Survivin as a Useful Adjunct Marker for the Grading of Papillary Urothelial Carcinoma

Ying-bei Chen, MD, PhD; Jiangling J. Tu, MD, PhD; Jean Kao, BS; Xi K. Zhou, PhD; Yao-Tseng Chen, MD, PhD

Context.—Distinguishing low-grade and high-grade non-invasive papillary urothelial carcinoma based on morphologic criteria can be challenging and adjunct markers are highly desirable. Survivin, presumably an antiapoptotic protein, was previously proposed as a prognostic marker for urothelial carcinoma.

Objective.—To assess interobserver variability by 2004 World Health Organization classification and the value of survivin and Ki-67 as potential markers for grading non-invasive papillary urothelial carcinoma.

Design.—Fifty-one bladder biopsies were graded blindly by 5 experienced general surgical pathologists. The protein and messenger RNA expression of survivin and Ki-67 was evaluated by immunohistochemistry and quantitative reverse transcription–polymerase chain reaction using paraffin-embedded tissue. The immunohistochemistry result was quantitatively analyzed using a computer-based color deconvolution module.

Results.—The diagnostic agreement among 5 pathologists was fair to poor, with 32% of the cases graded differently by at least 2 raters. All cases were divided into 3 groups: consensus low-grade, consensus high-grade, and indeterminate. The percentage of urothelial cells with positive survivin nuclear staining (survivin score) was significantly higher in the high-grade than in the low-grade group (P<.001). Survivin score outperformed Ki-67 in separating the high-grade group from the low-grade group and showed a significantly higher predictive accuracy for high-grade recurrence than the histologic grade. The disagreement of grading for the indeterminate group could be resolved by their survivin scores in most cases. Survivin messenger RNA level correlated well with survivin score by immunohistochemistry but was not a more discriminating marker.

Conclusions.—Significant interobserver variability exists in grading low-grade versus high-grade papillary urothelial carcinoma. Survivin immunohistochemical staining can be a useful adjunct tool for the grading of challenging cases.

(Arch Pathol Lab Med. 2008;132:224–231)
Expression of survivin in this phase may also allow cells to overcome an apoptotic checkpoint, thus favoring aberrant progression through mitosis. Survivin has been shown to be highly expressed in many types of cancer cells including urothelial carcinoma, but essentially undetectable in nonproliferating normal adult tissues, making it an attractive cancer therapeutic target. Despite some inconsistent results obtained on the prognostic value of survivin protein or messenger RNA (mRNA) level among several early studies in urothelial carcinoma, recent studies using immunohistochemistry (IHC) or real-time quantitative reverse transcription–polymerase chain reaction (qRT-PCR) have suggested survivin expression as a promising prognostic marker that correlated with unfavorable disease-specific survival and recurrence. In addition, survivin has also been proposed as a urine marker for early bladder cancer detection or recurrence follow-up. No studies, however, have been conducted to evaluate survivin as a diagnostic marker in grading papillary urothelial carcinoma. On the other hand, the expression of a widely used nuclear proliferation marker, Ki-67, has been reported to be significantly different between LG and HG carcinomas diagnosed using the 2004 WHO criteria but was unable to distinguish between grade 2 and grade 3 carcinomas of the 1973 WHO classification.

In the present study, we evaluated the interobserver variability of grading LG and HG urothelial carcinoma using the 2004 WHO system. The protein and mRNA expression profiles of survivin and Ki-67 were then examined by IHC and qRT-PCR to assess their potential value in helping grade papillary urothelial carcinomas that were difficult to reach consensus solely by morphologic criteria.

MATERIALS AND METHODS

Case Selection and Histologic Grading

Archived materials from 51 bladder biopsies performed during 2005 with a diagnosis of either LG or HG noninvasive papillary urothelial carcinoma were retrieved from the surgical pathology files of New York Presbyterian Hospital-Weill Cornell Medical Center (New York, NY), following an institutional review board–approved protocol. Cases of PUNLMP were excluded from case selection. These 51 biopsies were from 46 patients and were obtained at the time of initial or follow-up diagnosis. Roughly equal numbers of LG and HG cases were selected based on the issued diagnoses, and biopsies of minute tissue sizes (<0.3 cm in greatest dimension) were excluded. The cases were blinded and the hematoxylin-eosin–stained slides were reviewed by 5 experienced general surgical pathologists (pathology board certified for 5–34 years; mean, 17.2 years) from the same institution. Although no case was originally diagnosed as PUNLMP, each case was given the option to be graded as PUNLMP or LG or HG carcinoma. The pathologists were also given the choice of submitting additional comments if desired.

Immunohistochemistry

Immunohistochemical staining for survivin (D-8, Santa Cruz Biotechnology, Santa Cruz, Calif) and Ki-67 (MIB-1, Dako, Carpinteria, Calif) was performed on formalin-fixed, paraffin-embedded 5-μm tissue sections following appropriate antigen retrieval on the BondMax Autostainer (Vision BioSystems Inc, Northwell, Mass) using the Bond Polymer Define Secondary Detection System and using diaminobenzidine as the chromogen (Vision BioSystems). The selected survivin antibody is a mouse monoclonal antibody raised against amino acids 1 to 142 of human survivin. The immunostaining of survivin and Ki-67 was quantified using a color deconvolution module of ImageJ (National Institutes
Correlation between protein and mRNA levels was assessed using Spearman rank correlation coefficient ($r_s$).

**RESULTS**

**Interobserver Variability on Histologic Grading**

The grading of 51 bladder biopsy cases by 5 pathologists showed significant interobserver variability. Only 22 (43%) of the cases had a unanimous agreement by all 5 pathologists, whereas 16 (31%) of the cases did not reach an agreement by more than 3 pathologists. Pair-wise correlations between pathologists, evaluated by Cohen $\kappa$ values, ranged from 0.13 to 0.70, indicating poor to fair agreements ($>0.75$ as good, $<0.40$ as poor). The overall agreement among 5 pathologists was borderline fair with an overall $\kappa$ value of 0.41 ($SE = 0.042$). Although the original diagnoses of all cases were carcinoma, a diagnosis of PUNLMP was made by 3 pathologists in 2 biopsies. Combining PUNLMP and LG into an LG category did not substantially improve the agreement rate (overall $\kappa = 0.42$), indicating that, in this study, the main source of variation stemmed from the distinction between LG and HG.

To compare survivin and Ki-67 expression in lesions with different grades, we defined cases with the same grading by 4 or 5 pathologists as consensus cases. The 51 biopsies were found to fall into 3 groups: (1) consensus HG (18 biopsies), including all cases that 4 or 5 raters considered as HG; (2) consensus LG (17 biopsies), including all cases that 4 or 5 raters considered as LG, 1 case that 3 raters diagnosed as HG, and 1 case diagnosed as PUNLMP; and (3) indeterminate (IND) (16 cases), in which 3 or fewer raters reached agreement on either LG (including 1 case that 1 rater diagnosed as PUNLMP) or HG.

Review of the slides revealed 2 main reasons that might account for the interobserver variability. One is the existence of “focal HG” areas that were cited as the basis for grading as HG by some pathologists but dismissed by others, and the other is the “intermediate” morphology seen in some cases such as moderately high nuclear-cytoplasmic ratio, occasional mitotic activity, or some degree of nuclear pleomorphism.

**Survivin and Ki-67 Protein Expression**

The potential value of survivin and Ki-67 staining in this diagnostic setting was evaluated. As shown in Figure 1, A through I, the immunostaining pattern of survivin using the selected monoclonal antibody was almost exclusively nuclear, with rare cytoplasmic stain that was much weaker in intensity. In contrast to the adjacent benign urothelium that was essentially negative for survivin, both the LG and the HG groups showed specific nuclear staining in the urothelium (Figure 1, B, E, and H). In the LG cases, the staining was mostly limited to basal and para-basal layers, whereas the staining in HG lesions was more diffuse. The quantified mean percentage of urothelial cells with positive survivin nuclear staining (survivin scores) in LG lesions was $7.7\% \pm 7.5\%$ (mean $\pm$ SD), with a median value of 4.9%. In contrast, the HG lesions had significantly higher survivin scores ($22.7\% \pm 12.8\%$, $P < .001$) and median value (19.1%) (Figure 2, A).

In comparison to survivin, the difference in Ki-67 expression was less consistent between the consensus LG and HG lesions, and LG lesions were found to have either low or high Ki-67 expression (eg, Figure 1, C and F). As a result, the Ki-67 scores, although also significantly different between the 2 groups (LG, $18.7\% \pm 13.0\%$; HG, $36.7\% \pm 17.4\%$, $P = .003$), showed substantial overlap in ranges (Figure 2, B). Despite this, significant correlation was seen between survivin and Ki-67 scores (Figure 2, C; $r_s = 0.84$).

The expression of survivin and Ki-67 in the IND group was then evaluated. Survivin scores in the 16 biopsies of this group ($10.7\% \pm 11.4\%$) were found to be close to the LG group ($P = 0.20$) but significantly lower than those of the HG group ($P < .001$). Similarly, the Ki-67 scores in the IND group were indistinguishable from the LG group ($P = .87$) but significantly lower than the HG group ($P = .001$).

In addition, univariate logistic regression showed that both survivin and Ki-67 scores are significant predictors of pathologic grade ($P < .001$ for both) in the 51 cases. Excluding the controversial IND group, survivin (AUC = 0.90) outperformed Ki-67 (AUC = 0.79) as a good diagnostic marker to separate HG from LG, although their difference was not significant statistically in our limited samples ($P = .12$). By receiver operating characteristic curve analysis, a survivin score of 12% was found to be the cutoff value that would yield the best sensitivity and specificity in distinguishing consensus LG from HG lesions (Figure 2, A).

**The IND Group**

The 16 biopsies in the IND group were reviewed and found to segregate into 3 groups. The first group (5 cases) showed focal HG features in a mostly LG lesion, and these HG areas usually corresponded to areas with higher expression of survivin. The architectural features in these focal HG areas were often similar to the remainder of the biopsy, but higher nuclear-cytoplasmic ratio, nuclear hyperchromasia, and moderate nuclear pleomorphism could be seen in some tumor cells. Because of the focal nature, these cases had survivin scores, calculated as the average of 5 to 10 fields, only slightly higher or similar to the LG group (9.3%–13.9%), although the survivin immunoreactivity was focally as high as 13.7% to 19.2%.

Most of the IND cases (9/16) constituted the second group and had homogeneously low survivin levels (<9.2%). Histologically, most of these cases deviated from the typical LG tumors in that a few enlarged nuclei might be present that led to some degree of nuclear atypia (Figure 3, A and B), sometimes accompanied by a partial loss of polarity (Figure 3, B). Although these findings might be concerning and have likely led to their grading as HG by some pathologists, these cases did not fulfill all criteria for HG on review and should be graded as LG based on the strict WHO criteria, and the low survivin immunostaining would have helped lead to the more appropriate grading.

The third group, containing 2 remaining cases in the IND category, showed homogeneous high levels of survivin (survivin scores, 15.3% and 49.9%), yet still maintained overall LG architectural features and polarity, with low to intermediate nuclear pleomorphism but increased mitoses (Figure 4, A through F). Of note, both cases also had high Ki-67 scores (27.8% and 52.4%, respectively), consistent with the increased mitotic activity. The strong positivity of both markers suggests that these 2 lesions are potentially HG biologically, despite their incompletely developed HG morphologic features that led to their diagnosis as LG by some pathologists.
Correlation Between Survivin Scores and Subsequent Diseases

We next examined the relationship between survivin scores and the subsequent recurrence of an HG papillary carcinoma in the 38 biopsies for which histopathologic follow-up data were obtained. These included 13 cases in the consensus LG group, 14 in the consensus HG group, and 11 IND cases. In total, 10 showed HG recurrence, including 1 (8%) of 13, 3 (28%) of 11, and 6 (43%) of 14 biopsies from the consensus LG, IND, and consensus HG groups, respectively. Using the histologic grades agreed on by 3 pathologists as the final grades for cases in IND group, a significant correlation was seen between higher histologic grades and HG recurrence (AUC = 0.67, P = .03) in these 38 biopsies. In comparison, survivin scores demonstrated an even stronger correlation to HG recurrence (AUC = 0.90, P < .001) and a significantly better predictive accuracy than the histologic grade (P = .01). Noteworthy, the survivin outlier in the LG group (Figure 2, A) and 1 of the 2 IND cases with high survivin and Ki-67 scores (shown in Figure 4, D through F) were from the same patient who had 2 HG recurrences. The other IND case with high survivin score (Figure 4, A through C) also showed HG recurrence subsequently.

If the quantitative survivin scores were converted into a dichotomized survivin grade (<12% vs ≥12%), the resulted survivin grade would give a similarly significant correlation with the HG recurrence (AUC = 0.86, P < .001) as the survivin scores. Interestingly, although Ki-67 scores did not correlate to the histologic grades as well as the survivin scores (Figure 2, B), they also showed a similarly good correlation to the HG recurrences (AUC = 0.86, P < .001).
Survivin and Ki-67 mRNA Expression

We additionally evaluated whether survivin and Ki-67 mRNA expression, determined by qRT-PCR, might be quantitatively more accurate and correlate better to the histologic grades.

Using normalized Ct values in which a lower Ct value reflects higher mRNA expression, the consensus HG and LG groups were significantly different in their survivin mRNA expression \((P = .005)\) and in Ki-67 expression \((P = .009)\), with survivin again a more superior marker. As shown in Figure 5, a good correlation was seen between survivin mRNA levels and immunostaining scores \((r_s = 0.79)\), confirming the significant difference between LG and HG groups and the similarity between LG and IND groups (Figure 5, A). A similar correlation was also seen between Ki-67 mRNA levels and Ki-67 scores (Figure 5, B; \(r_s = 0.75\)). However, the distribution ranges of survivin mRNA levels between LG and HG lesions overlapped,
and survivin mRNA was not a more discriminating marker than survivin scores by IHC.

**COMMENT**

According to the 2004 WHO classification, most noninvasive papillary urothelial carcinomas are dichotomized into LG or HG, which means cases that used to be considered as a general intermediate grade (grade 2) in the 1973 WHO system now need to be segregated into these 2 categories that are managed differently in clinical practice. In this study, we assessed the interobserver variability in distinguishing LG and HG carcinomas in an academic setting in which surgical pathologists have been well exposed to the 2004 WHO classification. We found fairly low overall diagnostic agreement, and close to one third of the 51 selected bladder biopsies were deemed as IND as no consensus could be reached. We then examined the expression of survivin and Ki-67 by IHC and by qRT-PCR and tested whether either marker could be useful in differentiating HG from LG lesions. For immunohistochemical analysis, to our knowledge our study is the first to quantify the survivin and Ki-67 scores using a computer-based objective method to improve accuracy and reproducibility.

Several different survivin antibodies have been used in the previous literature, and nuclear as well as cytoplasmic staining has been reported. Survivin is known to be associated with mitotic apparatus, centromere/kinetochore of chromosomes, and mitochondria, and different splice variants were also reported to have unique nuclear or mitochondrial localizations. As polyclonal antibodies tend to generate nonspecific background staining, in this study we used a monoclonal antibody raised against the full-length human survivin protein to avoid this potential pitfall. The survivin IHC pattern we observed was predominantly nuclear, with focal weak cytoplasmic stain in some cases that we did not find to correlate with tumor grade. This clean staining pattern suggests that some of the previously reported cytoplasmic staining with polyclonal anti-survivin antibodies might have been background staining, and data correlating such cytoplasmic staining to prognosis should be interpreted extremely cautiously.

We found nuclear survivin score to be a good adjunct diagnostic marker in separating LG and HG carcinomas. It was a better marker than Ki-67, with less overlap in value ranges between 2 groups and higher discriminating ability to predict pathologic grade. The similar distributions of survivin and Ki-67 scores between the IND group and the LG group strongly suggest a possibility of overgrading in some cases. This is not unexpected, as pathologists often have the tendency to err on the more aggressive side when in doubt, particularly in diagnosing biopsy specimens. Further review of the IND cases indeed identified 9 (56%) of 16 cases with somewhat disordered architecture or mild cytologic atypia that would otherwise be more suitably graded as LG. This conclusion was further supported by the fact that none of the 9 cases in this subgroup had HG recurrences on subsequent biopsies. If survivin score is used as an adjunct marker, this tendency to overgrade could conceivably be significantly decreased. Excluding these presumably overgraded cases, most of

---

**Figure 4.** Two possible “intermediate” cases from the indeterminate group with largely maintained polarity but mild to moderate increase in nuclear-cytoplasmic ratio and nuclear polymorphism, both showing high survivin and Ki-67 scores (hematoxylin-eosin, original magnification ×200 [A and D]; survivin immunoperoxidase, original magnification ×200 [B and E]; and Ki-67 immunoperoxidase, original magnification ×200 [C and F]).
the remaining cases in the IND category were biopsies that showed predominantly LG histology but with focal HG areas. Correlating to the histologic findings, these HG areas were found to have higher percentage of survivin-positive cells, indicating that survivin staining could also be useful in such cases, serving as a tool to highlight the HG areas.

Although this finding of heterogeneous survivin staining in some IND cases would suggest that areas with the highest immunoreactivity should be most informative, we have, in this study, purposefully decided to use a computer-based image tool to calculate the average percentage of survivin-positive cells in 5 to 10 high-power fields per case because this would give us the most objective and accurate scoring. Not meant to be used in the daily practice, this methodology has allowed us to propose that in a histologically uniform papillary urothelial carcinoma, survivin positivity in less than ~10% of the cells would be suggestive of an LG lesion, whereas positive staining in more than 15% to 20% is strongly indicative of an HG lesion. In cases with heterogeneous staining one should focus on areas with high survivin-staining percentages, after excluding possibilities such as tangential sectioning of basal and parabasal layers in an LG lesion. We believe that this could potentially be a useful tool in grading non-invasive papillary urothelial carcinoma. Our finding that survivin scores correlated better and had a higher predictive value than histology for HG recurrence further supports the potential value of survivin as an adjunct test, and additional prospective studies with clinical follow-up in a larger cohort would be important to further validate these findings.

In contrast to overgrading, our results concluded that undergrading of these papillary carcinomas, which would be clinically more ominous, was much less common. However, we did identify unusual cases that maintained overall LG architectural features and polarity but showed mild to moderate nuclear pleomorphism and/or increased mitotic activity, with homogenously high survivin and Ki-67 scores. Such cases were found in 3 biopsies from 2 patients, including 2 biopsies in the IND group and the survivin outlier in the HG group. We suspect that the strong positivity of both markers suggests that these cases, although morphologically closer to LG than to HG based on current criteria, might be biologically more advanced than the true LG lesions. The finding that both patients had follow-up biopsies that showed HG carcinoma further supports this possibility.

Consistent with this notion was the observation that the distributions of survivin and Ki-67 scores among the examined carcinomas are continuous rather than dichotomous, indicating a continuous biologic spectrum rather than 2 distinct groups of lesions. This implies that separating noninvasive papillary carcinoma into HG and LG is convenient but artificial, and pathologists would inevitably encounter cases with “intermediate” morphologic features that cannot be easily classified using the 2004 WHO classification system. It would be important to identify more cases in this intermediate category and analyze their long-term prognosis. This type of study would help address whether the current grading system needs to be further refined and whether markers such as survivin could be used as prognostic indicators beyond histologic grading.

We thank Mohammed Akhtar, MD, Debra Beneck, MD, and Syed Hoda, MD, for their participation in the histologic grading of the cases in this study and for their valuable comments. We also thank Joan Jones, MD, and Ximing Yang, MD, PhD, for their critical review of the manuscript.

References


CAP ‘08 ABSTRACT PROGRAM

Abstract and case study submissions for the upcoming CAP ‘08 meeting will be accepted beginning on February 1, 2008 through March 28, 2008. Accepted submissions will be published in the September 2008 issue of the Archives. Visit www.CAP2008.org for additional program details and to submit abstracts and case studies.