70 (2.88-7.68)		<0.001
>5	16/81 (19.8%)	22/48 (45.8%)	2.44 (1.28-4.65)		0.005
Albumin (g/L)					
≥40	17/173 (9.8%)	33/85 (38.8%)	4.86 (2.70-8.72)		<0.001
<40	24/98 (24.5%)	34/60 (56.7%)	2.80 (1.66-4.74)		<0.001
CEA (ng/mL)					
≤5	30/244 (12.3%)	42/108 (38.9%)	3.82 (2.39-6.11)		<0.001
>5	11/27 (40.7%)	25/37 (67.6%)	2.00 (0.98-4.07)		0.056
CA19-9 (U/mL)					
≤37	31/241 (12.9%)	45/113 (39.8%)	3.71 (2.35-5.86)		<0.001
>37	10/30 (33.3%)	22/32 (68.8%)	2.54 (1.20-5.40)		0.015
CA125 (U/mL)					
≤35	38/262 (14.5%)	57/129 (44.2%)	3.78 (2.50-5.70)		<0.001
>35	3/9 (33.3%)	10/16 (62.5%)	2.10 (0.58-7.67)		0.261

0.5 1 2 4 8
Hazard ratio

Figure 4. Forest plot of overall survival according to subgroup analyses comparing patients with high-TIPLN and low-TIPLN. The 95% CI of the hazard ratio in each subgroup-analysis is depicted by a horizontal line crossing a dot. *P* values were calculated using two-sided log-rank test. TIPLN: tumor invasion proportion of lymph nodes.

positive lymph nodes (LODDS). We found that there was positive relationship between TIPLN and LNR (Supplementary Figure 1A) and LODDS (Supplementary Figure 1B). Additionally, high-TIPLN group had a more level of LNR and LODDS compared with low-TIPLN group (Supplementary Figure 1C, 1D).

Discussion

In this study, we have identified a histopathology criterion to predict the outcome of patients with N1 stage CRC based on the proportion of tumor cells invading in MLN, which directly analyzes pathological slides stained with H&E.
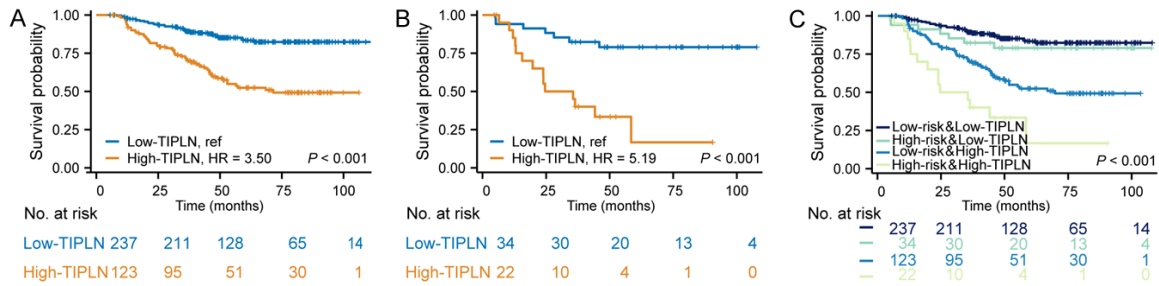
Figure 5. Prognostic implications of TIPLN in the low- and high-risk patients. K-M curves according to the TIPLN in Low-risk (A), High-risk (B), and combined (C) patient cohorts. TIPLN: tumor invasion proportion of lymph nodes; K-M: Kaplan-Meier.

Encouragingly, TIPLN demonstrated a favorable capacity to stratify patients in terms of their oncological prognosis. In accordance with the stratified analysis results, TIPLN was a robust predictor independent of clinical features and tumor stage, suggesting that TIPLN could be a valuable supplement to the existing markers and enhance the accuracy of risk stratification.

Nodal metastasis is a key component of the TNM staging system and acts as a powerful prognostic indicator for CRC patients without distant metastasis [22]. However, recent studies have raised concerns regarding the reliability of the traditional N staging scheme, which solely relies on the number of MLN [23, 24]. Various factors, including LNR [25-27], LODDS [28, 29], and tumor deposits [30, 31] have been proposed to supplement the current N staging system. While the predictive value of LNR and LODDS for CRC prognosis has been confirmed in previous studies, it is worth considering that the number of LNs examined can vary widely among patients due to the differences in surgical quality and techniques, thereby posing challenges in establishing a consensus on appropriate cut-off values [32, 33].

Additionally, the classifications mentioned above plainly omit potentially relevant biological details. As neoplasms progress through their evolution, they acquire the capacity for metastasis [34-36]. The involvement of LNs holds significant prognostic value, not only due to its indication of aggressive tumor biology, but also because persistent disease within the LNs can serve as a mediator for subsequent distant and lethal metastases [37]. The prognostic value of quantifying tumor cells in LNs has not been adequately assessed. Prior research has demonstrated that the ratio of tumor area to lymph node area is an independent predictor of prognosis for gastric cancer patients, especially at the N1 stage [13]. The findings are largely consistent with our study.

Clinically, tumor stage is used to inform discussions on adjuvant treatment options in stage II and III CRC. Recent studies have demonstrated that a 3-month treatment regimen may yield comparable survival outcomes to a 6-month treatment regimen, with the added benefit of diminished toxicity. Extended therapy may be advantageous for patients at high risk [38, 39]. Accordingly, the current study offers valuable insights for the further classification of N1 CRC patients into distinct risk groups. We found that TIPLN could identify patients with poor prognosis within both low-risk (patients with T1-3) and high-risk (patients with T4) groups. Unexpectedly, among the patients with low-TIPLN, there was no statistically significant difference in OS between low-risk and high-risk subgroups, while low-risk patients with high-TIPLN had a poorer prognosis than high-risk patients with low-TIPLN. Considering the demonstrated ability of TIPLN to stratify the prognostic risk in CRC patients, there is potential for combining TIPLN with current guidelines to assist clinicians in making more informed decisions regarding adjuvant chemotherapy regimens.

The mechanism underlying this observation has yet to be elucidated. Previous research has reported a tissue 3D imaging technique capable of quantifying the number of tumor cells within LNs. The results indicate that tumor cells exhibit a gradual accumulation of slow growth in the early stages, followed by a rapid expansion in the post-adaptation phase, which prepares them for distant metastasis [18]. As we

know, cancer cells infiltrate lymphatic capillaries and are passively transported to collecting lymphatic vessels. From there, they migrate to the sentinel LN (SLN - the first LN to which cancer cells spread from the primary tumor) and eventually enter the bloodstream through the subclavian vein. Prior to reaching the SLN, cancer cells secrete specific soluble factors to modulate its microenvironment (TiME), thereby establishing a favorable niche for successful colonization. Once colonized, cancer cells suppress anti-tumor immunity by recruiting regulatory T cells and myeloid-derived suppressor cells, inhibiting the activity of dendritic cells and CD8+ T cells, and promoting the release of immunosuppressive cytokines [40]. Recent results found LN metastases resist T cell-mediated cytotoxicity, induce antigen-specific regulatory T cells, and generate tumor-specific immune tolerance that subsequently facilitates distant tumor colonization [41, 42]. In our study, high ratio TIPLN might indicate more immunosuppressive TiME and greater invasive and metastasis ability. Here, we hypothesized that TIPLN functions as a staging signal, reflecting the progression of the tumor. Metastasis is more likely to occur when the accumulated invaded area of the tumor reaches a certain threshold. Alternatively, the quantity of tumor cells might reflect their aggressiveness and metastatic potential, thereby indicating a positive correlation between TIPLN and metastatic burden.

In node-positive CRC, tumor deposits (TD) and positive LNR have been demonstrated as independent prognostic predictors. The combination of TD and LNR could be used to identify them at high-risk of CRC deaths [43]. In addition, tumor microenvironment has been a robust prognostic factor for node-positive CRC. Recently, Hye-Yeong Jin et al. explored combinatory statuses of tumor-stromal percentage (TSP) and tumor-infiltrating lymphocytes (TILs) and found patients with low CD8+ TILs or high TSP, that is high tumor invasion, had worse prognoses [44]. Similar with their results, we found that more ratio tumor invasion proportion of lymph nodes (TIPLN) represented worse clinical outcomes. Another common prognostic factor, such as immune and nutritional index also investigated by many researchers [45, 46]. Compared with patients with a high immune and nutritional index, those with low levels had worse disease-free survival and OS [46].

Some limitations of the current study should be acknowledged. First, as a single-center retrospective study, the results might be subject to certain confounding factors. External validation in a large-scale and multicentric cohort is necessary to confirm the prognostic value of TIPLN. There are considerable differences in lymph node pathology slides among different hospitals. We have taken this into account repeatedly and, therefore, proposed a simple cutoff value for differentiation. Specifically, we use the tumor-to-lymph node area ratio: if this ratio exceeds 50%, we classify it as a high-risk group, which generally correlates with poorer prognosis. This simple and practical metric helps reduce variability between hospitals. Second, the study did not account for the location of the MLNs, which could introduce another potential source of bias. Furthermore, the impact of different chemotherapy regimens has not been adequately assessed; prospective, randomized controlled trials will be needed in the future.

In conclusion, TIPLN emerges as a robust, independent prognostic indicator for patients with N1 CRC, refining risk stratification within the TNM staging system. With an increasing body of research in the future, we anticipate the integration of TIPLN into the N staging system, thereby improving treatment decisions and outcomes for patients.

Disclosure of conflict of interest

None.

Address correspondence to: Ji-Peng Li and Jun Zhu, Department of Digestive Surgery, Xijing Hospital of Digestive Diseases, Fourth Military Medical University, No. 127 Changlexi Road, Xi'an 710032, Shaanxi, China. Tel: +86-29-84771509; E-mail: jipengli1974@aliyun.com; hy1977@fmmu.edu.cn (JPL); zjsty@fmmu.edu.cn (JZ)

References

[1]   Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global

cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[2] Edge SB and Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 2010; 17: 1471-1474.

[3] Shah MA, Renfro LA, Allegra CJ, André T, de Gramont A, Schmoll HJ, Haller DG, Alberts SR, Yothers G and Sargent DJ. Impact of patient factors on recurrence risk and time dependency of oxaliplatin benefit in patients with colon cancer: analysis from modern-era adjuvant studies in the Adjuvant Colon Cancer End Points (ACCENT) Database. J Clin Oncol 2016; 34: 843-853.

[4] Kim HJ and Choi GS. Clinical implications of lymph node metastasis in colorectal cancer: current status and future perspectives. Ann Coloproctol 2019; 35: 109-117.

[5] Ebbehøj AL, Smith HG, Jørgensen LN and Krarup PM. Prognostic factors for lymph node metastases in pT1 colorectal cancer differ according to tumor morphology: a nationwide cohort study. Ann Surg 2023; 277: 127-135.

[6] Greene FL. Current TNM staging of colorectal cancer. Lancet Oncol 2007; 8: 572-573.

[7] Hutter RV. At last--worldwide agreement on the staging of cancer. Arch Surg 1987; 122: 1235-1239.

[8] André T, Meyerhardt J, Iveson T, Sobrero A, Yoshino T, Souglakos I, Grothey A, Niedzwiecki D, Saunders M, Labianca R, Yamanaka T, Boukovinas I, Vernerey D, Meyers J, Harkin A, Torri V, Oki E, Georgoulias V, Taieb J, Shields A and Shi Q. Effect of duration of adjuvant chemotherapy for patients with stage III colon cancer (IDEA collaboration): final results from a prospective, pooled analysis of six randomised, phase 3 trials. Lancet Oncol 2020; 21: 1620-1629.

[9] Dimitriou N, Arandjelović O, Harrison DJ and Caie PD. A principled machine learning framework improves accuracy of stage II colorectal cancer prognosis. NPJ Digit Med 2018; 1: 52.

[10] Skrede OJ, De Raedt S, Kleppe A, Hveem TS, Liestøl K, Maddison J, Askautrud HA, Pradhan M, Nesheim JA, Albregtsen F, Farstad IN, Domingo E, Church DN, Nesbakken A, Shepherd NA, Tomlinson I, Kerr R, Novelli M, Kerr DJ and Danielsen HE. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. Lancet 2020; 395: 350-360.

[11] Kleppe A, Skrede OJ, De Raedt S, Hveem TS, Askautrud HA, Jacobsen JE, Church DN, Nesbakken A, Shepherd NA, Novelli M, Kerr R, Liestøl K, Kerr DJ and Danielsen HE. A clinical decision support system optimising adjuvant chemotherapy for colorectal cancers by integrating deep learning and pathological staging markers: a development and validation study. Lancet Oncol 2022; 23: 1221-1232.

[12] Tian Y, Wang X, Wu C, Qiao J, Jin H and Li H. A protracted war against cancer drug resistance. Cancer Cell Int 2024; 24: 326.

[13] Wang X, Chen Y, Gao Y, Zhang H, Guan Z, Dong Z, Zheng Y, Jiang J, Yang H, Wang L, Huang X, Ai L, Yu W, Li H, Dong C, Zhou Z, Liu X and Yu G. Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning. Nat Commun 2021; 12: 1637.

[14] Weiser MR. AJCC 8th edition: colorectal cancer. Ann Surg Oncol 2018; 25: 1454-1455.

[15] Chen HN, Shu Y, Liao F, Liao X, Zhang H, Qin Y, Wang Z, Luo M, Liu Q, Xue Z, Cao M, Zhang S, Zhang WH, Hou Q, Xia X, Luo H, Zhang Y, Yang L, Hu JK, Fu X, Liu B, Hu H, Huang C, Peng Y, Cheng W, Dai L, Yang L, Zhang W, Dong B, Li Y, Wei Y, Xu H and Zhou ZG. Genomic evolution and diverse models of systemic metastases in colorectal cancer. Gut 2022; 71: 322-332.

[16] Ulintz PJ, Greenson JK, Wu R, Fearon ER and Hardiman KM. Lymph node metastases in colon cancer are polyclonal. Clin Cancer Res 2018; 24: 2214-2224.

[17] Naxerova K, Reiter JG, Brachtel E, Lennerz JK, van de Wetering M, Rowan A, Cai T, Clevers H, Swanton C, Nowak MA, Elledge SJ and Jain RK. Origins of lymphatic and distant metastases in human colorectal cancer. Science 2017; 357: 55-60.

[18] Li J, Xu CJ, Tian GA, Li Q, Li DX, Yan F, Zhou YQ, Huang PQ, Xie JX, Wang X, Jiang SH, Wang YH, Song J, Zhang XL, Yi SQ, Hu LP, Xu Q, Li XW and Zhang ZG. Spatiotemporal quantification of metastatic tumour cell growth and distribution in lymph nodes by whole-mount tissue 3D imaging. Int J Biol Sci 2022; 18: 3993-4005.

[19] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC and Vandenbroucke JP; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007; 335: 806-808.

[20] Alcala N, Leblay N, Gabriel AAG, Mangiante L, Hervas D, Giffon T, Sertier AS, Ferrari A, Derks J, Ghantous A, Delhomme TM, Chabrier A, Cuenin C, Abedi-Ardekani B, Boland A, Olaso R, Meyer V, Altmuller J, Le Calvez-Kelm F, Durand G, Voegele C, Boyault S, Moonen L, Lemaitre N, Lorimier P, Toffart AC, Soltermann A, Clement JH, Saenger J, Field JK, Brevet M, Blanc-Fournier C, Galateau-Salle F, Le Stang N, Russell PA, Wright G, Sozzi G, Pastorino U, Lacomme S, Vignaud JM, Hofman V, Hofman P, Brustugun OT, Lund-Iversen M, Thomas de Montpreville V, Muscarella LA, Graziano P,

Popper H, Stojsic J, Deleuze JF, Herceg Z, Viari A, Nuernberg P, Pelosi G, Dingemans AMC, Milione M, Roz L, Brcic L, Volante M, Papotti MG, Caux C, Sandoval J, Hernandez-Vargas H, Brambilla E, Speel EJM, Girard N, Lantuejoul S, McKay JD, Foll M and Fernandez-Cuesta L. Integrative and comparative genomic analyses identify clinically relevant pulmonary carcinoid groups and unveil the supra-carcinoids. Nat Commun 2019; 10: 3407.

[21] Benson AB, Venook AP, Al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK, Cohen S, Cooper HS, Deming D, Engstrom PF, Garrido-Laguna I, Grem JL, Grothey A, Hochster HS, Hoffe S, Hunt S, Kamel A, Kirilcuk N, Krishnamurthi S, Messersmith WA, Meyerhardt J, Miller ED, Mulcahy MF, Murphy JD, Nurkin S, Saltz L, Sharma S, Shibata D, Skibber JM, Sofocleous CT, Stoffel EM, Stotsky-Himelfarb E, Willett CG, Wuthrick E, Gregory KM and Freedman-Cass DA. NCCN guidelines insights: colon cancer, version 2.2018. J Natl Compr Canc Netw 2018; 16: 359-369.

[22] Nagtegaal ID and Schmoll HJ. Colorectal cancer: what is the role of lymph node metastases in the progression of colorectal cancer? Nat Rev Gastroenterol Hepatol 2017; 14: 633-634.

[23] Veronese N, Nottegar A, Pea A, Solmi M, Stubbs B, Capelli P, Sergi G, Manzato E, Fassan M, Wood LD, Scarpa A and Luchini C. Prognostic impact and implications of extracapsular lymph node involvement in colorectal cancer: a systematic review with meta-analysis. Ann Oncol 2016; 27: 42-48.

[24] Nagtegaal ID, Knijn N, Hugen N, Marshall HC, Sugihara K, Tot T, Ueno H and Quirke P. Tumor deposits in colorectal cancer: improving the value of modern staging-a systematic review and meta-analysis. J Clin Oncol 2017; 35: 1119-1127.

[25] Rosenberg R, Engel J, Bruns C, Heitland W, Hermes N, Jauch KW, Kopp R, Pütterich E, Ruppert R, Schuster T, Friess H and Hölzel D. The prognostic value of lymph node ratio in a population-based collective of colorectal cancer patients. Ann Surg 2010; 251: 1070-1078.

[26] Parnaby CN, Scott NW, Ramsay G, MacKay C, Samuel L, Murray GI and Loudon MA. Prognostic value of lymph node ratio and extramural vascular invasion on survival for patients undergoing curative colon cancer resection. Br J Cancer 2015; 113: 212-219.

[27] Berger AC, Sigurdson ER, LeVoyer T, Hanlon A, Mayer RJ, Macdonald JS, Catalano PJ and Haller DG. Colon cancer survival is associated with decreasing ratio of metastatic to examined lymph nodes. J Clin Oncol 2005; 23: 8706-8712.

[28] Li T, Yang Y, Wu W, Fu Z, Cheng F, Qiu J, Li Q, Zhang K, Luo Z, Qiu Z and Huang C. Prognostic implications of ENE and LODDS in relation to lymph node-positive colorectal cancer location. Transl Oncol 2021; 14: 101190.

[29] Zhou YY, Du XJ, Zhang CH, Aparicio T, Zaanan A, Afchain P, Chen LP, Hu SK, Zhang PC, Wu M, Zhang QW and Wang H. Comparison of three lymph node staging schemes for predicting the outcome in patients with small bowel adenocarcinoma: a population-based cohort and international multicentre cohort study. EBioMedicine 2019; 41: 276-285.

[30] Agger E, Jörgren F, Jöud A, Lydrup ML and Buchwald P. Negative prognostic impact of tumor deposits in rectal cancer: a national study cohort. Ann Surg 2023; 278: e526-e533.

[31] Cohen R, Shi Q, Meyers J, Jin Z, Svrcek M, Fuchs C, Couture F, Kuebler P, Ciombor KK, Bendell J, De Jesus-Acosta A, Kumar P, Lewis D, Tan B, Bertagnolli MM, Philip P, Blanke C, O'Reilly EM, Shields A and Meyerhardt JA. Combining tumor deposits with the number of lymph node metastases to improve the prognostic accuracy in stage III colon cancer: a post hoc analysis of the CALGB/SWOG 80702 phase III study (Alliance)☆. Ann Oncol 2021; 32: 1267-1275.

[32] Jakob MO, Guller U, Ochsner A, Oertli D, Zuber M and Viehl CT. Lymph node ratio is inferior to pN-stage in predicting outcome in colon cancer patients with high numbers of analyzed lymph nodes. BMC Surg 2018; 18: 81.

[33] Mohan HM, Walsh C, Kennelly R, Ng CH, O'Connell PR, Hyland JM, Hanly A, Martin S, Gibbons D, Sheahan K and Winter DC. The lymph node ratio does not provide additional prognostic information compared with the N1/N2 classification in stage III colon cancer. Colorectal Dis 2017; 19: 165-171.

[34] Lambert AW, Pattabiraman DR and Weinberg RA. Emerging biological principles of metastasis. Cell 2017; 168: 670-691.

[35] Aparicio S and Caldas C. The implications of clonal genome evolution for cancer medicine. N Engl J Med 2013; 368: 842-851.

[36] Guo L, Wang Y, Yang W, Wang C, Guo T, Yang J, Shao Z, Cai G, Cai S, Zhang L, Hu X and Xu Y. Molecular profiling provides clinical insights into targeted and immunotherapies as well as colorectal cancer prognosis. Gastroenterology 2023; 165: 414-428, e417.

[37] Kawada K and Taketo MM. Significance and mechanism of lymph node metastasis in cancer progression. Cancer Res 2011; 71: 1214-1218.

[38] Iveson TJ, Kerr RS, Saunders MP, Cassidy J, Hollander NH, Tabernero J, Haydon A, Glimelius B, Harkin A, Allan K, McQueen J, Scudder C, Boyd KA, Briggs A, Waterston A, Medley L, Wilson C, Ellis R, Essapen S, Dhadda AS, Harrison M, Falk S, Raouf S, Rees C, Olesen

RK, Propper D, Bridgewater J, Azzabi A, Farrugia D, Webb A, Cunningham D, Hickish T, Weaver A, Gollins S, Wasan HS and Paul J. 3 versus 6 months of adjuvant oxaliplatin-fluoropyrimidine combination therapy for colorectal cancer (SCOT): an international, randomised, phase 3, non-inferiority trial. Lancet Oncol 2018; 19: 562-578.

[39] Grothey A, Sobrero AF, Shields AF, Yoshino T, Paul J, Taieb J, Souglakos J, Shi Q, Kerr R, Labianca R, Meyerhardt JA, Vernerey D, Yamanaka T, Boukovinas I, Meyers JP, Renfro LA, Niedzwiecki D, Watanabe T, Torri V, Saunders M, Sargent DJ, Andre T and Iveson T. Duration of adjuvant chemotherapy for stage III colon cancer. N Engl J Med 2018; 378: 1177-1188.

[40] Li YL and Hung WC. Reprogramming of sentinel lymph node microenvironment during tumor metastasis. J Biomed Sci 2022; 29: 84.

[41] Reticker-Flynn NE, Zhang W, Belk JA, Basto PA, Escalante NK, Pilarowski GOW, Bejnood A, Martins MM, Kenkel JA, Linde IL, Bagchi S, Yuan R, Chang S, Spitzer MH, Carmi Y, Cheng J, Tolentino LL, Choi O, Wu N, Kong CS, Gentles AJ, Sunwoo JB, Satpathy AT, Plevritis SK and Engleman EG. Lymph node colonization induces tumor-immune tolerance to promote distant metastasis. Cell 2022; 185: 1924-1942, e1923.

[42] Peng JM and Su YL. Lymph node metastasis and tumor-educated immune tolerance: potential therapeutic targets against distant metastasis. Biochem Pharmacol 2023; 215: 115731.

[43] Liu L, Ji J, Ge X, Ji Z, Li J, Wu J, Zhu J, Yao J, Zhu F, Mao B, Cao Z, Zhou J, Miao L, Ji G and Hang D. Prognostic value of tumor deposits and positive lymph node ratio in stage III colorectal cancer: a retrospective cohort study. Int J Surg 2024; 110: 3470-3479.

[44] Jin HY, Yoo SY, Lee JA, Wen X, Kim Y, Park HE, Kwak Y, Cho NY, Bae JM, Kim JH, Lee HS and Kang GH. Combinatory statuses of tumor stromal percentage and tumor infiltrating lymphocytes as prognostic factors in stage III colorectal cancers. J Gastroenterol Hepatol 2022; 37: 551-557.

[45] Xie H, Wei L, Liu M, Liang Y, Yuan G, Gao S, Wang Q, Lin X, Tang S and Gan J. Prognostic significance of preoperative prognostic immune and nutritional index in patients with stage I-III colorectal cancer. BMC Cancer 2022; 22: 1316.

[46] Xie H, Wei L, Liu M, Liang Y, Wang Q, Tang S and Gan J. The cancer inflammation prognostic index is a valuable biomarker for predicting the survival of patients with stage I-III colorectal cancer. Sci Rep 2023; 13: 18080.

Supplementary Table 1. Baseline characteristics in the study

| Characteristic	Overall (N = 416)
Age, years	62.0 (53.0-69.0)
Sex	
Male	235 (56.5%)
Female	181 (43.5%)
T stage	
T1	8 (1.9%)
T2	58 (13.9%)
T3	294 (70.7%)
T4	56 (13.5%)
N stage	
N1a	197 (47.4%)
N1b	219 (52.6%)
TNM stage	
IIIA	66 (15.9%)
IIIB	346 (83.1%)
IIIC	4 (1.0%)
Size, cm	
≤ 5	287 (69.0%)
> 5	129 (31.0%)
No. of lymph nodes retrieved	16.0 (13.0-19.0)
Albumin, g/L	43.7 (40.9-45.8)
CEA, ng/mL	2.5 (1.6-4.3)
CA19-9, U/mL	11.6 (7.5-22.5)
CA125, U/mL	11.1 (8.3-15.2)

Data are median (IQR) or n (%). TIPLN: tumor invasion proportion of lymph nodes; CEA: carcinoembryonic antigen; CA19-9: carbohydrate antigen 19-9; CA125: carbohydrate antigen 125.

Supplementary Table 2. Univariable and multivariable Cox analysis of continuous TIPLN

Variable	Univariate Cox analysis		Multivariate Cox analysis[a]	
	HR (95% CI)	*P* value	HR (95% CI)	*P* value
Age (> 60 vs ≤ 60 years)	1.10 (0.75-1.61)	0.625	-	-
Sex (female vs male)	1.13 (0.77-1.65)	0.529	-	-
T stage (T3-4 vs T1-2)	2.09 (1.12-3.89)	0.021	1.92 (1.01-3.63)	0.045
N stage (N1b vs N1a)	1.19 (0.81-1.75)	0.364	-	-
Size (> 5 vs ≤ 5 cm)	1.28 (0.87-1.91)	0.214	-	-
TIPLN (%)	1.02 (1.01-1.03)	< 0.001	1.02 (1.01-1.02)	< 0.001
Albumin (< 40 vs ≥ 40 g/L)	2.26 (1.55-3.30)	< 0.001	1.87 (1.27-2.74)	0.001
CEA (> 5 vs ≤ 5 ng/mL)	3.39 (2.27-5.07)	< 0.001	2.37 (1.48-3.80)	< 0.001
CA19-9 (> 37 vs ≤ 37 U/mL)	3.10 (2.05-4.69)	< 0.001	1.42 (0.86-2.33)	0.172
CA125 (> 35 vs ≤ 35 U/mL)	2.42 (1.35-4.32)	0.003	1.20 (0.65-2.21)	0.553

[a]Only significant prognostic factors in the univariate analysis were included in the multivariate model. TIPLN: tumor invasion proportion of lymph nodes; HR: hazard ratio; CI: confidence interval; CEA: carcinoembryonic antigen; CA19-9: carbohydrate antigen 19-9; CA125: carbohydrate antigen 125.
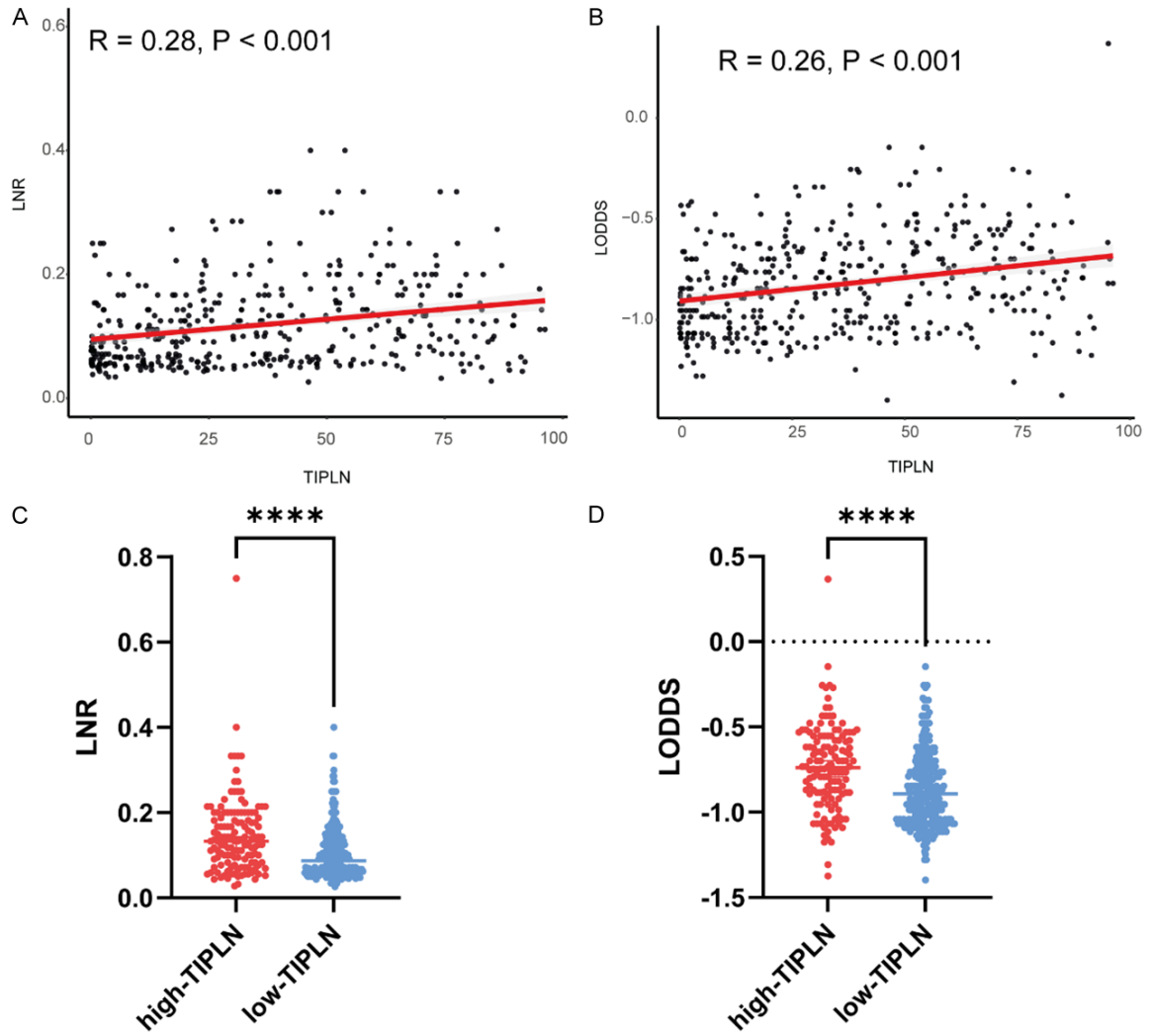
Supplementary Figure 1. A. The relationship between TIPLN and LNR; B. The relationship between TIPLN and LODDS; C, D. The distribution of LNR and LODDS in the high and low TIPLN groups. TIPLN: tumor invasion proportion of lymph nodes; LNR: lymph node ratio; LODDS: log odds of positive lymph nodes.