Improved objectivity of grading of TA\textsubscript{1} transitional cell carcinomas of the urinary bladder by quantitative nuclear and proliferation related features

M G W Bol, J P A Baak, P C de Bruin, S Rep, W Marx, S Bos, O Kisman

Abstract

Aim—To analyse whether the mean nuclear area of the 10 largest nuclei (MNA-10), the mitotic activity index (MAI), and Ki-67 immunoquantitative features have additional value to discriminate different grades of TA\textsubscript{1} transitional cell carcinoma (TCC) of the urinary bladder.

Materials/Methods—One hundred and fifty of 200 consecutive cases (75%) showing interobserver agreement on duplicate grade assessment by independent pathologists were studied. Using random numbers, the 150 cases were divided into sets for learning (n = 75) and testing (n = 75). Single and multivariate analyses were applied to discriminate the different grades in the learning set. The multivariate classifier developed in this way was evaluated in the test set (n = 75).

Results—With the MNA-10 alone, using the classification MNA-10 < 80 \mu m^2 = grade 1, 80 \mu m^2 < MNA-10 < 130 \mu m^2 = grade 2, MNA-10 > 130 \mu m^2 = grade 3, 71% of all 150 cases were correctly classified (69% of grade 1 v grade 2 and 76% of grade 2 v grade 3). With multivariate analysis, the best discriminating features in the learning set (17 grade 1, 30 grade 2, and 28 grade 3) between grades 1 and 2 were MNA-10 and MAI, and between grades 2 and 3 MAI and Ki-67. With these features, 94% of grade 1 v grade 2 and 97% of grade 2 v grade 3 were correctly classified in the learning set (overall, 95% correct, none of the grade 3 cases misclassified). In the test set the classification results were similar. When the three grades were entered at the same time for discrimination, Ki-67 area % and MAI was the best discriminating combination, both in the sets for learning and testing. Overall correct classification results in the sets for learning and testing were slightly lower, but still 94% and 92%. Most importantly, none of the grade 3 cases was misclassified; the classification shifts all occurred between grades 1 and 2.

Conclusions—The combination of MNA-10, MAI, and Ki-67 gives much better discrimination between grades 1, 2, and 3 in TA\textsubscript{1} TCC of the urinary bladder than MNA-10 alone. The similarity of the classification results of the learning set and test set are encouraging and this quantitative pathological grading model should be applied in a prospective study.

Keywords: urinary bladder; bladder tumours; transitional cell carcinoma; Ki-67; morphometry; proliferation

The incidence of transitional cell carcinoma (TCC) of the urinary bladder is high. There are two main forms of this cancer, superficial and invasive,\textsuperscript{1,2} and the superficial form (TA\textsubscript{1}) is the most frequent in the Western world. About 70% of TA\textsubscript{1} tumours recur after transurethral resection, both on the original site and new sites in the urinary bladder, and at recurrence approximately 20% show progression to a higher stage.\textsuperscript{3} In TA\textsubscript{1} tumours, grade is an important criterion for the prediction of outcome and also for therapeutic decision making. In many centres, patients with grade 1 and 2 TA\textsubscript{1} tumours are monitored on a regular basis (every three months for the first year), but grade 3 cancers are monitored even more intensively, and often also receive adjuvant local treatment. However, interobserver reproducibility of grading may be as low as 60–85%.\textsuperscript{4} Therefore, the need for objective grading criteria is high. Many groups have studied markers with the aim of improving the reproducibility of grading and the accuracy of the prediction of outcome. DNA ploidy is correlated with grade,\textsuperscript{5} but the assessment requires single cells to be prepared from the solid paraffin wax blocks. Mean nuclear area is easier to assess because standard histological sections can be used. The mean profile area of the 10 largest nuclei (MNA-10) found in a histological tumour section\textsuperscript{6} has proved to be valuable as an independent marker for grade and prognosis.\textsuperscript{7} In the study of Blomjous et al,\textsuperscript{9} DNA diploid or DNA aneuploid cases (measured by flow cytometry) had nearly the same survival as MNA-10 low (< 95 \mu m^2) and MNA-10 high (> 95 \mu m^2) TA\textsubscript{1} tumours, respectively. Thus, it is tempting to speculate that high MNA-10 values are strongly predictive of DNA aneuploidy, which may be understandable from a cell biological point of view—many studies have shown that nuclear area strongly correlates with DNA content; thus, aneuploid cells often have larger nuclei. Indeed, using image cytometry highly aneuploid nuclei often have very large nuclei.

However, proliferation is also important in grading,\textsuperscript{8} and this is not reflected in the MNA-10 value. In many cancers, Ki-67\textsuperscript{9–18} and the metaphase marker mitotic activity index (MAI)\textsuperscript{19–22} are widely accepted as proliferation associated markers. Although Pich and colleagues\textsuperscript{23} have studied Ki-67 among other parameters (argyrophilic nucleolar organiser...
Urinary bladder tumour grade and Ki-67 quantitation

regions, proliferation cell nuclear antigen, PC10, mean nuclear area, DNA flow cytometry, to our knowledge the values of MNA-10, MAI, and Ki-67 have not been studied simultaneously in superficial TCC. Such a multiparameter study could be useful for several reasons. First, these features are reliable, relatively cheap, and easy to assess, which is of practical value for a clinical grading tool. Second, discrimination between different grades may increase when these markers are combined in a model. Third, reproducibility of grading may improve.

These arguments were the reason for our present study. We will investigate whether these features in combination have stronger discriminating value between TCCs of different grades than MNA-10 alone.

Materials and methods

PATIENTS

Two hundred consecutive T_A,1 tumours were obtained from the 1996 archive of the department of pathology of the Medical Centre Alkmaar (MCA). The tissue was obtained either by transurethral resection or by biopsy performed at the departments of urology at the MCA and the Gemini Hospital, Den Helder, the Netherlands. The mean age of the patients at the time of diagnosis was 67 years (range, 54–79). The tumour tissue was fixed in 4% buffered formaldehyde, dehydrated, and embedded in paraffin wax. Haematoxylin and eosin (H&E) stained 4 µm histological sections were used. Two pathologists evaluated all of the 200 consecutive TCCs according to the World Health Organisation criteria. Care was taken not to confuse tangentially cut TCC parts as solid areas. Consensus on grade was reached in 150 (75%) of the original 200 cases. These 150 cases were used for further study (75 primary and 75 recurrent tumours). The worst differentiated area (measurement area, minimally 2 × 2 mm) was carefully demarcated with a black marker for mitoses counts and MNA-10 measurements. The measurement area was selected so as to avoid necrotic, damaged, tangentially cut, inflamed, and benign or lower grade parts of the section. Care was taken not to confuse tangentially cut TCC parts as solid areas.

IMMUNOHISTOCHEMISTRY

Paraffin wax sections of 4 µm thickness, adjacent to the H&E sections used for the MAI and MNA-10 grade assessment, were mounted on to Super Frost Plus slides (Menzel-Gläser, Braunschweig, Germany) and dried overnight at 37°C. The sections were dewaxed in xylene and rehydrated in alcohol. After rehydration, the endogenous peroxidase activity was blocked by 3% H2O2 in phosphate buffer at pH 6.0. The sections were incubated with biotinylated swine antirabbit antibody (Dako; 1/600 in PBS) for 30 minutes, followed by the third incubation step with 1/100 streptavidin–biotin peroxidase complex (Dako) for 30 minutes. Visualization of the complex was with diaminobenzidine/H2O2 for 10 minutes at room temperature. There were two washes in PBS after each incubation step. The counterstaining was with Mayer’s haematoxylin. Finally, sections were dehydrated in graded ethanol and xylene and mounted with DPX (Nustain, Nottingham, UK). The section cut immediately after the immunostained section was stained with H&E and the presence and location of the T_A,1 lesion was carefully controlled (“sandwich-technique”) to guarantee that the Ki-67 section contained the same lesion.

QUANTITATIVE ANALYSIS

MAI was assessed by counting the mitotic figures in the previously demarcated measurement area of the H&E stained sections. This was done in 10 consecutive fields of vision at a final magnification of ×400 (objective, ×40; numerical aperture, 0.75). A field of vision was only accepted if a minimum of 75% consisted of urothelial tumour cells and stroma. Mitotic figures were counted using well established criteria. The manually demarcated measurement area in both the H&E and the Ki-67 stained sections was electronically demarcated for the measurements by an interactive video overlay system (QPRODFIT, Version 6.1; Leica, Cambridge, UK). The system consists of a personal computer with a mouse, a video overlay board, a colour camera mounted on a standard microscope equipped with an automated scanning stage, and measurement software. The live microscopic image is displayed in full colour on the monitor of the personal computer. The MNA-10 measurements were performed by scanning the demarcated measurement area manually for the subjectively largest nuclei using a total magnification of ×50. These nuclei were subsequently projected on the monitor of the personal computer using the largest magnification (total magnification, ×1000) and their profile area was measured using the mouse. To allow the measurement to be representative, 20 nuclei were measured. The MNA-10 consists of the mean of the 10 largest nuclei of the 20 nuclei thus selected and measured. The Ki-67 immunohistoquantitation was performed at a final magnification of ×400 (objective, ×40; numerical aperture, 0.75) by means of the two class immunoscoring module of the system, with an electronic test grid graphically overlayed on the microscopic image and counters. In each field of vision, one and the same of the four endpoints of the smallest grid available—two horizontal lines—is used for the point counting. A sample of 300 points at the start of each measurement was used so that at least 150 nuclei were counted in each measurement area (average, 220). At the magnification used,
urothelial tumour cells could accurately be distinguished from other cells. Using these criteria, the sample point of the test grid could be Ki-67 positive, Ki-67 negative, or neither of the two. These last points were ignored. The Ki-67 area % was defined as ((Ki-67 positive)/(Ki-67 positive + Ki-67 negative)) × 100.

STATISTICS

Statistical analysis was performed with the SPSS for Windows software package version 8.0 (SPSS, Gorinchem, the Netherlands). Descriptive statistics were calculated and box plots and scatter plots were made of the quantitative features for the total group, for the separate groups of primary tumours and recurrences, and for the grade 1, grade 2, and grade 3 TCCs (primary and recurrences). The differences of the mean values of MNA-10, MAI, Ki-67, and age between the primary tumours and recurrences were analysed for the total group and for the separate groups of grades (1, 2, and 3) with the Mann-Whitney test (because the hypothesis that several of the quantitative features with higher grade. These differences were all highly significant between the grades (table 1). However, discrimination between the three grades was not perfect because there is considerable overlap. Figures 2 and 3 show the box plots of the other quantitative features studied and grade. As expected there was some discrimination but overlap was still large.

Selection by random numbers of cases for the learning and test sets for multivariate analysis resulted in the same number of different grades in each set: grade 1, n = 17; grade 2, n = 30; grade 3, n = 28. In the learning set, the best discriminating combination of features between grades 1 and 2 was MNA-10 and MAI (fig 4). Note that there seems to be a dichotomy in the grade 2 cases: those that strongly resemble grade 1 cases and those that are at some distance from the grade 1 TCCs, with slightly higher values. The percentage of correctly classified cases in the test set with this combination was 94.7% (table 2). The best discriminating set of features between grades 2 and 3 was Ki-67 area % and MAI (fig 5). Here, the percentage of correctly classified cases in the test set was 97%. When the three grades were entered at the same time, Ki-67 area % and MAI was the best discriminating combination.

Table 1 Descriptive statistics and probability of no difference (Mann-Whitney test) of the 150 grade 1, 2, and 3 transitional cell carcinomas of the urinary bladder

<table>
<thead>
<tr>
<th>Variable</th>
<th>Grade 1 Mean (SD) P</th>
<th>Grade 2 Mean (SD) P</th>
<th>Grade 3 Mean (SD) P</th>
<th>Grade 1 Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of cases</td>
<td>34</td>
<td>60</td>
<td>56</td>
<td>34</td>
</tr>
<tr>
<td>MAI</td>
<td>1.5 (2.00)</td>
<td>0.0003</td>
<td>9.7 (6.51)</td>
<td>0.0003</td>
</tr>
<tr>
<td>MNA-10</td>
<td>82.0 (25.26)</td>
<td>&lt;0.0000</td>
<td>101.4 (21.83)</td>
<td>&lt;0.0000</td>
</tr>
<tr>
<td>Ki-67 area %</td>
<td>4.6 (3.39)</td>
<td>&lt;0.0000</td>
<td>10.3 (5.93)</td>
<td>&lt;0.0000</td>
</tr>
</tbody>
</table>

MAI, mitotic activity index; MNA-10, mean nuclear area of the 10 largest nuclei; P, probability of no difference.
both in the set for learning and for testing (fig 6). Overall correct classification results with the three grades at the same time were slightly lower than with the separate analysis of two adjacent grades, but were still 94% in the learning set and 92% in the test set. Most importantly, none of the grade 3 cases was misclassified; the classification shifts all occurred between grades 1 and 2, and one grade 2 case was misclassified as grade 3.

Discussion

The purpose of our study on transitional cell carcinomas of the urinary bladder was to develop a reproducible method for grading that is simple and robust and thus may be used in daily patient care. We confirmed earlier findings that the mean of the largest 10 nuclear profiles subjectively detected at low magnification in a tissue section of TCCs can be used to distinguish three grades of differentiation, but with considerable overlap. At the start of our study, the question was whether discrimination between different grades would increase when quantitative features were combined. Indeed, the discriminating power of MNA-10 in combination with Ki-67 and MAI was found to be stronger than that of MNA-10 alone. A second important question was whether a multivariate combination of the quantitative features studied could be used routinely as a clinical grading tool. In this context, it was important that the classification results in the sets for learning and testing were similar. This shows that the classifying combination of features was robust. Our results are encouraging and this multivariate grading model should be applied in a prospective routine grading study. Such a study is essential to answer another important question: does the application of this multivariate quantitative model in routine practice increase the reproducibility of grading? Although the lack of observer reproducibility of grading is not an argument for leaving out quantitative features in clinical pathology, such a formal prospective reproducibility study of the quantitative grading model developed in our study is important.

In a previous study on Barrett’s oesophagus we found that the quantitative assessment of Ki-67 positivity method used here was the most reproducible and fastest out of several others analysed.24 This was in agreement with the reports of Fleege and colleagues25 and Gundersen et al.26 These studies demonstrated that point weighted sampling, as used in our study, is the most efficient estimator available, and thus the best theoretical and practical approach to assess volume percentages. However, although the best one available, the
method still slightly overestimates objects that are larger. Ki-67 positivity mainly occurs in late G1, and in early, mid, and late S phase cells. These nuclei may be larger than the Ki-67 negative G0 and early G1 nuclei. Thus, the method used may result in a slight positive bias or overestimation of Ki-67 positive nuclei, especially of the ones that are in very late S phase, which in general may be considerably larger. On the other hand, the Ki-67 positive nuclei of G1 and early S phase cells (which are nearly of the same size as G0, G1 Ki-67 negative nuclei) probably form most of the Ki-67 positive nuclei in many tumours. Theoretically, this reduces the degree of bias for large nuclei that are Ki-67 positive. Thus, it is unlikely that the theoretical overestimation of Ki-67 positive nuclei that might occur would have influenced the results significantly (see also the wide variation that occurs in the Ki-67/MNA scatter plot).

Our study has re-confirmed the findings of Blomjous et al., putting the spotlight again on MNA-10. There are two important questions that need to be answered. First, why is a simple tumour characteristic such as MNA-10 such a strong grade discriminator? Second, is the overall pattern of nuclear size, expressed in the mean nuclear area, not equally strong or stronger in TCCs than MNA-10? If so, the occurrence of cells with very large nuclear profile areas in histological tumour sections could be an irrelevant epiphenomenon. Although the prognostic studies of Blomjous and colleagues suggest that this is not the case and that MNA-10 is more important than the mean nuclear area of the overall population, more detailed investigations to answer this question are needed. Confirmation of the importance of MNA-10 might indicate that high MNA-10 values point to a clone of cells with a particularly aggressive nature, rather than being an incidental side effect. This immediately raises questions as to the molecular biological background of MNA-10. Many older publications in the 1960s and 1970s showed that nuclear area is strongly correlated with DNA content. Thus, aneuploid cells often have large nuclei. Indeed, with image cytometry, hypertetraploid highly aneuploid cells often have very large nuclei. Thus, it is reasonable to hypothesise that high MNA-10 values are related to aneuploidy and perhaps even identify highly aneuploid nuclei. Moreover, with increasing grade MNA-10 also increases. This coincides with the frequent occurrence of aneuploidy in grade 3 tumours. Grade 3 TCCs have a greater likelihood than grade 1 and 2 TCCs of invading the lamina propria. Certain aneuploid stem cell lines are less stable genetically.

Regarding the proliferation associated features, the fact that MAI and Ki-67 are both higher in grade 3 tumours (although not very strongly correlated in many) may be the result of a correlation with p53 mutations. Indeed, defects in chromosome 17p53 have been preferentially found in high grade lamina propria invasive TCCs. This suggests a propensity of these lesions to progress and invade the muscularis. Chromosome 9 changes have been associated with non-invasive papillary changes. Knowledge of the various chromosomal changes that determine the biological behaviour of different forms of superficial bladder cancer is at an embryonic stage and clearly more detailed genetic studies of TCCs are important. On the other hand, proliferation is also strongly correlated with neovascularisation. It is thus tempting to analyse p53 and microvessel density in relation to MAI and Ki-67. This may give more biological insight into the relation between these proliferation features and the different grades.

Supported in part by grant number 98-119 of the SDBM.


