Annals of Diagnostic Pathology 43 (2019) 151420

Contents lists available at ScienceDirect

Annals of Diagnostic Pathology

journal homepage: www.elsevier.com/locate/anndiagpath

Original Contributions

Tumor budding in colorectal carcinoma: An institutional interobserver reliability and prognostic study of colorectal adenocarcinoma cases

Sean Hacking[a,1], Mallorie Angert[a,1], Cao Jin[a], Myriam Kline[b], Neha Gupta[a], Margaret Cho[a], Rebecca Thomas[a], Lili Lee[a], Hector Chavarria[a], Mansoor Nasim[a,*]

[a] Department of Pathology and Laboratory Medicine, Zucker School of Medicine at Hofstra Northwell, United States of America
[b] Biostatistics Unit, Feinstein Institute for Medical Research, Northwell Health, United States of America

ARTICLE INFO

Keywords:
Colorectal carcinoma
Interobserver reliability
Tumor budding

ABSTRACT

Background: Colorectal carcinomas are one of the most commonly diagnosed malignancies. There are many prognostic factors relating to clinical course and disease progression, including tumor stage, metastasis, and tumor budding. In 2016, the International Tumor Budding Consensus Conference (ITBCC) created a system to uniformly assess tumor budding. This system includes a 3-tier system for the grading of tumor budding. In the past, there lacked uniform consensus, however the general grading practice was based on a 2-tiered system. Given that tumor budding is considered to have prognostic value, the accuracy and reproducibility of its assessment is vital. Our study aims to look at interobserver agreement in the scoring of tumor budding.

Design: A total of 233 cases of colorectal carcinoma diagnosed in our health system were retrospectively analyzed and routine H&E stained slides of these cases were collected. A representative slide for tumor budding was selected per case. Four investigators with different levels of experience and expertise evaluated the selected slide of each case for tumor budding. Scoring was based on the ITBCC protocol. Clinico-pathological data was collected for each case and analyzed with tumor budding scores. Tumor budding scores per individual investigator and consensus tumor budding score were compared to patient and tumor characteristics including patient survival, tumor grade, tumor stage, and lymph node status.

Results: Inter-observer agreement was calculated using Gwet's Agreement Coefficient (AC_1) and associated 95% confidence intervals was used to compare the ratings made by 4 pathologists. Overall, there was variation among pathologists in tumor budding score (Gwet's agreement coefficient = 0.25 and 0.326 for 3-tier and 2-tier grading system, respectively). Results show higher reliability with the 2-tier system compared to the 3-tier system. Tumor stage was significantly associated with budding score for all individual investigators and the consensus value (p value < 0.001).

Conclusion: There is low inter-observer agreement in the assessment of tumor budding in colorectal carcinoma. This suggests that it is difficult to uniformly grade tumor budding and that our classification system needs improvement. We found that the older 2-tier system (Hase et al.) results in slightly higher inter-observer agreement than the recently proposed 3-tier grading system (ITBCC, 2016), though both systems lead to suboptimal agreement. Worth noting is that observers with subspecialty GI training and more work experience had higher inter-observer agreement. Our results showed that subspecialty training tends to increase agreement more than overall work experience. In addition, our exploratory results showed that there is an association of tumor budding score to tumor stage. While increasing refinement in classification, the 3-tiered system resulted in decreased agreement in tumor budding assessment. Clearly, there is more work to be done in the identification and quantification of tumor buds.

1. Introduction

Colorectal carcinoma (CC) is one of the most frequently diagnosed cancers; being the second and third most common malignant cancer in women and men, respectively [1]. There are many prognostic factors relating to its clinical course and disease progression including tumor

* Corresponding author at: 6 Ohio Drive - Suite 202, Lake Success, New York 11042, United States of America.
E-mail address: Mnasim@northwell.edu (M. Nasim).
[1] Sean Hacking and Mallorie Angert contribute equally to this paper.

https://doi.org/10.1016/j.anndiagpath.2019.151420

stage, metastasis and tumor budding [2,3]. TNM classification is pragmatic in assessing prognosis and managing disease in CC [3]. However, there are flaws in the system and several subgroups show discrepancies in their expected clinical course compared to the TNM stage of their disease [4,5]. For example, one study found that higher stage rectal tumors (IIIA) actually have a more favorable survival outcome then lower stage tumors (stage II) [5].

In 2016, the International Tumor Budding Consensus Conference (ITBCC) proposed a system to uniformly assess tumor budding [6]. This classification includes a 3-tier system for the grading of tumor budding [6]. In the past, there lacked uniform consensus even though the general grading practice was based on the 2-tiered system originally proposed by Hase et al. [7]. In this system cases were divided into two groups according to the number of tumor budding foci in the densest field using a 20× objective lens [7]. Counts of 0–9 were classified as low-grade, while counts of 10+ were classified as high-grade budding [2,7]. Other more complex grading systems were proposed over previous years [8,9], however, they were not widely adopted [8,9]. Importantly, many studies have demonstrated tumor budding as a robust biomarker which provides prognostic information (independently from TNM classification) [10-12].

With regard to the assessment of tumor budding, various studies have reported different rates of inter-observer reliability [13,14], with most studies reporting low agreement [13,14]. Given that tumor budding is considered to have prognostic value as previously discussed, the accuracy and reproducibility of its assessment is vital. Our study aim is to examine inter-observer reliability in the scoring of tumor budding. Special consideration will be given regarding the effect of subspecialty training, work experience and the new three tier grading system (based on the International Tumor Budding Consensus Conference (ITBCC)).

2. Materials and methods

2.1. Design

A total of 233 cases of colorectal carcinoma diagnosed in our health system were retrospectively analyzed and routine H&E stained slides of these cases were collected. Worth mentioning is that neuroendocrine tumors/carcinomas were excluded from this study. Among patients with colorectal cancer, the procedures undertaken were as follows: 92 (40%) had a right hemicolectomy, 42 (18%) had a low anterior resection, 18 (7.8%) had a rectosigmoidectomy, 15(6.5%) had a left hemicolectomy, and 13 (5.6%) had a transverse colectomy. The remaining procedures accounted for < 5%. A representative slide for tumor budding was selected per case. Four investigators were selected based on different levels of experience and expertise. We defined work experience as follows: 'limited' for < 5 years, 'moderate' for 5–10 years, and 'extensive' for more than 10 years. Subspecialty training in Gastrointestinal (GI) pathology was defined as either completed (Fellowship trained) or not. The pathologists evaluated the selected slide for each case of tumor budding and were blinded to each other's results.

In the assessment, a detailed search was done for the area having the highest grade of tumor budding. After that, the counting of the buds took place in the hotspot region (20× objective lens). Scoring was preformed based on the 3-tier grading system (ITBCC protocol) and the 2-tier grading system [7]. The 3-tier grading system was divided into three grades: Bd1: 0–4 buds, Bd2: 5–9 buds and Bd3: 10 or more buds (Fig. 1). While the 2-tier system was divided into two grades: low grade (< 10 buds) and high-grade (≥10 buds).

Clinico-pathological data was collected for each case and analyzed in relation to tumor budding scores. In exploratory analyses [15], tumor budding scores per individual investigator and consensus tumor budding scores were compared to tumor stage and mortality. All samples were obtained with informed consent after approval by the Institutional Review Board of the respective hospitals they were treated in.

2.2. Statistical analysis

The primary objective of the statistical analysis was to calculate agreement among observers (using both the 2-tiered and 3-tiered grading systems) quantifying tumor budding. A second, exploratory objective, was to test whether grades given to tumor budding were predictive of selected clinical outcomes. The primary endpoints were the grades assigned to tumor budding. Secondary endpoints were pathological stage and mortality.

Descriptive statistics such as frequencies and percentages were calculated for the categorical variables; means and standard deviations were calculated for the numerical variables. Gwet's Agreement Coefficient (AC_1) was used to measure agreement among observers using the 2-tier and 3-tier grading systems to quantify tumor budding as presented on slides. Simple linear regression was used to determine if tumor budding was associated with pathologic stage; logistic regression was used to determine if tumor grade was associated with mortality. A p-value of < 0.05 was used to indicate statistical significance, however, a Bonferroni correction was used to adjust for multiple comparisons and a result was considered significant if $p < 0.01$. AgreeStat 2015 v6.1 was used to carry out the reliability analysis (2009–2016 Advanced Analytics, LLC). The data analysis for the secondary objective was generated using SAS/STAT software, Version 9.4 of the SAS System for Windows. Copyright © 2002–2012 SAS Institute Inc., Cary, NC, USA.

3. Results

3.1. Diagnostic reproducibility

The results of the reliability analysis revealed that there was 'fair' agreement among pathologists with respect to tumor budding score ($AC_1 = 0.25$ and $AC_1 = 0.326$ for 3-tier and 2-tier grading systems, respectively). Results demonstrated higher reliability with the 2-tier system compared to the 3-tier system (Table 1). Table 2 lists the agreement between individual observers and the consensus score (agreement is higher due to the inherent correlation built in between the single observer vs. the consensus score, which includes the scores of the single observer). Specifically, observers 3 and 4 had the highest inter-observer agreement and were GI pathology fellowship trained with moderate and extensive experience respectively. Observer 1 had fellowship training but limited work experience. Finally, observer 2 had no GI fellowship training but had extensive work experience. Included is a photo-micrograph from a case with marked discrepancy (based on the lack of agreement among the evaluators) (Fig. 2). Discrepancy in this particular case was likely precipitated by marked inflammation with indistinct borders of the inflammatory/tumor cells.

3.2. Prognostic relevance

There was statistical significance predicting tumor budding score to tumor stage for all individual investigators and the consensus value (p value < 0.001, Table 3). A higher tumor score was associated with a higher pathologic stage. With regard to mortality, inspection of the results of the individual investigators revealed that none of the ratings of the observers were significantly associated with mortality.

4. Discussion

Tumor budding is an important and well known adverse prognostic factor in colorectal carcinoma [10-12]. Based on this knowledge, it stands to reason that high inter-observer agreement is crucial for clinical decision-making. Yet, the results of our study showed that agreement among 4 observers on the assessment of tumor budding using either the 2-tiered or 3-tiered systems was less than high. In our exploratory analysis, when tumor budding scores were modeled to predict tumor stage, a significant association was found for all individual

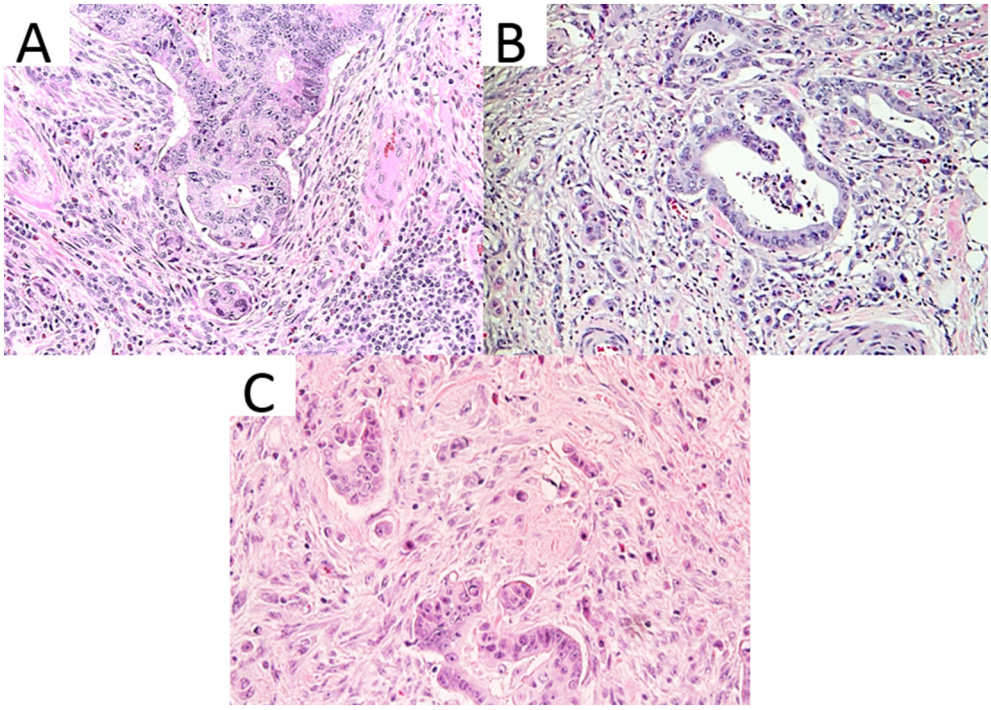

Fig. 1. H&E stained slides (20×) representative of tumor budding grades. A. Low grade tumor budding (Bd1); B. intermediate grade tumor budding (Bd2); C. high grade tumor budding (Bd3).

Table 1
Inter-observer results of the 2-tier and 3-tier tumor budding systems.

	3-Tier tumor budding rating system	2-Tier tumor budding rating system
Gwet's AC_1 (95% CI)	0.25 (0.19 to 0.31)	0.329 (0.25 to 0.40)

Table 2
Agreement by observer vs. consensus.

Agreement by observer vs. consensus	3-Tier tumor budding rating
	233 slides
	Gwet's AC_1 (95% CI)
Subspecialty trained, limited work experience (1) vs. consensus	0.501 (0.409 to 0.593)
Not subspecialty trained, extensive work experience (2) vs. consensus	0.443 (0.349 to 0.536)
Subspecialty trained, moderate work experience (3) vs. consensus	0.630 (0.545 to 0.714)
Subspecialty trained, extensive work experience (4) vs. consensus	0.653 (0.571 to 0.735)

observers (and the consensus value). These findings are supported by other studies, which have also found that a higher grade of tumor budding is correlated with a higher CC tumor stage [14]. However, when examining mortality, none of the ratings of the observers were able to demonstrate an association between tumor budding score and death.

As discussed previously, our study found low agreement in tumor budding scores among pathologists. Reliability was higher for the 2-tier system compared to the 3- tier system. Many studies have highlighted the importance of tumor budding in CC [10-16]. In some cases, presence of prominent tumor budding in an adenocarcinoma may have implications for management, such as additional resection [17], while

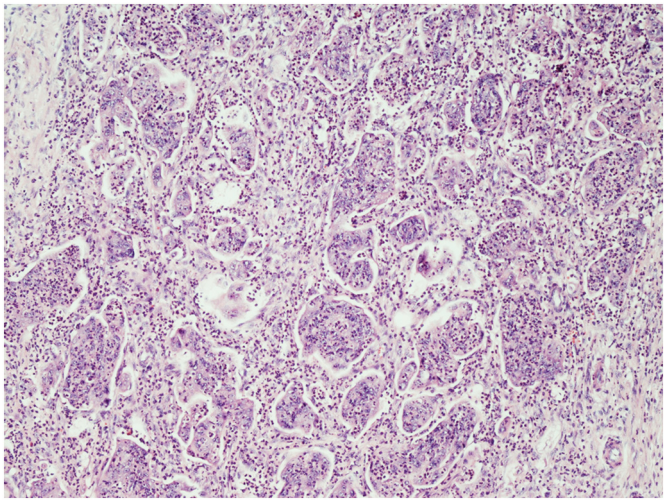

Fig. 2. H&E stained slide (10×) representative of a case with marked discrepancies, likely precipitated by marked inflammation with indistinct borders of the inflammatory/tumor cells.

Table 3
Tumor stage as a function of budding score per observer.

	Parameter estimate	Standard error	Tvalue	p-Value
Observer 1	0.43848	0.10705	4.10	< 0.0001
Observer 2	0.55427	0.09596	5.78	< 0.0001
Observer 3	0.35013	0.09698	3.61	0.0004
Observer 4	0.39733	0.10433	3.81	0.0002
Consensus	0.41603	0.09524	4.37	< 0.0001

in other cases it may serve as a prognostic factor alone [17]. Given these clinical implications, the ability to accurately and consistently assess tumor budding is of great importance. However, regardless of a 2-tier or 3-tier tumor budding assessment system, visualization and

quantification of tumor budding is proving elusive, as demonstrated by the “fair” level of agreement found among the observers in this study. These results persisted despite credentials, training, or experience. Observers are, for the most part, not quantifying tumor buds similarly when asked to assess the identical case slides. The source of this inter-observer variability is likely multifactorial but the findings of our study have shown that rather than to improve the assessment of tumor budding, increasing the number of categories from 2 to 3 has resulted in a worsening of reliability.

As our results underscore, the assessment of tumor budding is a difficult process [12,14]. A possible explanation for the lack of reproducibility is the current education and training process (used to access tumor budding on H&E slides). Some of the observers in our study received sub-specialty training in GI pathology, other observers did not. Additionally, observers had different amounts of work experience. Our results showed that observers with extensive or moderate work experience and subspecialty training tended to perform better. Comparing the remaining two observers: one had extensive work experience and no fellowship training, while the other had limited work experience and GI fellowship training. Our findings suggest that subspecialty training may be more useful in attaining higher diagnostic accuracy than overall work experience.

A suggestion for improvement could be the development of high quality and free online teaching videos. Additionally, the College of American Pathologists (CAP) could offer a course in tumor budding assessment. An improvement in tumor budding assessment may lead to improved clinical outcomes and it would be important for a future study to evaluate the association between tumor buds and clinical outcomes.

5. Conclusion

There is low inter-observer agreement in the assessment of tumor budding in colorectal carcinoma. This suggests that it is difficult to uniformly grade tumor budding and that our classification system needs improvement. We found that the older 2-tier system [7] results in slightly higher inter-observer agreement than the recently proposed 3-tier grading system (International Tumor Budding Consensus Conference 2016 [6], though both systems lead to suboptimal agreement. Worth noting is that observers with subspecialty GI training and more work experience had higher inter-observer agreement. Our results showed that subspecialty training tends to increase agreement more than overall work experience. In addition, our exploratory results showed that there is an association of tumor budding score to tumor stage. Tumor budding has been shown to have prognostic implications [10-12]. Therefore, we proffer that it is essential to increase inter-observer agreement and reproducibility first. At this time, there is a need for further training and education to more precisely detect tumor budding. While increasing refinement in classification, the 3-tiered system resulted in decreased agreement in tumor budding assessment. Clearly, there is more work to be done in the identification and quantification of tumor buds.

Declaration of competing interest

The authors report no conflicts of interest in this work.

Acknowledgments

We thank Claudine Alexis and Kathy Quinn for their role as research coordinators.

References

[1] Marley AR, Nan H. Epidemiology of colorectal cancer. Int J Mol Epidemiol Genet 2016;7(3):105–14. [Published 2016 Sep 30].
[2] Koelzer VH, Langer R, Zlobec I, Lugli A. Tumor budding in upper gastrointestinal carcinomas. Front Oncol 2014;4(216). https://doi.org/10.3389/fonc.2014.00216. [Published 2014 Aug 14].
[3] Kulendran M, Stebbing JF, Marks CG, Rockall TA. Predictive and prognostic factors in colorectal cancer: a personalized approach. Cancers (Basel) 2011;3(2):1622–38. https://doi.org/10.3390/cancers3021622. [Published 2011 Mar 29].
[4] Elfaedy O, Owens P, Aakif M, Mansour E. Discrepancy in colorectal cancer staging: a single center experience. World J Surg Surgical Res 2018;1:1054.
[5] Huang B, Mo S, Zhu L, Xu T, Cai G. The survival and clinicopathological differences between patients with stage IIIA and stage II rectal cancer: an analysis of 12,036 patients in the SEER database. Oncotarget 2016;7(48):79787–96. https://doi.org/10.18632/oncotarget.12970.
[6] Validation of the International Tumor Budding Consensus Conference 2016 recommendations on tumor budding in stage I-IV colorectal cancer. Hum Pathol 2019 Mar;85:145–51.
[7] Hase K, Shatney C, Johnson D, Trollope M, Vierra M. Prognostic value of tumor “budding” in patients with colorectal cancer. Dis Colon Rectum Jul 1993;36(7):627–35.
[8] Ueno H, Murphy J, Jass JR, Mochizuki H, Talbot IC. Tumour ‘budding’ as an index to estimate the potential of aggressiveness in rectal cancer. Histopathology Feb 2002;40(2):127–32.
[9] Nakamura T, Mitomi H, Kikuchi S, Ohtani Y, Sato K. Evaluation of the usefulness of tumor budding on the prediction of metastasis to the lung and liver after curative excision of colorectal cancer. Hepatogastroenterology Sep-Oct 2005;52(65):1432–5.
[10] Grizzi F, Celesti G, Basso G, Laghi L. Tumor budding as a potential histopathological biomarker in colorectal cancer: hype or hope? World J Gastroenterol 2012;18(45):6532–6. https://doi.org/10.3748/wjg.v18.i45.6532.
[11] Graham RP, Vierkant RA, Tillmans LS, et al. Tumor budding in colorectal carcinoma: confirmation of prognostic significance and histologic cutoff in a population-based cohort. Am J Surg Pathol 2015;39(10):1340–6. https://doi.org/10.1097/PAS.0000000000000504.
[12] Dawson Heather, Assarzadegan Naziheh, Riddell Robert, Kirsch Richard, Blank Annika, Zlobec Inti, et al. Tumor budding is a strong predictor of disease-free survival in stage II colorectal cancer: validation study based on the International Tumor Budding Consensus Conference (ITBCC) recommendations. J Clin Oncol 2017;35:594. https://doi.org/10.1200/JCO.2017.35.4_suppl.594.
[13] Rogers AC, Winter DC, Heeney A, et al. Systematic review and meta-analysis of the impact of tumour budding in colorectal cancer. Br J Cancer 2016;115(7):831–40. https://doi.org/10.1038/bjc.2016.274.
[14] Martin B, Schäfer E, Jakubowicz E, et al. Interobserver variability in the H&E-based assessment of tumor budding in pT3/4 colon cancer: does it affect the prognostic relevance? Virchows Arch Aug 2018;473(2):189–97. https://doi.org/10.1007/s00428-018-2341-1.
[15] Kai K, Aishima S, Aoki S, Takase Y, Uchihashi K, Masuda M, et al. Cytokeratin immunohistochemistry improves interobserver variability between unskilled pathologists in the evaluation of tumor budding in T1 colorectal cancer. Pathol Int 2016;66:75–82.
[16] Horcic M, Koelzer VH, Karamitopoulou E, et al. Tumor budding score based on 10 high-power fields is a promising basis for a standardized prognostic scoring system in stage II colorectal cancer. Hum Pathol 2013;44(5):697–705.
[17] Cho S-J, Kakar S. Tumor budding in colorectal carcinoma: translating a morphologic score into clinically meaningful results. Arch Pathol Lab Med Aug 2018;142(8):952–7. https://doi.org/10.5858/arpa.2018-0082-RA.