Preprints.org

Article

Assessment of the Tumor-Stroma Ratio and Tumor Infiltrating Lymphocytes in Colorectal Cancer: Inter-Observer Agreement Evaluation

Azar Kazemi , Masoumeh Gharib , Nema Mohammadian Roshan , Shirin Traz Jamshid , Fabian Stögbauer , Saeid Eslami * , Peter Schüffler

Article

Assessment of the Tumor-Stroma Ratio and Tumor Infiltrating Lymphocytes in Colorectal Cancer: Inter-Observer Agreement Evaluation

Azar Kazemi[1,2], Masoumeh Gharib[3], Nema Mohammadian Roshan[3], Shirin Taraz Jamshid[3], Fabian Stögbauer[1], Saeid Eslami[2,4,5,*], Peter J. Schüffler[1]

[1]Institute of General and Surgical Pathology, Technical University of Munich, 81675 Munich, Germany; ORCID: 0000-0003-3197-0491. Email: azarkazemi.mi@gmail.com (A.K.); ORCID: 0000-0002-8897-5165.Email: fabian.stoegbauer@tum.de (F.S.); ORCID: 0000-0002-1353-8921. Email: peter.schueffler@tum.de. (P.J.S.)
[2]Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, 9177948564 Mashhad, Iran. ORCID: 0000-0003-3197-0491.; ORCID: 0000-0003-3755-1212. Email: EslamiS@mums.ac.ir. Tel: +98 (0)51 3800-2429 (S.E.)

[3]Department of Pathology, Faculty of Medicine, Mashhad University of Medical Sciences, 9137913316 Mashhad, Iran. ORCID: 0000-0003-2501-2684. Email: Gharibm@mums.ac.ir (M.G.)**;** ORCID: 0000-0002-7862-7140. Email: Roshann@mums.ac.ir (N.M.R.)**;** ORCID: 0000-0002-5954-493X. Email: TJamshidiSH@mums.ac.ir (S.T.J.)

[4]Pharmaceutical Sciences Research Center, Institute of Pharmaceutical Technology, Mashhad University of Medical Sciences, 9177948954 Mashhad, Iran.

[5]Department of Medical Informatics, University of Amsterdam, 1105 AZ Amsterdam, the Netherlands. ORCID: 0000-0003-3755-1212. Email: EslamiS@mums.ac.ir. Tel: +98 (0)51 3800-2429.

[*]**Corresponding author:**
Saeid Eslami is responsible for communicating with the journal and responding to questions.

Simple Summary: We assessed the inter-observer agreement of stromal tumor-infiltrating lymphocytes (TILs), an immune biomarker, in intratumoral and tumor-front and tumor-stroma ratio (TSR) on 86 colorectal cancer (CRC) hematoxylin and eosin (HE) slides. Although an international TILs working group (ITILsWG) has proposed a methodology for assessing TILs in breast cancer, using the methodology for scoring the TILs in CRC has received less attention, and also introducing safe and feasible cut-off values for reporting the biomarkers in clinical setting still is a controversial subject. Results showed that poor to moderate agreement for TILs estimating and moderate agreement in TSR approximating among pathologists obtained. However, the two-tier cut-off classification was more agreed upon than the three-tier classification.

Abstract:

1

Background: To implement the new marker in clinical practice, reliability assessment, validation, and standardization of utilization must be applied. This study evaluated the reliability of TILs and TSR assessment through conventional microscopy by comparing observers' estimations. Methods: Intratumoral and tumor-front stromal TILs, and TSR were assessed by three pathologists using 86 HE slides CRC. TSR and TILs were categorized using one and four different proposed cut-off systems, respectively and agreement was assessed using the intraclass coefficient (ICC) and Cohen's kappa statistics. Pairwise evaluation of agreement was performed using the Fliess Kappa statistic and the concordance rate and it was visualized by Bland-Altman plots. Results: For the evaluation of intratumoral stromal TILs, ICC of 0.505 (95% CI:0.35-0.64) was obtained, Kappa values were in the range of 0.21 to 0.38, and concordance rates in the range of 0.61 to 0.72. For the evaluation of tumor-front TILs, ICC was 0.52 (95% CI:0.32 - 0.67), the overall kappa value ranged from 0.24 to 0.30, and the concordance rate ranged from 0.66 to 0.72. For estimating the TSR, the ICC was 0.48 (95% CI:0.35 - 0.60), the kappa value was 0.49 and the concordance rate was 0.76. Conclusion: The agreement between pathologists in estimating these markers corresponds to poor to the moderate agreement, however, implementing immune scores in daily practice requires more concentration in inter-observer agreements.

Keywords: tumor-infiltrating lymphocytes; tumor-stroma ratio; colorectal cancer; inter-observer

Introduction

Diagnosis and treatment decisions are vital, and prognosis prediction is crucial for enhancing clinical outcomes in patients with CRC [1]. Although the incidence of CRC is stable or decreasing in some countries, Asian and European countries show an increasing trend in CRC incidence [2]. Also, incidence and mortality arising from CRC are increasing among young people around the world [2]. The survival of CRC patients is improved in countries where screening programs are implemented but differs in countries with low-standing screening programs [3].

Despite these improvements, the burden of CRC remains one of the deadliest cancers for patients [4]. Therefore, questions about early diagnosis and more accurate prognosis have forced the study of new prognostic biomarkers. Tumor microenvironment assessment and immune cell detection have grown following clinical success trends in therapeutic strategies for some solid malignancies. Tumor-infiltrating

2

lymphocytes (TILs) and tumor-stroma ratio (TSR) are two new biomarkers that convey crucial information for the diagnosis, prognosis, and treatment response prediction of various solid cancers [5-9].

TILs that show a host response to the tumor are stronger predictors of patient survival than TNM classification [10]. Therefore, it can be used as a prognostic factor across solid tumor types, including CRC [11-13]. The ITILsWG standardized the evaluation of TILs on HE-stained sections in breast cancer [14] and recommended using the same methodology for other solid malignancies to incorporate TILs quantification into clinical practice and research.

Currently, there is no validated cutoff system to report TILs in clinical practice. Therefore, recent studies should contribute to determining the safe threshold to provide more evidence for introducing an ideal cutoff system for daily practice [15]. However, according to ITILsWG recommendations [16], this new biomarker has been reported as a continuous variable. To implement a clinically relevant cutoff system, researchers who have applied the ITILsWG standardized scoring methodology have investigated different cutoffs to classify TILs into low to high-grade categories in colorectal [17, 18] and breast cancers [14, 19] and have evaluated the prognostic value of TILs based on these cutoff systems. Iseki et al. [17] evaluated the ITILsWG method to standardize the TILs estimation and prognostic utility assessment of TILs in CRC. This study reported that the methodology proposed by the ITILsWG can be useful in assessing TILs in CRC and can be used as a prognostic factor. A recent study by Fuchs et al. [18] validated the efficacy of ITILsWG methodology in a large cohort of CRC patients. This study showed that compared with intraepithelial TILs, stromal TILs assessed using the ITILsWG approach have a better prognostic value in CRC. As these cells are dense in some parts of the tissue, visual assessment by pathologists is prone to error [20]. The study assessed inter-observer agreement in face-to-face and non-intensive training models and reported good and moderate inter-observer agreements, respectively [18].

TSR, as a percentage of the stroma area compared to the tumor cell area, is a histological feature assessed based on the morphological evaluation of HE sections. TSR, which indicates the quantity of the stromal

component surrounding cancer cells in a tissue section, has been identified as a determinant of progression [21] and a predictor of prognosis in CRC [22, 23]. Based on a meta-analysis [22], in most studies that validated the prognostic value of TSR in patients with CRC, a higher percentage of stroma (>50%) was associated with the worst prognosis.
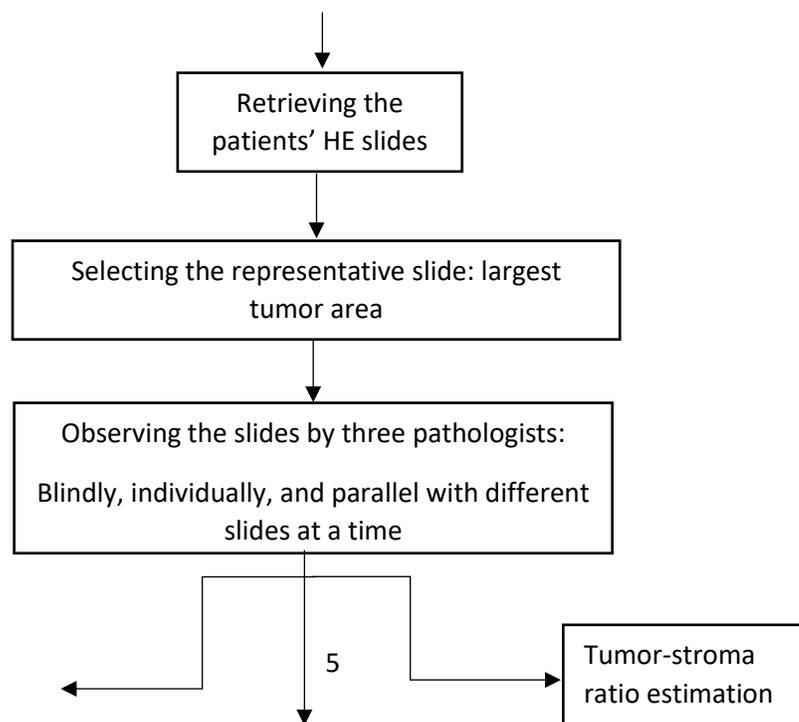
These studies suggest that estimating stromal components and investigating new biomarkers can be helpful in clinical practice by incorporating these bioscores into diagnosis and treatment strategies. However, concerns have been expressed regarding the reliability of the approximation of stromal component entities. Although few studies have shown the efficacy of ITILsWG proposed guidelines for predicting outcomes in CRC, the concordance rates in quantifying TILs showed variable agreement [18]. Likewise, semi-quantitative reporting of TSR is under study to investigate the variation in concordance rates. Recently, Souza et al. [23] showed that TSR estimation with good concordance was feasible. Our study aimed to determine the inter-observer agreement of TSR assessment and TIL scoring in CRC HE-stained sections using the proposed methods.

Material and Methods

Patients

This retrospective study was conducted on patients who underwent surgical resection for grades II and III of colorectal adenocarcinoma between 2010 and 2017 in the Omid oncology hospital affiliated with the Mashhad University of Medical Sciences, Mashhad, Iran. All patients of the study signed an informed consent for using their data and tissues in the research. All the procedures in the study were performed according with the institutional and national ethical standards and rules of the Declaration of Helsinki of 1975 and later comments. This study was approved by the Mashhad University Medical Sciences Research Ethics Committee (approval ID: IR.MUMS.MEDICAL.REC.1399.557).

Histological slides were collected from the archives of the pathology department. To use of TILs and TSR in clinical setting, standardization of detection and quantification of biomarkers on HE slides, as a standard tissue section staining, in research setting can facilitate diagnosis procedures especially for developing

Selecting the patients who underwent surgical resection from 2010 to 2017

countries. Therefore, 101 cases with histological features of colorectal adenocarcinoma and available HE-stained sections were collected. HE slides were reviewed via light microscopy by a pathologist to select one representative HE slide (4 μm thick) for each patient. Patients were excluded if they were treated before surgery, if no tumor mass or invasion was found in the tissue section, or if the tumor was entirely necrotic. Finally, the study cohort included eighty-six cases in whose HE slides were assessable. Three pathologists (M.G, N.M.R, S.T.J) with a range of 8 to 17 years of clinical and surgical pathology experience participated in the study. Two pathologists (observer 1: OBS1 and observer 2: OBS2) were experienced in TIL scoring and the other pathologist (observer 3: OBS3) had less experience in TILs counting. No grouping training session was provided, but the ITILsWG guidelines and materials were provided to participating pathologists [15] and any questions regarding observation and scoring TILs and TSR clarified by OBS1. Scheme I shows the study method.

> **Retrieving the patients' HE slides**

> **Selecting the representative slide: largest tumor area**

> **Observing the slides by three pathologists:**
> **Blindly, individually, and parallel with different slides at a time**

5

> **Tumor-stroma ratio estimation**

Intratumoral stromal
TILs estimation

Tumor-front
stromal TILs
estimation

Inter-observer agreement assessment according to
four cutoff system:

- System 1 (TILs density <10%, 10-50%, >50%)
- System 2 (TILs density <10%, 11-39%, ≥40%)
- System 3 (TILs density ≤20%, 21-9%, ≥50%)
- System 4 (TILs density <42%, ≥42%)

Inter-observer agreement
assessment according to
single cut-off system:
Stroma ≤50%, >50%

Comparing and interpreting the
ICCs, Kappa values, concordance
rates, discordance, consensus
rates and pairwise evaluation

Scheme I. Study methodology

Quantification of stromal TILs

According to the ITILsWG's proposed methodology of 2014 [14], the density of TILs on HE slides was

evaluated. As the study recommended, visual measuring of stromal TILs which are a superior parameter

using standard HE-stained sections considered as a primary aim of TILs quantification. In summary,

mononuclear cells, including lymphocytes and plasma cells, within the stromal area should be evaluated.

Other cells such as granulocytes and polymorphonuclear leukocytes should be excluded. TILs outside of

invasive tumor boundaries should also be excluded. The average stromal TILs density within the borders

of the invasive tumor over the entire whole HE section should be scored. Crush artifacts, necrosis, or fibrosis

should be excluded. Figure 1.A-C shows three levels of TILs density according to the methodology.

A) Low, 100X   B) Intermediate, 100X   C) High, 100X

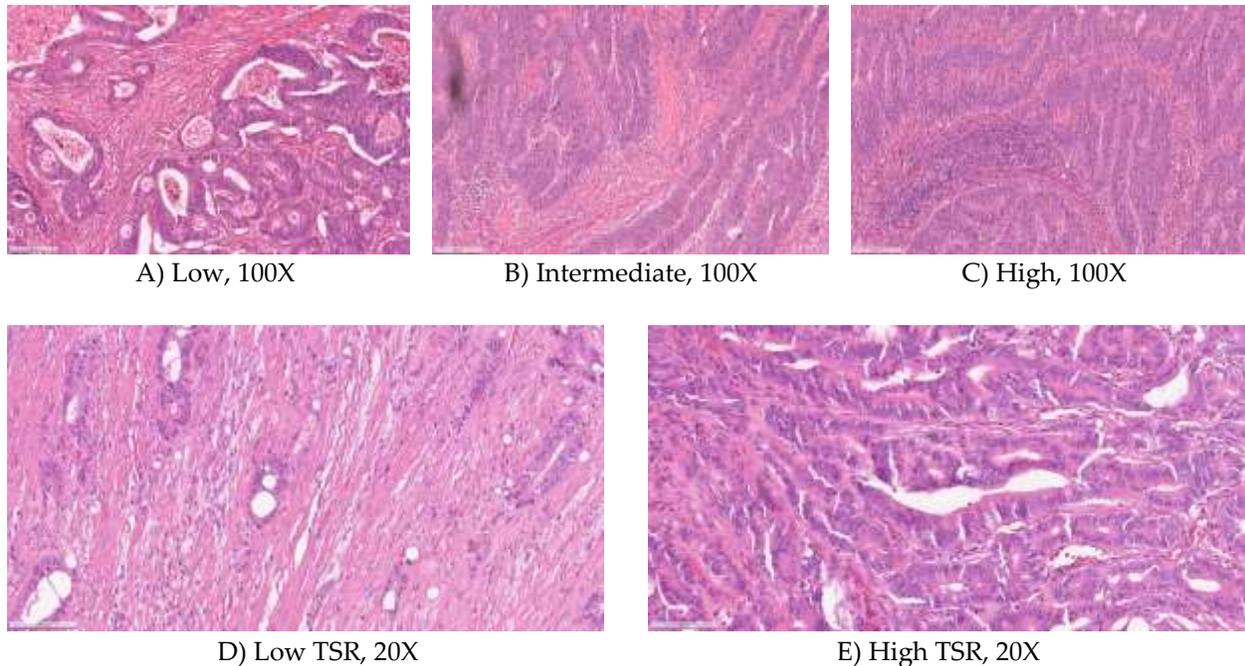D) Low TSR, 20X   E) High TSR, 20X

Figure 1. Examples of grades of TILs and TSR. A-C) Examples of different grades (low/intermediate/high) of intra-tumoral stromal TILs density in colorectal cancer HE stained slides scored according to the ITILsWG methodology. D-E) Example of cases with low and high stroma in CRC HE-stained slides.

In a blinded manner, three observers independently estimated the density of TILs in the invasive tumor-front and intratumoral TILs. To avoid any bias, all pathologists were blinded to other clinical and histopathological details and then stored the scores in a columnar form created for reporting. The scores were reported as continuous values and rounded to the nearest 0.05. All reports were collected by a data analyzer who was blinded to other patients' data during the observation phase. All results were kept confidential, with no feedback to the participants. Using the ITILsWG method [14] and the cutoffs suggested in previous studies [17-19], the continuous TILs scores were categorized into two or three grades: low to high. The suggested grading systems for the quantification of intratumoral and tumor-front stromal TILs are shown in Tables 1 and 2.

Table 1. Inter-observer agreement among pathologists for estimating intratumoral TILs using four cutoff systems

Intraclass coefficient (ICC)	0.51 (CI*: 0.35 - 0.64)
Fleiss' kappa values (P-value)	
System 1 (<10%, 10-50%, >50%)	0.32 (<0.001)
System 2 (<10%, 11-39%, ≥40%)	0.21 (<0.001)
System 3 (≤20%, 21-9%, ≥50%)	0.27 (<0.001)

System 4 (<42%, ≥42%)	0.38 (<0.001)
Concordance rates	
System 1	0.70 (Range: 0.63-0.77)
System 2	0.61 (Range: 0.56-0.71)
System 3	0.62 (Range: 0.60-0.69)
System 4	0.72 (Range: 0.69-0.78)

* CI: 95% confidence interval.

Table 2. Inter-observer agreement for estimating tumor-front TILs using different cutoff systems

Intraclass coefficient (ICC)	0.52 (CI*: 0.32 - 0.67)
Fleiss' kappa values (P-value)	
System 1 (<10%, 10-50%, >50%)	0.26 (<0.001)
System 2 (<10%, 11-39%, ≥40%)	0.24 (<0.001)
System 3 (≤20%, 21-9%, ≥50%)	0.29 (<0.001)
System 4 (<42%, ≥42%)	0.30 (<0.001)
Concordance rates	
System 1	0.71 (range: 0.53-0.77)
System 2	0.68 (range: 0.58-0.7)
System 3	0.66 (range: 0.58-0.67)
System 4	0.72 (range: 0.63-0.76)

* CI: 95% confidence interval.

Tumor-stroma ratio estimation

TSR was quantified visually by three pathologists on the tissue slides used in routine diagnostic pathology using conventional light microscopy according to the approximation of TSR in the Sullivan et al. study [24]. The estimations were recorded as a tumor-to-stroma ratio in a blinded manner, and inter-observer agreement in the estimation of the TSR was assessed by three pathologists. In areas of necrosis, artifacts, and crush artifacts, stromal cells were not estimated. According to previous studies, which were reviewed in a meta-analysis by Zhu et al. [22], a cutoff of 50% was used for TSR scores. Cases with ≤ 50% stroma were considered stroma-poor, and cases with >50% stroma were considered stroma-rich. Two examples of stroma-poor and stroma-rich cases are shown in Figure 1.D-E.

Statistical analysis

Statistical analyses were performed using R version 4.2.1. Missing values were imputed using the mean of sample scores. The intra-class correlation coefficient(ICC) was calculated using a mixed model to assess the inter-observer agreement for the continuous scores of TILs and TSR. After categorizing the scores, Cohen's Kappa was used for pairwise inter-observer reliability between two observers, and Fleiss kappa as an adaptation of Cohen's Kappa [25] was used to measure inter-rater agreement among all pathologists.

Concordance rate analysis was performed as the overall agreement between each observer and the median value for each sample and the Bland-Altman plots were drowned to compare the intratumoral TILs scores discordance. To assess the pairwise correlation between pathologists, the Spearman correlation coefficient was applied to study the raw biomarker scores because the scores were not normally distributed. The McNemar test was performed to determine whether there was a difference in the frequency of categories scored by the two pathologists. Furthermore, we assumed that the sample was difficult to estimate if all pathologists disagreed (extremely different scoring for a sample) and showed the differences through a heatmap graph. Statistical significance was set at P ≤0.05.

Results

The study population included eighty-six patients aged between 21 and 87 years, with a mean age of 56.74 years. Of the patients, 54.65% were male.

Intratumoral stromal TILs

The ICC for estimating intratumoral TILs was 0.505 (95% CI:0.35-0.64) (Table 1). The overall kappa values for intratumoral stromal TIL evaluation using different cutoff systems were in the range of 0.21 to 0.38 (Table 1). The concordance rates show the overall agreement among each pathologist's estimation for each case compared with the same case intratumoral TILs median value. The overall concordance for quantifying the density of TILs in the intratumoral ranged from 0.61 to 0.72 (Table 1).

9

To visualize the discordance among pathologists, based on the TILs density scores estimated by each observer and OBS_Median (median calculated for each sample as a gold standard observer), Bland-Altman plots were drawn (Figure 2.A-C). While OBS2 and OBS3 tended to estimate intratumoral TILs densities lower than OBS-median, OBS1 gave higher TILs scores than OBS_Median (Figure 2.A-C and Figure 3.A).
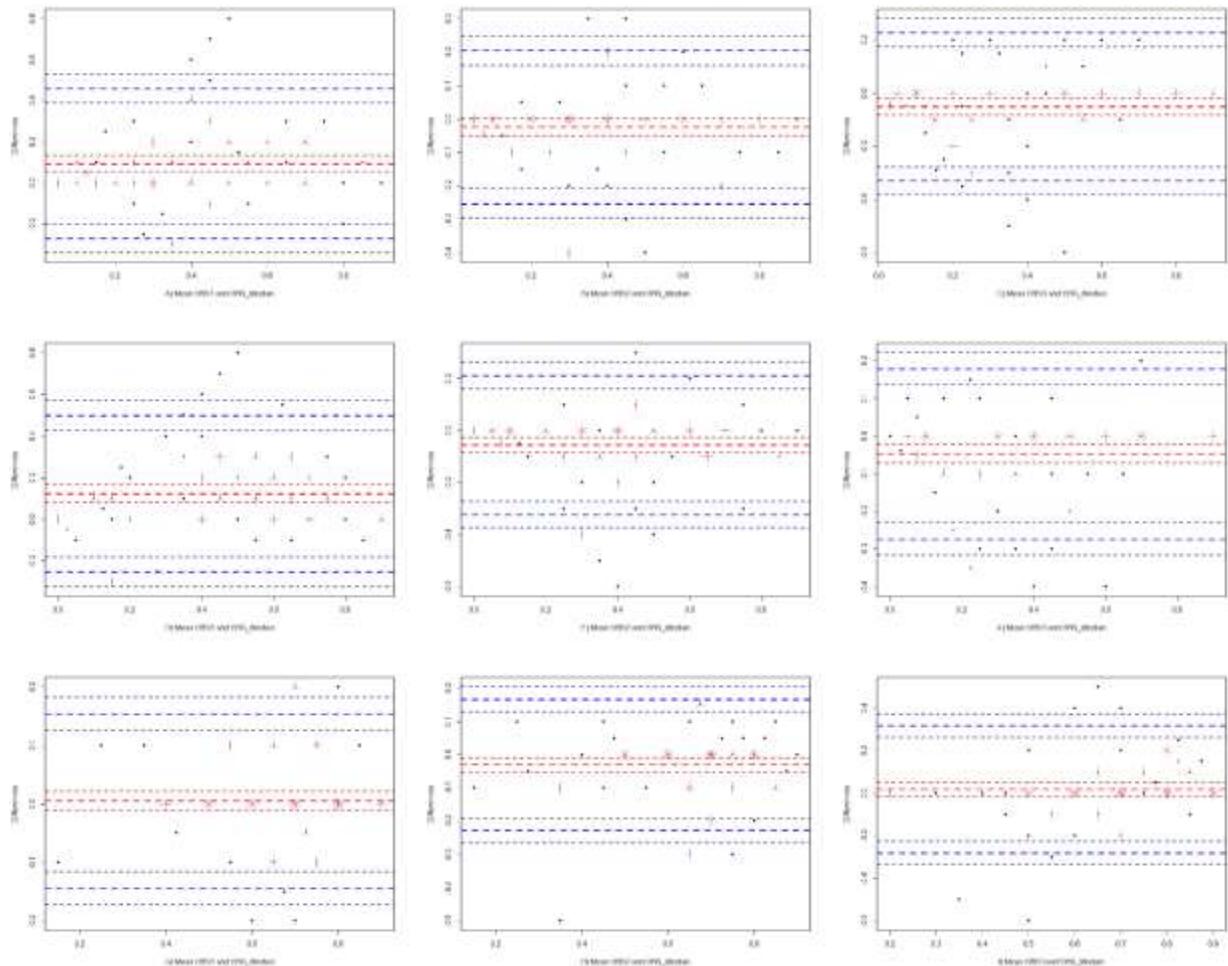

Figure 2. Bland-Altman plots were drowned to compare the intratumoral TILs scores (A-C), tumor-front TILs (D-F), and TSR scores (G-I) of three observers (OBS1, OBS2, and OBS3) to the median of their scores (OBS-Median). The y-axis represents the difference between the observer and OBS-Median scores, while the x-axis shows their mean values. The red dashed line denotes the mean difference, and the blue dashed lines show the upper and lower limits of agreement. Additionally, the plot includes dashed dot red lines and dashed dot dark blue lines, which represent the 95% confidence interval limits for the mean difference and agreement limits.
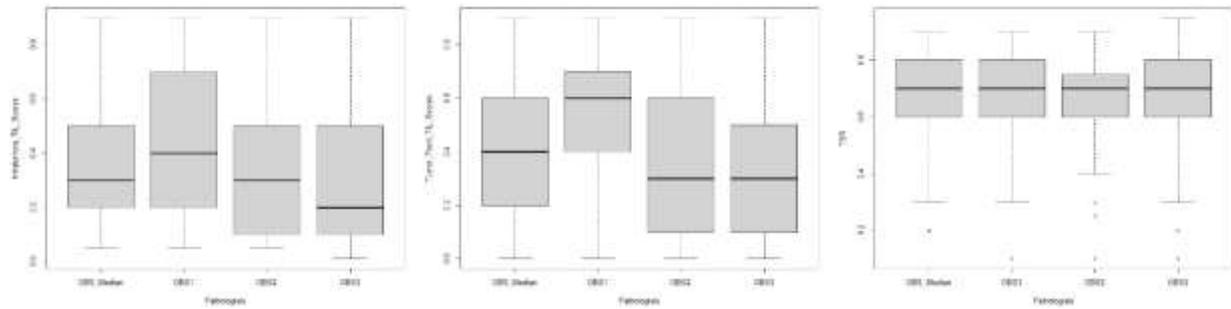
A) Intratumoral TILs scores distribution per pathologist	B) Tumor front TILs scores distribution per pathologist	C) TSR distribution per pathologists

Figure 3. The distribution of biomarkers scores given by each pathologist. The box plot of OBS_Median shows the distribution of the median of biomarkers scored by three observers (OBS1, OBS2, and OBS3).

The consensus for different cutoff categories of TILs is shown in Table S1. A positive consensus meant that all pathologists agreed on the TIL score. Likewise, a partially positive consensus meant that two of the three pathologists agreed, and a negative consensus can be interpreted as a complete disagreement between all pathologists in evaluating the same case. The frequencies of cases in each category of intratumoral stromal TILs recorded by the pathologists are shown in Table S1. Results indicate that compared to three-tiered systems, using two-tiered systems causes more positive consensus among pathologists.

Correlations between pathologists are in a range of 0.48 to 0.64 (Table 2). Pairwise agreement evaluation for estimation of stromal TILs using different cutoff systems showed that there was a significant difference between OBS1 and the other two pathologists. The pairwise consensus shows the percentage of cases in which the two pathologists agree (Table S2). A heatmap graph was drawn to represent slides that are difficult to score and cause under-or overestimation of TILs by pathologists. Figure 4.A) shows that for the same cases, some observers reported scores higher than those of other observers.

A)
 Cases

    1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
    39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73
    74 75 76 77 78 79 80 81 82 83 84 85 86

Mean

B)
Cases

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86

C)
Cases

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86
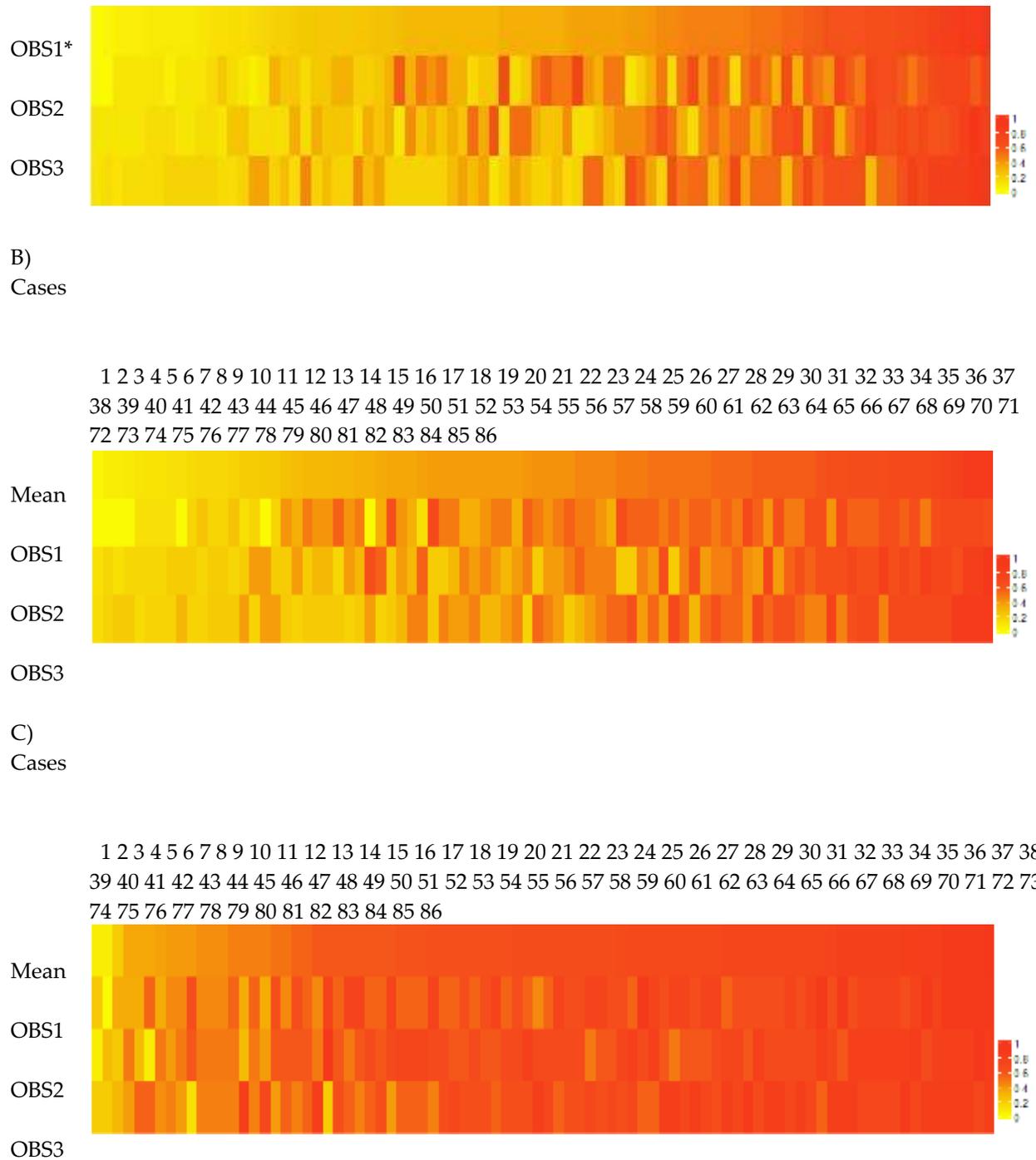

Figure 4. A) The first row in the heat map shows the mean value of intratumoral TILs records for each case scored by all pathologists according to ascending from left to right; the other rows represent three observers' intratumoral TILs records for the same case. Color-coding indicates the density of TIL quantified by each pathologist. The comparison shows there are some outlier values in low-level (yellow) to high-level (red) TIL densities. B) The first row in this heat map shows the mean value of scored TILs in the tumor-front for each case according to ascending from left to right; the other rows represent three observers' TILs estimated scores for cases. C) The first row in the heat map shows the tumor-stroma ratio (TSR) mean values from left to right according to ascending order for each case, the other rows represent three observers' TSR records.

TILs in the tumor-front

The intraclass coefficient, kappa statistic, and concordance rate among observers for estimating the tumor-front TILs were calculated according to different cutoff systems, and the results are shown in Table 3.

Table 3. Inter-observer agreement for estimating tumor-front TILs using different cutoff systems

Intraclass coefficient (ICC)	0.52 (CI*: 0.32 - 0.67)
Fleiss' kappa values (P-value)	
System 1 (<10%, 10-50%, >50%)	0.26 (<0.001)
System 2 (<10%, 11-39%, ≥40%)	0.24 (<0.001)
System 3 (≤20%, 21-9%, ≥50%)	0.29 (<0.001)
System 4 (<42%, ≥42%)	0.30 (<0.001)
Concordance rates	
System 1	0.71 (range: 0.53-0.77)
System 2	0.68 (range: 0.58-0.7)
System 3	0.66 (range: 0.58-0.67)
System 4	0.72 (range: 0.63-0.76)

* CI: 95% confidence interval.

The ICC for evaluation of the stromal TILs in the tumor-front was 0.52 (95% CI:0.32 - 0.67). The overall kappa value for the different categories of TILs density in the tumor-front ranged from 0.24 to 0.30. The overall agreement between each pathologist's estimation and the TILs median value for the same case in different categories ranged from 0.66 to 0.72.To determine the number of slides that are more difficult to evaluate the stromal TILs in the tumor-front, the consensus for different cutoff categories of stromal TILs in the tumor-front shows that pathologists had consensus on at least 37% of cases. According to the recorded TILs in the tumor-front, at least 58% of cases were estimated by OBS1 categorized in high density of TILs (Table S3).

The correlation among pathologists showed that the correlation between TILs scored by OBS2 and OBS3 was higher than the correlation with TILs scored by OBS1 (Table 2). There were no significant differences between OBS2 and OBS3 in all grading systems for scoring TILs in the tumor-front (Table S4). Kappa values were obtained from 0.16 to 0.47, also in comparison with three-tiered categories, according to cutoff system

4, which is a two- tiered categorization, between OBS1 and OBS3, who are experienced in scoring tumor-front TILs, the higher agreement is observed (consensus:0.66 (95% CI:0.55 - 0.76); kappa value:0.40) (Table S4). The heatmap indicates the challenging slides for evaluation that led pathologists to under-or overestimate stromal TILs in the tumor-front (Figure 4.B). It seems OBS1 estimated the tumor-front TILs with a higher threshold, and patterns of tumor-front TILs estimated by OBS2 and OBS3 showed that they gave lower scores (Figure 2.D-F, Figure 3.B). In total, OBS1 estimated the TILs with a higher threshold.

Tumor-stroma ratio

According to the Spearman coefficient analysis results, for raw continuous TSR scores (Table 2), the TSR scored by OBS1 correlated with the TSR scores estimated by OBS2. Stromal estimations were divided into two categories: stroma-rich and stroma-poor. Evaluation of TSR led to an ICC of 0.48 (95% CI:0.35 - 0.60), and an overall kappa value of 0.49. However, the concordance rate between pathologists was 0.76, which was good (Table 4). Furthermore, a positive consensus existed over the 75 slides for estimating TSR (Table S5). In addition, the table shows the frequency of cases estimated by the pathologists in each category.

Table 4. Inter-observer agreement values for estimating tumor-stroma ratio

Intraclass coefficient (ICC)	0.48 (CI*: 0.35 - 0.60)
Fleiss' kappa values (P-value)	
≤50% , >50%	0.49 (<0.001)
Concordance rates	
≤50% , >50%	0.76 (range: 0.72-0.78)

* CI: 95% confidence interval.

The agreement between pathologists for estimating the TSR ranged from 0.26 to 0.72 and there was no significant difference between OBS1 and OBS2 (consensus: 0.95, kappa value: 0.72) (Table S6). The heatmap for TSR estimation illustrates how some observers assessed TSR differently from other observers. It was assumed that samples with disparate scores could not be easily estimated (Figure 3.C). The distribution of the TSR scores by pathologists for the 86 slides is shown in Figure 3.C. OBS1 and OBS3 assessed the TSR as close to OBS_Median, but OBS2 estimated the TSR to be relatively lower than OBS_Median. Additionally,

Figure 2.G-I, which visualized discordance among pathologists for estimating TSR, showed that the variability in stroma-rich cases was higher than that in stroma-poor cases.

Discussion

The current study evaluated inter-observer agreement among pathologists estimating the density of stromal TILs in intratumoral and tumor-front compartments and the tumor-stroma ratio in CRC using the ITILsWG recommended methodology and proposed cutoff values in previous studies. In the current study, HE-stained slides as the most common histological staining for cancer diagnostic purposes around the world reviewed to measure stromal TILs and TSR. For scoring intratumoral and tumor-front TILs densities, TILs scored by OBS2 and OBS3 were more correlated, and the results showed for estimation TSR, the estimated ratio of tumor-stroma by OBS1 and OBS2 were more correlated.

Intratumoral stromal TILs

According to the criteria of Koo and Li [26], the ICC for assessing intratumoral stromal TILs can be interpreted as poor to moderate agreement. According to these criteria, the 95% confidence interval should be considered as a basis to evaluate the reliability level which is categorized into values <0.5 as poor, 0.5 to 0.75 as moderate, 0.75 to 0.9 as good, and values greater than 0.9 as excellent reliability [26]. According to Landis and Koch's interpretation criteria [27], the overall kappa values were interpreted as a fair agreement. According to these criteria, the value of zero is interpreted as poor agreement, 0 to 0.2 as slight, 0.2 to 0.4 fair, 0.4 to 0.6 as moderate, 0.6 to 0.8 substantial, and 0.8 to 1 as perfect agreement. The results indicate a lower agreement was obtained if we categorized the intratumoral TILs densities into three-tiered categories and would improve to a higher inter-observer agreement (kappa value:0.38, concordance rate:0.72, and consensus rate:55.81%) when categorizing TILs into two-tiered groups. Tramm et al. [19] also showed that when categorizing TILs density into two levels, the inter-observer agreement would increase.

The pairwise evaluation indicated fair to moderate agreement among pathologists estimating intratumoral TILs using the ITILsWG proposed methodology cutoffs (system 1) [14] and Iseki et al. [17] suggested a cutoff (system 4). There is a moderate agreement among OBS2 and OBS3 in 72.1% of cases (kappa value:0.44) using the ITILsWG cutoff system (system 1) and agreed on TILs scores of 74.4% of cases (kappa value:0.43) categorized according to the suggested cutoff system 4. There is no evidence of significant differences between OBS2 and OBS3 in estimating intratumoral TILs scores. The Spearman correlation coefficient for scoring intratumoral TILs was moderately positive among the pathologists. However, the TILs scored by OBS2 and OBS3 were more correlated. All evidence indicates that scores recorded by OBS1, who have experience in training ITILsWG methodology for estimating the density of TILs, are significantly different from the two other pathologists. Overall, the variability in cases with an intermediate density is higher than that in cases with a low or high density of TILs.

This study shows that in cases with a moderate mean density of intratumoral TILs, pathologists deal with challenges in scoring these cases, and TIL densities are variably estimated. However, tissues with low or high mean densities of TILs were estimated in the same range. The current study emphasizes the results of a previous study [19] that the source of inter-observer disagreement could be the variable threshold between pathologists. Obtaining aid from software in the agreement study process is a probable solution for dealing with disagreement factors. Denkert et al. [28] have shown that although using software in the assessment of TILs makes the assessment approach complex, optimizing the reproducibility is likely.

In some studies [5, 14, 15, 17, 19, 23, 29], the suggested cutoffs for grading stromal TILs have varied, and in most of these studies, the cutoff was estimated for grading TILs in breast cancer [5, 15, 19, 29]. By comparing different cutoff systems categorizing the TILs, although none of the cutoff systems showed a moderate agreement among all pathologists, the ITILsWG [14] suggested cutoff values (<10%, 10-50%, >50%) and Iseki et al. [17] suggested cutoff values (<42%, ≥42%) are more reliable. The concordance for each pathologist also improved when we used the two-tiered system (system 4), which was suggested based on a CRC study

[17], indicating that categorizing tumors with two levels of TILs as low vs. high can facilitate the assessment of TILs in practice. However, according to Denkert et al. [28] first ring of the inter-observer agreement study, in which 32 pathologists participated in scoring the TILs in 60 breast cancer cases, showed agreement between pathologists in comparison with dichotomous grading systems. According to their study, grading TILs into two-tiered cutoffs does not guarantee better agreement, and setting precise cutoffs is crucial. In our study, dichotomous cutoff groups were used according to the CRC study [17], which slightly improved the overall and pairwise inter-observer agreement.

Furthermore, this study is in accordance with previous studies [17, 18] that showed that although the methodology recommended by the ITILsWG proposed for scoring TILs in breast cancer, using the recommended methodology for detecting and assessing TILs can be a useful methodology in CRC. However, according to our results, the suggested cutoffs [14] as a three-tiered grading system (system 1), which has been used in most studies to evaluate inter-observer reproducibility of TIL scoring among pathologists [18, 19, 29] separating the tumors into three-tiered levels of TILs, can be safe for scoring TILs in breast cancer but needs to be assessed more in CRC. To the best of our knowledge, the only studies that have evaluated the methodology of ITILsWG in CRC are Iseki et.al [17] and Fuchs et.al [18] studies, that according to these studies, this methodology was a strong predictor of survival in patients with CRC.

TILs in the tumor-front

According to the results of this study, the correlation among pathologists for estimating the tumor-front TILs density is moderately positive, which in more detail, the correlations between OBS2 and OBS3 were higher than the correlation with OBS1.

According to Koo and Li [26] criteria, ICC for evaluating tumor-front TILs can be interpreted as poor to moderate agreement between pathologists. The overall kappa value for the different categories of TILs density in the tumor-front corresponds to a poor agreement among pathologists [27]. This study shows that,

in comparison with the three-tiered level of TILs density in the tumor-front, the concordance for each pathologist was improved when using a two-tiered level of TILs in the tumor-front (system 4), indicating that estimating tumor-front TILs in tumors with low vs. high levels of TILs density is more reliable in clinical practice. In comparison with three-tiered categories, according to cutoff system 4, which is a two-grade categorization, between OBS1 and OBS3 who are experienced in scoring TILs, a higher agreement is observed. The kappa statistic and concordance rate showed that the ITILsWG proposed methodology [14] is slightly safe for identifying TILs in the tumor-front. However, the positive consensus rate in system 1 was slightly lower than that in the other cutoff systems. System 4, which is a three-tiered grading system, is safer to use in daily practice, indicating an appropriate cutoff system for categorizing TILs density in CRC.

The percentage of cases in which two pathologists agreed to estimate the TILs density in the tumor-front. Kappa statistics for all cutoff systems indicate between OBS2 and OBS3 is a fair or moderate agreement. In general, there were no significant differences between OBS2 and OBS3 in all grading systems in scoring TILs in the tumor-front.

The frequencies of cases in each category of TILs in the tumor-front recorded by pathologists showed that OBS1 estimated the tumor-front TILs with a higher threshold. The heat map indicates the challenging slides for evaluation that led pathologists to underestimate or overestimate TILs at the tumor-front. It is more challenging for pathologists to estimate TIL density in cases with a moderate mean value of TILs in the tumor-front. Given that the standardized methodology of ITILsWG has been proposed for scoring intratumoral TILs in breast cancer, few studies show using it for scoring intratumoral TILs [18] and tumor-front TILs [17] in CRC, this study showed equal agreement can obtain by using the methodology for scoring intratumoral and tumor-front TILs in CRC. Furthermore, Iseki et al. showed that evaluating TILs by the observer was correlated with the software assessment of TILs [17]. Computational assessment of TILs in the tumor-front can validate the usefulness of the recommended methodology and suggest cutoff values. In addition, it seems OBS1 estimated the tumor-front TILs with a higher threshold, and patterns of tumor-

front TILs estimated by OBS2 and OBS3 showed that they gave lower scores. This evidence showed the source of variability in TIL scoring is the different thresholds among observers [19].

Tumor-stroma ratio

According to Spearman coefficient analysis, the correlation between OBS1 and OBS2 had higher values than the correlation with OBS3. The ICC value was indicative of poor reliability between pathologists in estimating TSR. The ICC value indicates that more than 50% of the variance in the results can be attributed to the variance caused by non-stromal attributes; the remaining variance is due to stromal proportion differences. The overall kappa value showed moderate agreement between all pathologists in estimating TSR in CRC tissue sections [30]. In comparison with the study by Souza et al. [23], the kappa value obtained in our study showed less agreement. Therefore, there is a varied conclusion for disagreements among pathologists. Therefore, these evidence shows for reducing the disagreement between pathologists for estimating TSR more research is needed.

However, a good concordance rate between the pathologists was found, and a positive consensus was obtained in 87.2% of the cases. These results show that even if the estimation of the TSR is visual and subjective, pathologists agree on approximating the percentage of the stromal area compared to the tumor area in CRC tissue. Pairwise agreement evaluation indicated that the agreement between pathologists for estimating the TSR ranged from fair to good; therefore, there was no difference between pathologists in approximating the TSR. The results vouch for pairwise agreement among pathologists in the scoring of TSR on CRC slides. Additionally, the variability in approximating the stroma-rich cases was higher than that in stroma-poor cases.

Limitations

This study had some limitations. First, this was a retrospective study with a small sample size. Given that Mashhad is a medical referral city in the northeast of the country, patients from different cities and places

have come to undergo surgery, in addition, health unintegrated systems led to a lack of access to survival status data for some patients. Therefore, prospective studies with larger sample size are required to confirm the results of the present study. Second, Lack of access to IHC-stained sections made it impossible to compare TILs score and Immunoscore on HE and IHC, respectively. Third, sample size was not sufficient to clarify whether using the recommended methodology is associated with better patient clinical outcomes. Further studies are needed to validate cut-offs of TILs scored in intratumoral and invasive front according to the ITILsWG methodology to evaluate its association with patient survival. Finally, the observers in the study had different levels of experience in clinical and surgical pathology, although observer 1 answered different questions regarding the aim of the study, TILs and TSR estimating, and recording the scores, the subjectivity of the measuring by the human remained. Therefore it seems that the computerized diagnostic assistant on the HE slides could aid pathologists to overcome the subjectivity of scoring in clinical settings.

Conclusion

In conclusion, the present study confirmed that the evaluation of TILs in CRC using HE slides according to the methodology recommended by the ITILsWG provides poor to the moderate agreement among study pathologists. Despite the discordance among observers for assessing TILs, the two-tiered cutoff system is safer to apply in routine practice because of less disagreement among pathologists. Furthermore, although the assessment of TSR using the proposed methodology is appropriate, a moderate inter-observer agreement was observed among pathologists with different experiences in pathology professions. In the future, we aim to automate biomarkers estimation using artificial intelligence that can aid pathologists in detecting biomarkers on HE-stained sections to smooth the disagreements between pathologists in the routine clinical setting. Applying multi-phasing training-observing to the study may be beneficial in improving the agreement between pathologists to create ground truth. However, larger studies to emphasize the agreement between pathologists are required before using biomarkers for CRC in practice settings.

Supplementary Materials

Table S1. Consensus rate among pathologists in assessing the density of intratumoral TILs, and the frequency of TIL density levels estimated by each pathologist categorized in different cut-off systems.

Table S2. Pairwise agreement assessment for intratumoral stromal TILs scoring

Table S3. Consensus rate between pathologists for estimating the tumor front TILs, and the frequency of TIL density levels scored by each pathologist in the different cut-off systems.

Table S4. Pairwise agreement evaluation for estimating stromal TILs in tumor front

Table S5. Consensus rate between pathologists for estimating the TSR and frequency of TSR levels estimated by each pathologist.

Table S6. Pairwise agreement evaluation between pathologists for assessing the TSR

Author contributions

A.K.: contributed to all research steps, including study design, data collection, statistical analysis, and writing of the manuscript. M.G.: participated in the study design, data collection, and observation of the slides, writing the manuscript, and supervising the pathological side of the study. N.M.R. and S.T.J.: participated in observing the slides. F.S.: contributed to preparing and writing the manuscript. S.E.: supervised the A.K in the study. P.J.S.: supervised A.K. in the study, participated in data interpretation. All authors have read and reviewed the manuscript and agreed to their individual contributions prior to submission and agreed to the published version of the manuscript.

Funding

Institutional Review Board Statement

The study was approved by the Medical Ethics Committee (approval ID:

IR.MUMS.MEDICAL.REC.1399.557)

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

All data are stored and available upon request to first author. Azarkazemi.mi@gmail.com

Conflicts of Interest

All authors declared that they have no conflict of interest.

References:

1.　　Bychkov, D.; Linder, N.; Turkki, R.; Nordling, S.; Kovanen, P. E.; Verrill, C.; Walliander, M.; Lundin, M.; Haglund, C.; Lundin, J., Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep* **2018,** *8* (1), 3395.
2.　　Wong, M. C. S.; Huang, J.; Lok, V.; Wang, J.; Fung, F.; Ding, H.; Zheng, Z. J., Differences in Incidence and Mortality Trends of Colorectal Cancer Worldwide Based on Sex, Age, and Anatomic Location. *Clin Gastroenterol Hepatol* **2021,** *19* (5), 955-966 e61.
3.　　Cardoso, R.; Guo, F.; Heisser, T.; Hackl, M.; Ihle, P.; De Schutter, H.; Van Damme, N.; Valerianova, Z.; Atanasov, T.; Majek, O.; Muzik, J.; Nilbert, M. C.; Tybjerg, A. J.; Innos, K.; Magi, M.; Malila, N.; Bouvier, A. M.; Bouvier, V.; Launoy, G.; Woronoff, A. S.; Cariou, M.; Robaszkiewicz, M.; Delafosse, P.; Poncet, F.; Katalinic, A.; Walsh, P. M.; Senore, C.; Rosso, S.; Vincerzevskiene, I.; Lemmens, V.; Elferink, M. A. G.; Johannesen, T. B.; Korner, H.; Pfeffer, F.; Bento, M. J.; Rodrigues, J.; Alves da Costa, F.; Miranda, A.; Zadnik, V.; Zagar, T.; Lopez de Munain Marques, A.; Marcos-Gragera, R.; Puigdemont, M.; Galceran, J.; Carulla, M.; Chirlaque, M. D.; Ballesta, M.; Sundquist, K.; Sundquist, J.;

Weber, M.; Jordan, A.; Herrmann, C.; Mousavi, M.; Ryzhov, A.; Hoffmeister, M.; Brenner, H., Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. *Lancet Oncol* **2021,** *22* (7), 1002-1013.

4.   Favoriti, P.; Carbone, G.; Greco, M.; Pirozzi, F.; Pirozzi, R. E.; Corcione, F., Worldwide burden of colorectal cancer: a review. *Updates Surg* **2016,** *68* (1), 7-11.

5.   Denkert, C.; von Minckwitz, G.; Darb-Esfahani, S.; Lederer, B.; Heppner, B. I.; Weber, K. E.; Budczies, J.; Huober, J.; Klauschen, F.; Furlanetto, J.; Schmitt, W. D.; Blohmer, J. U.; Karn, T.; Pfitzner, B. M.; Kummel, S.; Engels, K.; Schneeweiss, A.; Hartmann, A.; Noske, A.; Fasching, P. A.; Jackisch, C.; van Mackelenbergh, M.; Sinn, P.; Schem, C.; Hanusch, C.; Untch, M.; Loibl, S., Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol* **2018,** *19* (1), 40-50.

6.   Maibach, F.; Sadozai, H.; Seyed Jafari, S. M.; Hunger, R. E.; Schenk, M., Tumor-Infiltrating Lymphocytes and Their Prognostic Value in Cutaneous Melanoma. *Front Immunol* **2020,** *11*, 2105.

7.   Paijens, S. T.; Vledder, A.; de Bruyn, M.; Nijman, H. W., Tumor-infiltrating lymphocytes in the immunotherapy era. *Cell Mol Immunol* **2021,** *18* (4), 842-859.

8.   Zakhartseva, L. M.; Yanovytska, M. A., Prognostic Value of Tumor Stroma Ratio in Triple Negative Breast Cancer. *Wiad Lek* **2021,** *74* (3 cz 2), 565-571.

9.   Kang, J.; Su, M.; Xu, Q.; Wang, C.; Yuan, X.; Han, Z., Tumour-stroma ratio is a valuable prognostic factor for oral tongue squamous cell carcinoma. *Oral Dis* **2023,** *29* (2), 628-638.

10.   Galon, J.; Mlecnik, B.; Bindea, G.; Angell, H. K.; Berger, A.; Lagorce, C.; Lugli, A.; Zlobec, I.; Hartmann, A.; Bifulco, C.; Nagtegaal, I. D.; Palmqvist, R.; Masucci, G. V.; Botti, G.; Tatangelo, F.; Delrio, P.; Maio, M.; Laghi, L.; Grizzi, F.; Asslaber, M.; D'Arrigo, C.; Vidal-Vanaclocha, F.; Zavadova, E.; Chouchane, L.; Ohashi, P. S.; Hafezi-Bakhtiari, S.; Wouters, B. G.; Roehrl, M.; Nguyen, L.; Kawakami, Y.; Hazama, S.; Okuno, K.; Ogino, S.; Gibbs, P.; Waring, P.; Sato, N.; Torigoe, T.; Itoh, K.; Patel, P. S.; Shukla, S. N.; Wang, Y.; Kopetz, S.; Sinicrope, F. A.; Scripcariu, V.; Ascierto, P. A.; Marincola, F. M.; Fox, B. A.; Pages, F., Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. *J Pathol* **2014,** *232* (2), 199-209.

11.   Francini, E.; Ou, F. S.; Lazzi, S.; Petrioli, R.; Multari, A. G.; Pesola, G.; Messuti, L.; Colombo, E.; Livellara, V.; Bazzurri, S.; Cherri, S.; Miano, S. T.; Wolfe, E. G.; Alberts, S. R.; Hubbard, J. M.; Yoon, H. H.; Francini, G., The prognostic value of CD3+ tumor-infiltrating lymphocytes for stage II colon cancer according to use of adjuvant chemotherapy: A large single-institution cohort study. *Transl Oncol* **2021,** *14* (2), 100973.

12.   Eriksen, A. C.; Sorensen, F. B.; Lindebjerg, J.; Hager, H.; dePont Christensen, R.; Kjaer-Frifeldt, S.; Hansen, T. F., The Prognostic Value of Tumor-Infiltrating lymphocytes in Stage II Colon Cancer. A Nationwide Population-Based Study. *Transl Oncol* **2018,** *11* (4), 979-987.

13.   Ye, L.; Zhang, T.; Kang, Z.; Guo, G.; Sun, Y.; Lin, K.; Huang, Q.; Shi, X.; Ni, Z.; Ding, N.; Zhao, K. N.; Chang, W.; Wang, J.; Lin, F.; Xue, X., Tumor-Infiltrating Immune Cells Act as a Marker for Prognosis in Colorectal Cancer. *Front Immunol* **2019,** *10* (1664-3224 (Electronic)), 2368.

14.   Salgado, R.; Denkert, C.; Demaria, S.; Sirtaine, N.; Klauschen, F.; Pruneri, G.; Wienert, S.; Van den Eynden, G.; Baehner, F. L.; Penault-Llorca, F.; Perez, E. A.; Thompson, E. A.; Symmans, W. F.; Richardson, A. L.; Brock, J.; Criscitiello, C.; Bailey, H.; Ignatiadis, M.; Floris, G.; Sparano, J.; Kos, Z.; Nielsen, T.; Rimm, D. L.; Allison, K. H.; Reis-Filho, J. S.; Loibl, S.; Sotiriou, C.; Viale, G.; Badve, S.; Adams, S.; Willard-Gallo, K.; Loi, S.; International, T. W. G., The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann Oncol* **2015,** *26* (2), 259-71.

15.   Van Bockstal, M. R.; Francois, A.; Altinay, S.; Arnould, L.; Balkenhol, M.; Broeckx, G.; Burgues, O.; Colpaert, C.; Dedeurwaerdere, F.; Dessauvagie, B.; Duwel, V.; Floris, G.; Fox, S.; Gerosa, C.; Hastir, D.; Jaffer, S.; Kurpershoek, E.; Lacroix-Triki, M.; Laka, A.; Lambein, K.; MacGrogan, G. M.; Marchio, C.;

Martin Martinez, M. D.; Nofech-Mozes, S.; Peeters, D.; Ravarino, A.; Reisenbichler, E.; Resetkova, E.; Sanati, S.; Schelfhout, A. M.; Schelfhout, V.; Shaaban, A.; Sinke, R.; Stanciu-Pop, C. M.; van Deurzen, C. H. M.; Van de Vijver, K. K.; Van Rompuy, A. S.; Vincent-Salomon, A.; Wen, H. Y.; Wong, S.; Bouzin, C.; Galant, C., Interobserver variability in the assessment of stromal tumor-infiltrating lymphocytes (sTILs) in triple-negative invasive breast carcinoma influences the association with pathological complete response: the IVITA study. *Mod Pathol* **2021,** *34* (12), 2130-2140.

16.   Hendry, S.; Salgado, R.; Gevaert, T.; Russell, P. A.; John, T.; Thapa, B.; Christie, M.; Van De Vijver, K.; Estrada, M. V.; Gonzalez-Ericsson, P. I., Assessing tumor infiltrating lymphocytes in solid tumors: A practical review for pathologists and proposal for a standardized method from the International Immuno-Oncology Biomarkers Working Group: Part 1: Assessing the host immune response, TILs in invasive breast carcinoma and ductal carcinoma in situ, metastatic tumor deposits and areas for further research. *Advances in anatomic pathology* **2017,** *24* (5), 235.

17.   Iseki, Y.; Shibutani, M.; Maeda, K.; Nagahara, H.; Fukuoka, T.; Matsutani, S.; Kashiwagi, S.; Tanaka, H.; Hirakawa, K.; Ohira, M., A new method for evaluating tumor-infiltrating lymphocytes (TILs) in colorectal cancer using hematoxylin and eosin (H-E)-stained tumor sections. *PLoS One* **2018,** *13* (4), e0192744.

18.   Fuchs, T. L.; Sioson, L.; Sheen, A.; Jafari-Nejad, K.; Renaud, C. J.; Andrici, J.; Ahadi, M.; Chou, A.; Gill, A. J., Assessment of Tumor-infiltrating Lymphocytes Using International TILs Working Group (ITWG) System Is a Strong Predictor of Overall Survival in Colorectal Carcinoma: A Study of 1034 Patients. *Am J Surg Pathol* **2020,** *44* (4), 536-544.

19.   Tramm, T.; Di Caterino, T.; Jylling, A. B.; Lelkaitis, G.; Laenkholm, A. V.; Rago, P.; Tabor, T. P.; Talman, M. M.; Vouza, E.; Scientific Committee of Pathology, D. B. C. G., Standardized assessment of tumor-infiltrating lymphocytes in breast cancer: an evaluation of inter-observer agreement between pathologists. *Acta Oncol* **2018,** *57* (1), 90-94.

20.   Sari, C. T.; Gunduz-Demir, C., Unsupervised Feature Extraction via Deep Learning for Histopathological Classification of Colon Tissue Images. *IEEE Trans Med Imaging* **2019,** *38* (5), 1139-1149.

21.   Ao, T.; Kajiwara, Y.; Yonemura, K.; Shinto, E.; Mochizuki, S.; Okamoto, K.; Kishi, Y.; Ueno, H., Morphological consistency of desmoplastic reactions between the primary colorectal cancer lesion and associated metastatic lesions. *Virchows Arch* **2020,** *477* (1), 47-55.

22.   Zhu, Y.; Jin, Z.; Qian, Y.; Shen, Y.; Wang, Z., Prognostic Value of Tumor-Stroma Ratio in Rectal Cancer: A Systematic Review and Meta-analysis. *Front Oncol* **2021,** *11*, 685570.

23.   Souza da Silva, R. M.; Queiroga, E. M.; Paz, A. R.; Neves, F. F. P.; Cunha, K. S.; Dias, E. P., Standardized Assessment of the Tumor-Stroma Ratio in Colorectal Cancer: Interobserver Validation and Reproducibility of a Potential Prognostic Factor. *Clin Pathol* **2021,** *14*, 2632010X21989686.

24.   Sullivan, L.; Pacheco, R. R.; Kmeid, M.; Chen, A.; Lee, H., Tumor Stroma Ratio and Its Significance in Locally Advanced Colorectal Cancer. *Curr Oncol* **2022,** *29* (5), 3232-3241.

25.   Gisev, N.; Bell, J. S.; Chen, T. F., Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Res Social Adm Pharm* **2013,** *9* (3), 330-8.

26.   Koo, T. K.; Li, M. Y., A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* **2016,** *15* (2), 155-63.

27.   Landis, J. R.; Koch, G. G., The measurement of observer agreement for categorical data. *Biometrics* **1977,** *33* (1), 159-74.

28.   Denkert, C.; Wienert, S.; Poterie, A.; Loibl, S.; Budczies, J.; Badve, S.; Bago-Horvath, Z.; Bane, A.; Bedri, S.; Brock, J.; Chmielik, E.; Christgen, M.; Colpaert, C.; Demaria, S.; Van den Eynden, G.; Floris, G.; Fox, S. B.; Gao, D.; Ingold Heppner, B.; Kim, S. R.; Kos, Z.; Kreipe, H. H.; Lakhani, S. R.; Penault-Llorca, F.; Pruneri, G.; Radosevic-Robin, N.; Rimm, D. L.; Schnitt, S. J.; Sinn, B. V.; Sinn, P.; Sirtaine, N.; O'Toole, S. A.; Viale, G.; Van de Vijver, K.; de Wind, R.; von Minckwitz, G.; Klauschen, F.; Untch, M.; Fasching, P. A.; Reimer, T.; Willard-Gallo, K.; Michiels, S.; Loi, S.; Salgado, R., Standardized evaluation of

tumor-infiltrating lymphocytes in breast cancer: results of the ring studies of the international immuno-oncology biomarker working group. *Mod Pathol* **2016,** *29* (10), 1155-64.

29.   Swisher, S. K.; Wu, Y.; Castaneda, C. A.; Lyons, G. R.; Yang, F.; Tapia, C.; Wang, X.; Casavilca, S. A.; Bassett, R.; Castillo, M.; Sahin, A.; Mittendorf, E. A., Interobserver Agreement Between Pathologists Assessing Tumor-Infiltrating Lymphocytes (TILs) in Breast Cancer Using Methodology Proposed by the International TILs Working Group. *Ann Surg Oncol* **2016,** *23* (7), 2242-8.

30.   Dang, H. A.-O.; van Pelt, G. W.; Haasnoot, K. J.; Backes, Y.; Elias, S. G.; Seerden, T. C.; Schwartz, M. P.; Spanier, B. W.; de Vos Tot Nederveen Cappel, W. H.; van Bergeijk, J. D.; Kessels, K. A.-O.; Geesing, J. M.; Groen, J. N.; Ter Borg, F.; Wolfhagen, F. H.; Seldenrijk, C. A.; Raicu, M. G.; Milne, A. N.; van Lent, A. U.; Brosens, L. A.; Offerhaus, G. J. A.; Siersema, P. D.; Tollenaar, R. A.; Hardwick, J. A.-O.; Hawinkels, L. J.; Moons, L. M.; Lacle, M. M.; Mesker, W. E.; Boonstra, J. J., Tumour-stroma ratio has poor prognostic value in non-pedunculated T1 colorectal cancer: A multi-centre case-cohort study. (2050-6414 (Electronic)).