Paving the way to tumor budding assessment using digital pathology: a pilot study in Timișoara City (Romania)

ALIS LILIANA CARMEN DEMA1), SORINA TĂBAN1), AURA JURESCU1), ADELINA-ROXANA GHEJU1), ADRIAN OVIDIU VĂDVA1), CIPRIAN CONSTANTIN DUȚĂ2), OCTAVIAN FULGER LAZĂR2)

1)Department of Pathology, “Victor Babeș” University of Medicine and Pharmacy, Timișoara, Romania
2)Department of Surgery II, “Victor Babeș” University of Medicine and Pharmacy, Timișoara, Romania

Abstract
Background and Aim: The outcome for some colorectal cancer patients correlates poorly with classical prognostic factors, like tumor stage. Tumor budding (TB) is a promising and intensely studied new prognostic factor. We aimed to evaluate the reliability of bud counting on Hematoxylin–Eosin (HE)-stained and immunohistochemically (IHC)-stained scanned slides. Materials and Methods: We evaluated 21 cases of robotic surgery colorectal cancer specimens that were submitted to the Department of Pathology, Emergency County Hospital, Timișoara, Romania. TB was assessed by one experienced (R3) and two junior pathologists (R1, R2), in 10 circular areas at 20× (0.785 mm²) on scanned HE-stained and IHC-stained [cytokeratin (CK) AE1/AE3] slides. Interobserver agreement (Cohen’s kappa) and intraclass correlation coefficient (ICC) were calculated. Results: In the case of HE-stained slides, the inter-item correlation matrix showed values between 0.632 and 0.84, while the ICC on average measures for consistency showed very good correlation [ICC: 0.887, 95% confidence interval (CI): 0.765–0.95]. The inter-item correlation matrix for IHC-stained slides comprised values between 0.864 and 0.921, while the ICC for average measures for consistency yielded an excellent value (ICC: 0.95, 95% CI: 0.896–0.978). We identified higher values for budding scores on IHC-stained slides, in comparison to the HE-stained slides: in 19/21 cases for R1 (average increase of 234.85%), 16/21 cases for R2 (average increase of 114.14%), and 20/21 cases for R3 (average increase of 66.92%). Conclusions: We consider the method of buds counting in 10 microscopic fields on scanned slides to be reliable and valuable. TB counts are higher on IHC-stained slides and associate a better interobserver agreement.

Keywords: colorectal carcinoma, tumor budding, immunohistochemistry, digital slides.

Introduction
The colorectal cancer (CRC) incidence ranks it as the third most common type of cancer in men and second in women, placing it on the fourth place on the most common cancer-related death list [1]. Recent European estimates place it as the second most frequent cancer, behind breast malignancies, with 447,000 new cases reported annually, ranking second also on the most common cancer-related death list, following lung cancer [2].

As with other cancer sites, the stage of the tumor, assessed according to the American Joint Committee on Cancer/International Union against Cancer (AJCC/IUAC) TNM system, is still considered the main prognostic factor in CRC [3, 4]. The real value of this parameter is arguable, since certain studies have shown that some stage I/II CRC behave aggressively [5], while a significant number of stage III CRC have a favorable outcome, despite being metastatic to the lymph nodes [6].

In this setting, the need arises to identify novel prognostic factors that can improve the stratification of the patients with CRC, to facilitate a better treatment regime, thus avoiding under- or over-treatment of certain cases.

We are in the age of molecular classifications, where these assessments represent useful diagnostic, predictive and/or prognostic factors. Although these classifications now play major roles in breast cancer [7], tumors of the central nervous system [8] and even CRC [9, 10], a series of histopathological elements are the subject of highly discussed and intensely studied prognostic and predictive factors. They can be evaluated on Hematoxylin–Eosin (HE) and immunohistochemically (IHC)-stained slides, thus making them more accessible, easier to evaluate and cheaper to obtain. Some of these factors like the poorly differentiated clusters (PDC) of cancer cells [11–14], the tumor budding (TB) [15–18], the tumor border configuration/invasion pattern [19, 20], lymph node micro-metastases [21], or the polyploid giant tumor cells [22] could supplement or even replace the classical prognostic factors in CRC.

Tumor buds are defined as small clusters of up to five cells at the invasive front of the tumor [23]. From all the previously mentioned elements, a fair amount of research grants potential value to these as a prognostic factor in CRC, arguing for the necessity to introduce it in the pathological report and in the TNM classification of CRC [4, 24]. To achieve this, there is a need to set up a reproducible, widely accepted and used system for quantification and reporting of TB.

Microscopy analysis techniques are evolving, and digital pathology is gaining more ground in the diagnostic field, e.g., external quality assurances schemes [25], as a learning and training tool for grading systems like Gleason score in prostate cancer [26], providing a virtual environment for remote consultation [27], for teaching purposes [28], as well as in the field of prognosis and therapy. The latter is nicely described in the computer-aided assessments of predictive and prognostic factors on digital slides in breast cancer [29].
Whole slide imaging (WSI), which implies digital scanning of glass slides, is recurrently used in CRC to evaluate some prognostic factors like TB [4, 30, 31], PDC, lymphatic vessel invasion, lymphatic vessel density [31, 32], intra- and peritumoral lymphocytic infiltrate [33]. WSI, based on computer-assisted image analysis, allows automatic quantification of these parameters, leading to increased accuracy of the results. Thus, time can be saved and the interobserver variability improved, given that a standardization of assessment and quantification is provided.

The aim of this study was to assess TB in a CRC patients group that underwent robotic surgery, with the prospect of implementing it as a prognostic factor in the pathology report. We analyzed TB from the perspective of the assessment modality and the interobserver reliability for counting buds on HE and IHC-stained scanned slides.

Materials and Methods

We performed a retrospective study on CRC cases that were previously diagnosed on endobiopsy specimens and then resected at the “Pius Brînzeu” Emergency County Hospital, Timișoara, Romania. The study included cases that were diagnosed according to the World Health Organization (WHO) guidelines [34] and underwent robotic surgery (da Vinci Xi® Surgical System) between 07/2015–07/2016. We excluded patients that underwent chemotherapy or radiotherapy before surgery.

Patient demographics and clinical data were collected from the specimen submission slip, patient’s chart and hospital records.

We considered right-sided CRC the tumors located at the following levels: caecum, ascending colon, hepatic flexure and proximal two-thirds of the transverse colon (derived from the hindgut) and left-sided CRC the tumors at the distal third of the transverse colon, splenic flexure, descending colon, sigmoid colon and rectum (derived from the midgut) [35].

The histopathology slides cut from the corresponding paraffin blocks were reevaluated, independently, by two senior pathologists (ST and AD), to establish the tumor subtype, histological grade (G), depth of invasion (pT), the invasion pattern (infiltrative or pushing), the presence of perineural invasion (PnI), lymphovascular invasion (LVI) and lymph node metastasis (pN).

To circumvent the issues related to the problematic grading when using the four-tier system, we decided to reclassify all the tumors in a two-tier system, based on gland formation: low grade, with >50% gland formation, corresponding to G1–G2, and high grade, ≤50% gland formation (G3–G4) [3, 36].

The pathological stage of the CRC cases was established according to the seventh edition of AJCC Cancer Staging Manual [37].

We assessed TB on all of the cases included in the study according to the International Tumor Budding Consensus Conference (ITBCC) recommendation [38]: as a single tumor cell or a cell cluster of up to four tumor cells (1–4 tumor cells) that did not represent areas of glandular disruption made by the inflammatory infiltrate.

The first step in TB assessment was performed by a junior pathologist. She evaluated all the HE-stained slides and selected, for each of the cases, the slide and corresponding paraffin block, on which the highest TB count was identified on the invasion front.

New slides were then cut from the paraffin blocks at 4 μm thickness and IHC labeled with anti-cytokeratin (CK) AE1/AE3 antibody (Novocstra, ready-to-use, code PA0909). Immunoreactions were developed using a polymer-based detection kit (Novolink, Novocastra code RE7280-K), visualized with 3,3’-Diaminobenzidine (DAB) and counterstained with Hematoxylin. All IHC stainings were performed on a Leica Bond Max autostainer according to the manufacturer’s protocol with heat-induced epitope retrieval at pH 9.

Assessment of tumor budding on digitized slides

The TB analysis was performed by two junior pathologists (AJ – R1 and AG – R2) and one experienced gastrointestinal pathologist (ST – R3). The assessment was performed on WSI, at the tumor invasion front. Each pathologist first examined the HE dataset, and then the corresponding IHC counterpart, according to the following protocol:

The 21 HE-stained and corresponding IHC-stained slides (CK AE1/AE3) were scanned on a Leica Aperio AT2 scanner, using a 40× magnification setting. The resulting digital slides were then stored on a local server running Leica Digital Image Hub, a software which allows remote access, via web interface, to all available digital slides located on the server. For each independent evaluator, a unique login was created to have limited access to his/her set of digital slides. Each evaluator first drew 10 circles to simulate the area covered by the default 21× objective (0.785 mm²) and placed them along the tumor invasion front, thus selecting areas in which to count buds. All counts were collected in an MS Excel generated spreadsheet. For all slides, the first count was performed in the area that was visually deemed to have the highest count (visual hotspot).

Bias management – evaluator blinding

The name of each digital slide file was set as the default consecutive number, as per scanner setting. No rename was performed to directly identify the case number. There was no direct labeling linking the HE and corresponding IHC slide. Each evaluator had their own individual set of scanned slides, and all annotations were viewable only by the person generating them and the system administrator. All the counts were collected in a single file and the link between digital slide number and case number was created.

The classification of the cases in respect to the average count of TB was performed according to a model similar to the one proposed by Kawachi et al. [39]: tumors with ≥10 TB were graded as G3Bd, those with 6–9 TB as G2Bd, while tumors with ≤5 TB were graded as G1Bd. Following this, we also graded the tumors on a two-tier system, similar to Kawachi et al. [39]: low GBd, when ≤9TB were identified and high GBd, with ≥10TB.
Statistical analysis

Interobserver agreement (Cohen’s kappa) and intraclass correlation coefficient (ICC) were calculated using IBM Statistical Package for the Social Sciences (SPSS) v. 20 statistics software. The following grading system for these coefficients was used: 0–0.3 values indicated lack of agreement, 0.31–0.5 – weak agreement, 0.51–0.7 – moderate agreement, 0.71–0.9 – very good agreement and 0.91–1 – excellent agreement. A p-value less than 0.05 was considered statistically significant.

Ethics statements

This study was carried out according to the principles of the Declaration of Helsinki of good clinical practice. The use of material for this study was approved by the Research Ethics Committee of the “Victor Babeș” University of Medicine and Pharmacy, Timișoara. Informed consent was obtained from each patient for using the resected specimens for scientific studies.

Results

We identified 21 cases that fit our inclusion criteria described above. The patients were between 42 and 85 years old (average age: 63.4 years), 57% (12/21) being men. One CRC case was right sided, the rest were in the left colon (3/21) and the rectum (17/21).

Histologically, all the cases were non-mucinous adenocarcinomas. The cases were staged according to the WHO 2010 – TNM system, as follows: six (28%) cases stage IA, eight (38%) cases stage IIA, six (29%) cases stage IIIB and one case stage IIIC. In respect to the tumor extension, the following distribution was identified: two (9%) cases pT1, five (24%) cases pT2 and 14 (67%) cases pT3. Lymph node metastasis evaluation classified 13 (62%) cases as N0, seven (33%) cases as N1 and one case (5%) as N2.

We graded the CRC tumors according to the WHO criteria and noticed that 2/21 were G1, 17/21 (81%) were G2 and 2/21 (9.5%) were G3. When we applied the above-mentioned two-tier grading system, 19/21 (90%) tumors were classified as low grade (Figure 1).

The analysis of the tumor invasion pattern showed that 1/21 (5%) cases was pushing type, 12/21 (57%) cases were infiltrative type and 8/21 (38%) cases of mixed type.

All the 21 CRC cases presented TB, with an average value between 0.4 and 55.3 buds in 10 microscopic fields on HE-stained slides, while on the IHC-stained slides the values were between 1.1 and 66.6 (Table 1).
between 0.006 and 0.0001). IHC staining improved the interobserver reliability (values between 0.125 and 0.384) on the HE-stained slides. The reclassification occurred in 17/21 (80%) cases for R1, as follows: in 19/21 cases for R1 (average increase of 234.85%), 16/21 cases for R2 (average increase of 114.14%), and 20/21 cases for R3 (average increase of 66.92%).

Interestingly, for the R1 and R2 junior pathologists, in two, respectively five cases, the average number of TB counted on IHC-stained slides was lower than the one determined on the HE-stained slides. For R3, the senior pathologist, all the bud counts were higher on IHC-stained slides in comparison to the corresponding HE-stained slides.

Based on the number of TB counted on HE-stained slides, the following classification was obtained: 13/3/10 cases as G1Bd, as reported by R1, R2 and R3, 3/9/5 cases as G2Bd and 5/9/6 as G3Bd. On the IHC-stained slides evaluation, the following classification was noted: 3/2/6 as G1Bd, 3/3/4 as G2Bd and 15/16/11 as G3Bd, as reported by R1, R2 and R3.

We used the previously obtained average values to compute the ICC on both HE and IHC bud counts (Table 2). In the case of HE-stained slides, the interitem correlation matrix showed values between 0.632 and 0.84, while the ICC on average measures for consistency showed very good correlation [ICC: 0.887, 95% confidence interval (CI): 0.765–0.95]). The IHC values evaluation showed better results: the interitem correlation matrix comprising values between 0.864 and 0.921, while the ICC for average measures for consistency yielded an excellent value (ICC: 0.95, 95% CI: 0.896–0.978).

### Table 1 – The average values for TB counts/case/staining for each of the reviewers (R1, R2, R3)

<table>
<thead>
<tr>
<th>No.</th>
<th>HE</th>
<th>IHC</th>
<th>HE</th>
<th>IHC</th>
<th>HE</th>
<th>IHC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.25</td>
<td>17.5</td>
<td>4.5</td>
<td>12.5</td>
<td>2.2</td>
<td>3.8</td>
</tr>
<tr>
<td>2</td>
<td>3.71</td>
<td>11.4</td>
<td>6.9</td>
<td>14.33</td>
<td>4.33</td>
<td>3.67</td>
</tr>
<tr>
<td>3</td>
<td>2.67</td>
<td>20</td>
<td>14.9</td>
<td>21.7</td>
<td>8.4</td>
<td>21.4</td>
</tr>
<tr>
<td>4</td>
<td>5.1</td>
<td>31.7</td>
<td>13</td>
<td>39.3</td>
<td>13.8</td>
<td>25.1</td>
</tr>
<tr>
<td>5</td>
<td>2.63</td>
<td>7.5</td>
<td>8.6</td>
<td>14.5</td>
<td>5.4</td>
<td>6.7</td>
</tr>
<tr>
<td>6</td>
<td>8.67</td>
<td>26.3</td>
<td>6.9</td>
<td>22.2</td>
<td>18.6</td>
<td>21.3</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>23.7</td>
<td>6.7</td>
<td>19.9</td>
<td>6.8</td>
<td>14.7</td>
</tr>
<tr>
<td>8</td>
<td>8.8</td>
<td>14.5</td>
<td>6.6</td>
<td>17.8</td>
<td>10.6</td>
<td>13.3</td>
</tr>
<tr>
<td>9</td>
<td>3.9</td>
<td>10.22</td>
<td>7.1</td>
<td>12.6</td>
<td>3.6</td>
<td>5.8</td>
</tr>
<tr>
<td>10</td>
<td>2.9</td>
<td>7.67</td>
<td>8.8</td>
<td>7.6</td>
<td>4.8</td>
<td>7.6</td>
</tr>
<tr>
<td>11</td>
<td>14.5</td>
<td>25</td>
<td>10.8</td>
<td>29.6</td>
<td>9.8</td>
<td>14</td>
</tr>
<tr>
<td>12</td>
<td>4.3</td>
<td>14.4</td>
<td>7.4</td>
<td>11.1</td>
<td>8.8</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>11.6</td>
<td>16.3</td>
<td>12.3</td>
<td>17.6</td>
<td>3.8</td>
<td>13.9</td>
</tr>
<tr>
<td>14</td>
<td>11</td>
<td>6.2</td>
<td>9.9</td>
<td>26.1</td>
<td>6.8</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>1</td>
<td>5.3</td>
<td>6.75</td>
<td>0.4</td>
<td>0.8</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>2.25</td>
<td>3.2</td>
<td>1.6</td>
<td>0.8</td>
<td>1.1</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>24.33</td>
<td>10.2</td>
<td>6</td>
<td>4.6</td>
<td>5.9</td>
</tr>
<tr>
<td>18</td>
<td>2.17</td>
<td>2.33</td>
<td>11.4</td>
<td>3.71</td>
<td>1</td>
<td>2.6</td>
</tr>
<tr>
<td>19</td>
<td>8.8</td>
<td>14.7</td>
<td>9.2</td>
<td>12.7</td>
<td>10.4</td>
<td>13</td>
</tr>
<tr>
<td>20</td>
<td>9.5</td>
<td>16.1</td>
<td>33.4</td>
<td>23.9</td>
<td>6.4</td>
<td>13.1</td>
</tr>
<tr>
<td>21</td>
<td>55.3</td>
<td>59.5</td>
<td>41.1</td>
<td>66.6</td>
<td>32.2</td>
<td>42.6</td>
</tr>
</tbody>
</table>

TB: Tumor budding; HE: Hematoxylin–Eosin; IHC: Immunohistochemistry.

We identified higher values for budding counts on IHC-stained slides, in comparison to the HE-stained slides.

### Table 2 – Interitem correlation matrix and Cohen’s kappa reflecting the differences in TB assessment between the three observers

<table>
<thead>
<tr>
<th>Interitem correlation matrix</th>
<th>HE three-tier</th>
<th>IHC three-tier</th>
<th>HE two-tier</th>
<th>IHC two-tier</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE IHC</td>
<td>Kappa p</td>
<td>Kappa p</td>
<td>Kappa p</td>
<td>Kappa p</td>
</tr>
<tr>
<td>R1 vs. R2</td>
<td>0.785</td>
<td>0.864</td>
<td>0.364</td>
<td>0.001</td>
</tr>
<tr>
<td>R1 vs. R3</td>
<td>0.84</td>
<td>0.918</td>
<td>0.211</td>
<td>0.174</td>
</tr>
<tr>
<td>R2 vs. R3</td>
<td>0.632</td>
<td>0.921</td>
<td>0.136</td>
<td>0.34</td>
</tr>
</tbody>
</table>

TB: Tumor budding; HE: Hematoxylin–Eosin; IHC: Immunohistochemistry.

#### Tumor grading correlation evaluation using Cohen’s kappa (interobserver reliability test)

We firstly used the current recommended three-tier grading system for TB to classify each case. There was poor correlation between the observers (kappa values between 0.125 and 0.384) on the HE-stained slides. The IHC staining improved the interobserver reliability (kappa between 0.39 and 0.548, with statistical significance between 0.006 and <0.0001).

For the two-tier system, the HE interobserver reliability test showed minor improvement (kappa values between 0.14 and 0.512). The IHC staining evaluation showed similar changes (kappa values between 0.31 and 0.618, with statistical significance p-values between 0.05 and 0.005).

**Buds grade vs. WHO grade**

We wanted to see if there are changes in the way the tumors would be classified according to the new GBd. As shown in Table 3, for the three-tier systems, the reclassification occurred in 17/21 (80%) cases for R1, 12/21 (57.1%) cases for R2 and 15/21 (71.4%) cases for R3. On the two-tier classification systems, we identified a class concordance in 14/21 (66.6%) cases for R1, 11/21 (52.4%) for R2 and 15/21 (71.4%) for R3.

We analyzed the difference in classification when considering the maximum bud count value, in the hotspot, instead of the average value in the 10 microscopic fields. The results showed an increase in the grade on HE-stained slides as follows: in 11/21 (52.4%) cases for R1, 11/21 (52.4%) cases for R2 and 9/21 (42.8%) cases for R3. On IHC slides, the total number of class changes was lower: 3/21 (14.3%) cases for R1, 4/21 (19%) cases for R2 and 8/21 (38%) cases for R3.

### Table 3 – Correlation between classical G and GBd

<table>
<thead>
<tr>
<th>GBd&lt; G</th>
<th>GBd= G</th>
<th>GBd&gt; G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-tier GS</td>
<td>Two-tier GS</td>
<td>Three-tier GS</td>
</tr>
<tr>
<td>R1</td>
<td>R2</td>
<td>R3</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

GS: Grading system; G: Histological grade; Bd: Tumor budding.
Discussions

The tumor stage, evaluated according to the TNM system, remains, even today, in 2018, the base parameter for stratifying CRC patients in prognostic groups that benefit from adapted therapeutic measures [3]. Alas, according to some authors, for some tumor stages and for a certain patient, the TNM stage is not enough to predict the prognosis [4, 40, 41].

Together with the stage, the tumor grade is considered an important prognostic factor in most of the malignant tumors. In CRC, there is a fairly large interobserver variability in grading these malignancies, as opposed to other sites, like the renal clear cell carcinomas, where the grade, according to the WHO system [42], is established on more objective criteria. This variability is largely attributed to the imprecise and subjective character of the evaluation and to the intratumor heterogeneity [43]. Furthermore, there are tumor subtypes of CRC, like the mucinous, medullary and micropapillary carcinomas, for which the tumor grade has an arguable clinical utility, due to the limited prognostic value [44]. Hence, the prognostic value of tumor grading in CRC, assessed according to the WHO, may become questionable.

Although the analysis of certain molecular factors seems useful from the perspective of the prognostic and predictive value for therapy response in CRC patients [45], the histopathological and IHC investigations, which by far are easier and cheaper to perform, readily available in most pathology labs, could offer solutions for a more accurate classification. In this sense, the aspect of the tumor invasion front (infiltrative or pushing) [46, 47], TB [15–18, 48], PDC [11, 13, 14, 49], the immune peri- and intra-tumoral response [50] represent intensely studied parameters, showing promising results. From all the previously mentioned, the TB and PDC stand out as potential aggressiveness markers, which could complete or replace some well-known prognostic factors in CRC [12, 14, 38, 51].

TBs are classified as peritumoral TBs, which can be evaluated in endoscopic and surgical resection specimens, and intratumoral TBs, identifiable in endobiopsy specimens, which usually do not contain the invasion front [52, 53].

TBs are formed because of the detachment of isolated or small groups of tumor cells, at the invasion front, in their attempt to evade the local defense mechanisms, represented by inflammatory cells, and ultimately result in blood and lymphatic vessel invasion [17]. TB would represent the histomorphological expression of epithelial-to-mesenchymal transition (EMT), in which the tumor cells gain or increase their aggressiveness by acquiring a mesenchymal phenotype, thus fostering a higher capacity for migration and invasion, reduced proliferative activity and resistance to apoptosis [54, 55].

The majority of scientific studies report a negative prognostic impact for the presence and high numbers of TB, as it correlates with perineural and lympho-vascular invasion, lymph node and distant metastasis, along with the infiltrative aspect of the invasion front [15, 18, 24, 30, 56–63], even giving this parameter the value of an independent prognostic marker in CRC N0 [24, 59]. The 2012 guides of the European Society of Medical Oncology (ESMO) recognize the TB as a potential prognostic factor for early CRC [64]. TB would act as a good predictor for the presence of isolated tumor cells in pN0 patients [65, 66]. The variance in the prognostic value attributed by different studies may be explained by the heterogeneity of the study groups: consecutive vs. selected cases; colon vs. rectal vs. colorectal cancers; the number of cases evaluated – tens vs. hundreds; different cancer stages – early vs. advanced vs. all stages; presence or absence of distant metastasis; different tumor subtypes, etc.

TBs are important also from the perspective of the therapy in the following circumstances: (i) in polyps with malignant transformation (pT1), endoscopically resected, the lymph node metastases are identified in up to 15% of the cases [39], thus TB could be used as an independent predictor for its presence [38], allowing the selection of specific cases for surgical resection and avoiding excessive treatment in 85% of the cases; (ii) in stage II CRC, a tumor group that presents a highly variable evolution, the presence of TB is associated with a high risk of relapse and distant metastasis [16, 41, 59, 67]. TB proved to be an independent predictor for patient survival in this group [56], justifying, according to some authors, the need for adjuvant therapy instatement [16], while other agree that further larger studies are required to establish this necessity [41]; (iii) in the diagnostic preoperative biopsies of CRC, the presence of TB correlates with a larger extent of the tumor, with lympho-vascular invasion detectable on the resection specimen, with lymph node and distant metastasis [41], thus prompting the need for neoadjuvant therapy [16, 30, 38, 68]; (iv) TB tend to correlate with the response to anti-epidermal growth factor receptor (EGFR) therapies in patients with metastatic CRC, in a similar fashion as the status of the K-RAS gene mutation [55].

Despite the importance attributed to TB in respect to the prognosis and therapy, in the Western countries it is not mandatory to report this parameter, due to the poor reproducibility of the evaluation as there is no standardization in the definition, assessment and reporting of TB [41, 69]. In the attempt of standardization and generalization of this parameter, the ITBCC, which took place in 2016 in Bern, was tasked with gaining a consensus regarding an evidence-based standardized scoring system for TB that can be used in international guidelines for CRC and daily practice [38].

In our study, we defined TB as isolated or a cell clusters of up to four tumor cells that do not represent areas of glandular disruption. We chose the upper limit of four cells for TB because ITBCC recommends it, on one hand, and to avoid the confusion for overlap of the limit value of five cells between TB and PDC, on the other hand. PDCs are defined as groups/solid nests of more than five tumor cells and represent another promising parameter in the evaluation of CRC [70].

We identified in the literature that there is a somewhat unclear definition of TB in respect to the number of cells: four tumor cells or less, up to four tumor cells [38], less/fewer than five cells [71], up to five cells [23, 72], ≤5 cells [67], 1–5 tumor cells [16, 31, 52], aspect that seems to have no important consequence, since Caie et al. have shown that TB are associated with unfavorable prognosis, regardless of the size of TB (1–2 cells, 1–5...
cells and >5 cells) [31]. Nevertheless, PDCs are similar to TB, correlate with TB, being considered a sequential step in tumor progression [14], and for some time they were quantified by some authors as TB [52]. Therefore, the comparison of the results of different studies dedicated to TB and PDC is difficult.

Our results show the presence of TB in all of the 21 analyzed cases, in contrast to the literature data which shows that TB are present in very variable percentage, ranging from 20% [41] up to 78% of the cases of CRC [18], depending on the evaluated specimen: HE vs. IHC (38% vs. 63%) [62], method of evaluation: the whole invasion front vs. hotspot – the microscopic field with the highest TB count (42% vs. 38%) [4], the molecular subtype: microsatellite stable CRC vs. microsatellite unstable CRC (50% vs. 25%) [41]. This variability could be dependent on the study group, ours comprising robotic surgery specimens, and not consecutive or a particular stage of CRC. Furthermore, in the definition for TB positive cases, there is a discrepancy in how they are reported as having/not having TB. Morodomi et al. [52] consider that the cases presenting 0–4 TB are TB-negative, while others consider the presence of any TB as a TB-positive CRC [73]. On the other hand, the high number of TB-positive specimens that we identified correlates with a high number of cases (95%) presenting infiltrative and mixed type pattern of invasion, in concordance with the previously described association of high TB count and other unfavorable prognostic factors like the infiltrative type invasion front [19, 20].

Considering the heterogeneity of CRC, both in the degree of differentiation and in the histological subtype, it is expected that the TB count would vary inside the same tumor, therefore the correct selection of the tumor area to be evaluated is extremely important. There is a high variability in different studies, regarding the area in which to perform the TB evaluation, the method of counting and the reporting of this parameter. Unlike the majority of authors that evaluate only one section per tumor [15, 18, 24, 30, 56–60, 62, 63, 65] a rather limited number of researchers evaluated the entire tumor [15, 59]. In most of the studies, TBs are evaluated at the invasion front [15, 18, 24, 30, 56–60, 62, 63, 65], while in others TB are evaluated also inside the tumor [52, 53, 68]. The quantification of TB was performed differently over the years: Hase et al. [15] proposed a subjective evaluation at 5× objective, with consecutive allocation as none, mild, moderate and severe and further grouping into two classes: none or minimal vs. moderate or severe; Nakamura et al. [56] proposed a method that evaluates the proportion of the invasion front presenting TB, with the following classification: none, mild: <1/3, moderate: 1/3–2/3, marked: >2/3; Ueno et al. [72] proposed a quantitative method of evaluation by assessing the number of TB in one microscopic field, that presented the highest number of TB (hotspot, 250× magnification). Karamitopoulou et al. [57] and Horcic et al. [74] recommend the evaluation of TB in 10 microscopic fields at 400× magnification, while Caie et al. [31] proposes 15 fields at 200×. The last ones have the advantage of taking into consideration the intratumoral heterogeneity, reducing the interobserver variability and employing a method already known and used in quantification, e.g., mitosis in other tumor types. Furthermore, the invasion front in malignant polyps is a lot smaller, and the quantification using 10 or 15 microscopic fields is not feasible for this tumor category [4]. Contrary to this approach, there is the concern that the evaluation of TB in 10–15 microscopic fields would lead to an underestimate of the TB count, not taking into consideration the field with the highest number of TB [38]. ITBCC recommends the following method for TB assessment: select the field with the highest TB density on a 10× objective on an HE-stained slide, then count the TB on a 0.785 mm² circular area, corresponding to a 20× objective lens with a 20 mm eyepiece field number diameter. Considering the variability in the size of the field due to the optics of individual microscopes, a conversion table was developed to adjust the number of TB to the area of 0.785 mm² [38]. Similarly, the Japanese Society for Cancer of the Colon and Rectum recommends the evaluation of a single microscopic field at the invasion front, containing the highest TB count (hotspot method), at 200× magnification, on HE-stained slides [71].

We chose to quantify TB in 10 circular areas of 0.785 mm², considering that the 200×250× magnification is the most widely used one for determining TB in CRC [18, 39, 58, 63, 65, 72]. As our evaluators did not have any previous experience in TB evaluation and implicitly identifying the hotspot, we considered to evaluate 10 microscopic fields on both HE- and IHC-stained slides.

All of our evaluators identified a higher number of TB on IHC-stained slides, when compared to HE, similarly to previously published studies [4, 55, 56, 63, 67, 74–78], emphasizing the utility of CK staining for accurate evaluation of TB. The cells that build up the TB seem less differentiated than the ones in the tumor mass, as the electron microscopy studies show, a process known as tumor dedifferentiation [79] and, unlike PDC, when they are isolated of constitute small TB, they are difficult to identify on HE-stained slides, as they can be mistaken for histiocytes belonging to the inflammatory cell infiltrate or for reactive stromal cells [67]. Sometimes they can be masked by the inflammatory cells [57, 72, 80]. IHC methods identify even the tumor cells which appear consecutive to the glandular disruption by the inflammatory infiltrate – these should not be counted as TB [23], or the apoptotic bodies, cellular debris and cytoplasm fragments of tumor cells, the latter as an expression of the tumor cell motility, explaining, among others, the high number of TB counts on the CK-stained slides [48]. We think it is possible that the higher count of TB on HE-stained than on the IHC-stained slides determined by our junior pathologists in two and five cases, respectively, is explained by the erroneous counting of histiocytes or cells dissociated from glandular structures due to inflammation. This aspect would constitute an argument for using IHC-stained slides by the pathologist with limited experience when they are involved in TB counting.

We are witnessing a variability in reporting TB counts, dependent on the clone of the antibody that is used (CK AE1/AE3, MNF116, Cam5.2), the method of quantification (hotspot vs. 10 fields), and the magnification employed (×200, ×400).
There is controversy surrounding the role of IHC in evaluating TB. On one hand, considering that TB is a prognostic marker without diagnostic value, that TB counts on HE-stained and IHC-stained slides have similar prognostic value [16, 24] and that IHC staining for every CRC case implies increased running costs and the need for proper lab equipment, there is no justification for routine CK staining for TB evaluation of all the cases [18, 24], only for the selected ones. On the other hand, if TB is to be implemented as a parameter that will contribute or dictate therapeutic decisions of surgical resection, adjuvant or neoadjuvant therapy, the costs of IHC staining are negligible in comparison to the costs of the therapy. The ITBCC recommendation, voted favorable by 86% of the participants, is a reasonable approach, as it states that TB counts should be made on HE-stained slides, and that IHC should be used only in difficult cases, e.g., with marked inflammation, glandular fragmentation, etc. [38].

Analyzing the interobserver variability in TB quantification on HE-stained slides, we noticed a moderate to very good correlation, which is in accord with other studies [81]. Wang et al. show a very good interobserver correlation between experienced pathologists and consider that it can increase through training [59]. These results are different from what Puppa et al. reported, namely a low interobserver agreement for TB evaluation on HE-stained slides, but better for incipient tumors and in a relationship with the expertise of the pathologist [4].

In our hands, the interobserver agreement for IHC TB quantification showed excellent values, higher than on the HE-stained, and similar to the findings of Karamitopoulou et al. [57]. Based on the literature findings, De Smedt et al. [41] recommend evaluating TB in 10 microscopic fields on IHC stained slides, as this method has the best interobserver agreement, recommendation that our study confirms. On a different perspective, Puppa et al. [4] notice a similar inter- and intraobserver variability for HE- and IHC-stained slides, and surprisingly, a higher agreement on HE-stained slides for experienced pathologists.

Interobserver agreement is certainly better when TB are evaluated as a numeric count – quantitative [17, 74, 82] vs. qualitative assessment [74].

The analysis of tumor classification according to the classical degree of differentiation – G, with three grades (G1, G2, G3) and two grades (low vs. high grade), and the grading based on TB counts, on the three-tier system (G1Bd, G2Bd and G3Bd) and two-tier system (low GBd and high GBd), showed that the two grading systems (G and GBd) correlate better when using two degrees than three.

The mismatched allocation in the GBd categories in comparison to the classical grading system, for all our reviewers, supports the idea that TB grading is different than the well-established WHO grade, as it is also shown in the ITBCC recommendations [38]. This difference in grading emphasizes the idea that TB might be able to explain the variability in the evolution of the tumors from the same stage/grade.

Furthermore, the comparison of different studies regarding the tumor grading according to GBd is extremely difficult, as there is no standardization in the determination, evaluation and reporting. There is no generally accepted cut-off that separates high GBd tumors from low grade ones. Märkl et al. proposed the cut-off of ≥30 buds/in a 20× objective field (1.3 mm²) for classifying as high GBd tumors [5], while Sy et al. proposed nine TB in a 20× objective field as the limit for separating the low GBd from high GBd [83]. Even for the cut-off that we used, there are slightly different approaches: some advocating for >10 TB [24, 58, 65], while other say ≥10 TB [18, 19, 72] in a 20× objective field. In the semi-quantitative evaluations, high GBd grade is set for tumors showing TB positivity in more than 50% of the examined area [59].

In our hands, the interobserver reliability improved when the TB were counted on the IHC-stained slides, when compared to the HE-stained one, similarly to the previous reports [57, 74]. There was also better agreement between the observers when classification was performed on the two-tier system vs. three-tier system, upholding the idea that interobserver concordance can be improved by decreasing the number of grades in classifications [72]. All things considered, the ITBCC recommendation for case stratification in respect to the TB is to be performed on a three-tier system: 0–4 – low budding (Bd1); 5–9 buds – intermediate budding (Bd2); and ≥10 buds – high budding (Bd3) [38], system that is currently used by Japanese Society for Cancer of the Colon and Rectum [71]. We used a similar system, which differed in the threshold for the low budding (0–5) and intermediate budding (6–9).

It is possible that the very good results that we noted for the interobserver agreement are related to the fact that we used WSI for quantification and not the classical optical microscope, outcome that was also identified by Lugli et al. [17], who showed better concordance rates when using IHC-stained slides and digital pathology evaluation of these, for both hotspot and 10 fields methods, and proposed the evaluation of TB to be performed on IHC-stained slides, on 10 microscopic fields for resection specimens and hotspot method for smaller specimens.

Scanned slides are used more often than ever for quantifying some histopathological factors with prognostic and predictive value [84] and for establishing interobserver agreement [85], providing several advantages: it assures a standardized measure for the area of tissue to be examined, eliminating discrepancies that can arise in usual microscopes related to different optics properties; it decreases the possibility of having overlapping quantification or missed areas and, even better, allows the setting of fixed areas to be used for quantification by several observers, improving consistency in the study method; not less important is that it might be more ergonomic than using a regular microscope, as one will be examining a digital screen in a more relaxed position; it improves the reliability of counting in fields by adding grids on top of the image; it allows the evaluation of the entire lesion on several slides and annotate them [30]; last but not least important, one can zoom in the image without losing the size and location of the field to be examined [4]. We identified several of these advantages in our study and we consider the most important one was the ability to draw 10 fixed circular areas of 0.785 mm² each.
The quantification methods based on computer-aided image analysis gain more ground as the technology progresses over time, offering a series of advantages for evaluation of some histological parameters, among them the TB [32]. Caie et al. [31] managed to demonstrate the possibility of semiautomated evaluation of TB along with two more factors: lympho-vascular density and lympho-vascular invasion, employing computer-driven image analysis of fluorescently stained slides for pan-CK and D2-40, a lymphatic endothelial maker. This method significantly reduced the interobserver variability, but it has disadvantages: the need for a fluorescence microscope along with proper image analysis software, which are not readily available in all laboratories and can be quite expensive, and for special training of the examiner, which will be subject to increased visual effort.

We consider our method to be useful as a first step in standardizing TB assessment in pathology labs, where the slides can be scanned or photographed digitally. This 10 fields’ method of 0.785 mm² surface area provides a good interobserver agreement and facilitates teaching and training of pathologists in TB quantification.

The limits of our study are related to the reduced number of cases included and the lack of follow-up data on the patients. We aimed it to be a pilot study, a first step in TB quantification in CRC, for setting up a method agreeable to all the pathologists involved in TB assessment. Following this experience, we plan to implement this method in the pathology service of the “Pius Brînzeu” Emergency County Hospital, Timișoara, Romania, as the first edition of the European Guidelines for Quality Assurance in CRC Screening and Diagnosis, recommend mentioning TB in the pathology report, adding an informative comment that TB is suggested to be a prognostic factor for CRC [69].

Conclusions

TB represents a parameter that needs to be known and the pathologist must be familiar with the methods of identification, quantification and reporting of TB, as the standardization of this assessment will lead to the implementation of this prognostic and predictive factor in current diagnosis of CRC. We consider the method of buds counting in 10 microscopic fields to be reliable and valuable. TB counts are higher on IHC slides and associate a better interobserver agreement. The evaluation of scanned slides is labor intensive, time consuming, but it might be able to increase the accuracy of the results and to reduce the interobserver variability.

Conflict of interests

The authors declare that they have no conflict of interests.

Acknowledgments

This paper is supported by the Project “Multicenter study on the use of a robotic approach in order to correlate a decrease of circulating adipokine level in obesity with the risk of pelvic cancer, “ROBOCAPE-”” POSCCE – ID 1846/Code SMIS 48478 – No. 669/01.09.2014.

Authors’ contribution

A.L.C. Dema and S. Tában elaborated the study design, coordinated its development, contributed to literature review, the analysis and data interpretation, and writing the paper. S. Tában performed TB counting on HE- and IHC-stained slides. O.F. Lazăr and C.C. Duță performed the robotic surgeries, coordinated the clinical data collection and finally approved the manuscript. A.R. Gheju and A. Jurescu selected the cases, the paraffin blocks, coordinated the IHC staining and counted TB on HE- and IHC-stained slides. A.O. Văduva scanned the slides, performed the statistical analysis of the results and contributed to writing of the paper.

References

Paving the way to tumor budding assessment using digital pathology: a pilot study in Timisoara City (Romania)


Horcic M, Koelzer VH, Karamitopoulou E, Terracciano L, Puppa G, Zlobec I, Lugli A. 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(3):697–705.


Paving the way to tumor budding assessment using digital pathology: a pilot study in Timișoara City (Romania)


Corresponding author
Adrian Ovidiu Vaduva, Assistant Professor, MD, PhD, Department of Microscopic Morphology – Pathology, "Victor Babes" University of Medicine and Pharmacy, 2 Eftimie Murgu Square, 300041 Timișoara, Romania; Phone/Fax +40256–204 476, int. 1446, e-mail: vaduva.adrian@umft.ro

Received: July 3, 2018
Accepted: November 28, 2018