Reproducibility of Fuhrman Nuclear Grading of Renal Cell Carcinoma: A Preliminary Study*

Background: The aim of this study was to evaluate intraobserver and interobserver reproducibility of Fuhrman nuclear grading of renal cell carcinoma.

Methods: Pathology slides from 46 cases of renal cell carcinoma were rescored by 2 pathologists according to the Fuhrman system. Both intraobserver and interobserver reproducibility were assessed using kappa statistics.

Results: The initial Fuhrman grade was grade 1 in 4 of the cases (8.7%), grade 2 in 30 (65.2%), grade 3 in 11 (23.9%), and grade 4 in 1 (2.2%). After reviewing the slides by the same pathologist, grades were reassigned as follows: grade 1 in 8 cases (17.4%), grade 2 in 23 (50%), grade 3 in 14 (30.4%), and grade 4 in 1 (2.2%). Intraobserver reproducibility of the Fuhrman system was substantial (κ = 0.66). Fuhrman grading by the second pathologist was grade 1 in 11 cases (23.9%), grade 2 in 27 (58.7%), grade 3 in 7 (15.2%), and grade 4 in 1 (2.2%). Interobserver reproducibility of the Fuhrman system was moderate (κ = 0.42).

Conclusions: Despite substantial intraobserver reproducibility of Fuhrman grading, moderate interobserver reproducibility and low agreement for grade 3 should be a consideration.

Key Words: Fuhrman nuclear grade, renal cell carcinoma, reproducibility

Renal Hücreli Karsinomda Fuhrman Nükleer Derecenin Tekrarlanabilirliği: Ön Çalışma

Amaç: Bu çalışmada, renal hücreli karsinomda Fuhrman nükleer derece sisteminin gözlemci içi ve gözlemciler arasındaki tekrarlanabilirliğini değerlendirmek amaçlanmıştır.


Bulgular: Birinci patoloğun ilk değerlendirmesinde; derece 1: dört (% 8,7), derece 2: otuz (% 65,2), derece 3: on bir (% 23,9) ve derece 4: bir (% 2,2) olduğu belirlenmiştir. Aynı patoloğun ikinci değerlendirmesinde; derece 1: sekiz (% 17,4), derece 2: yirmi üç (% 50), derece 3: on dört (% 30,4) ve derece 4: bir (% 2,2) olsaqua saptanmıştır. Fuhrman sisteminin gözlemci içi tekrarlanabilirliği güçlü düzeyde bulunmaktadır (κ = 0,66). İkinci patoloğun Fuhrman derecelendirmesinde; derece 1: on bir (% 23,9), derece 2: yirmi üç (% 58,7), derece 3: yedi (% 15,2) ve derece 4: bir (% 2,2) olduğu belirlenmiştir. Fuhrman sisteminin gözlemciler arasındaki tekrarlanabilirliği ise orta düzeyde bulunmaktadır (κ = 0,42).

Sonuç: Fuhrman derecenin gözlemci içi güçlü tekrarlanabilirliğine rağmen, gözlemciler arasında orta düzeydeki tekrarlanabilirliği ve derece 3’teki düşük uyum göz önünde tutulmalıdır.

Anahtar Sözcükler: Fuhrman nükleer derece, renal hücreli karsinom, tekrarlanabilirlik

Introduction

Renal cell carcinoma (RCC) accounts for 90% of all primary malignant renal tumors in adults of both sexes (1-3). Staging and nuclear grading of RCC are considered important predictors of survival (1-5). Numerous grading systems have been developed for RCC. Skinner et al. (6) were the first to propose a grading system based solely on nuclear morphology, in 1971, which was later simplified by Fuhrman et al. in 1982 (7). Currently, the Fuhrman system is the most widely used nuclear grading system for...
grading RCC (4,8,9). The system is based on nuclear size, shape, and prominence of nucleoli (7). Despite wide acceptance of the Fuhrman nuclear grading system, its reproducibility has been questioned. There is a limited number of series that have assessed intraobserver and interobserver variability of the Fuhrman system. The aim of the present preliminary study was to assess intraobserver and interobserver reproducibility among 2 pathologists of the Fuhrman nuclear grading system for RCC.

Materials and Methods
This study included 46 patients that had undergone radical nephrectomy for RCC at the pathology departments of the Faculty of Medicine, Zonguldak Karaelmas University and Lutfi Kirdar Kartal Training and Research Hospital between 2005 and 2007. The histopathological types of the 46 RCCs were conventional (n = 37, 80.4%), chromophobe (n = 5, 10.9%), and papillary (n = 4, 8.7%). Clinical features were obtained from hospital records. All histopathological sections were fixed in formalin, embedded in paraffin, cut into 5 μm sections, and stained with hematoxylin and eosin (H&E).

Slides of all the cases were rescored according to the Fuhrman nuclear grading system by 2 pathologists blinded to the original scores. There was a 2-month interval between each grading round. Tumors were graded as follows: grade 1 tumors were composed of cells with small (~10 μm), round, uniform nuclei and inconspicuous or absent nucleoli; grade 2 tumor cells had larger (~15 μm) nuclei with irregular outlines and nucleoli that were visible under high-power (400x) microscopy; grade 3 tumor cells had even larger nuclei (~20 μm) with obviously irregular outlines and prominent nucleoli, even under low-power (100x) microscopy; grade 4 tumors exhibited features similar to grade 3 tumors, but also had bizarre and multilobed nuclei, and clumped chromatin. The nuclear grade was assigned to each tumor according to its least differentiated area (7).

Intraobserver variation was determined as the first pathologist’s grading at 2 different times. Interobserver variation was determined by the second pathologist, who did not know the first pathologist’s results. Data were analyzed using SPSS v.11.0 (SPSS, Chicago, IL, USA). Kappa statistics was used to evaluate the concordance between the original and reviewed nuclear grades (fair agreement, κ = 0.00-0.20; moderate agreement, κ = 0.21-0.45; substantial agreement, κ = 0.46-0.75; near-perfect agreement, κ = 0.76-0.99; perfect agreement, κ = 1.00). Slides of all the cases were pathologically staged by the first pathologist according to the 1997 TNM criteria. Tumor sizes were also recorded. The relationship between Fuhrman nuclear grade and pathologic stage were determined by correlation analysis with Pearson’s correlation coefficients.

Results
Mean age of the patients was 59.37 ± 11.89 years (range: 32-80 years). Sixteen patients (34.8%) were female and 30 (65.2%) were male. Maximum tumor diameter ranged from 2 to 14 cm (mean: 6.78 ± 3.14) (Table 1).

The initial Fuhrman nuclear grade was grade 1 in 4 patients (8.7%), grade 2 in 30 (65.2%), grade 3 in 11 (23.9%), and grade 4 in 1 (2.2%). The coexistence of 2 different grades in the same tumor was observed in 47.8% of the cases. After reviewing slides by the same pathologist, nuclear grades were reassigned as follows: grade 1 in 8 patients (17.4%), grade 2 in 23 (50%), grade 3 in 14 (30.4%), and grade 4 in 1 (2.2%).

Intraobserver reproducibility of Fuhrman nuclear grading was substantial (κ = 0.66); 37 (80.4%) of the cases were given concordant grades by the first pathologist. Tumors originally classified as grade 1 were upgraded by 1 grade in 25% of the cases. Tumors originally classified as grade 2 were upgraded by 1 grade in 10% of the cases. Tumors originally classified as grade 3 and grade 4 remained unchanged (100%) (Table 2).

Fuhrman nuclear grading of the second pathologist was grade 1 in 11 patients (23.9%), grade 2 in 23 (58.7%), grade 3 in 7 (15.2%), and grade 4 in 1 (2.2%). Interobserver reproducibility of Fuhrman nuclear grading was moderate (κ = 0.42); the same grade was obtained by the 2 pathologists in 31 of the cases (67.4%). Tumors classified as grade 1 by the first pathologist were upgraded by 1 grade in 25% of the cases. Tumors classified as grade 2 by the first pathologist were upgraded by 1 grade in 3.3% of the cases. Tumors classified as grade 3 by the first pathologist were under graded by 1 grade in 10% of the cases. Tumors originally classified as grade 4 remained unchanged (100%) (Table 3).
Tumors were pathological stage 1 in 29 patients (63%), stage 2 in 13 patients (23.3%), stage 3 in 2 patients (4.3%), and stage 4 in 2 patients (4.3%). Pathologic stage was significantly correlated with initial ($r = 0.66, P = 0.01$) and reviewed ($r = 0.50, P = 0.01$) Fuhrman nuclear grades ($r = 0.57, P = 0.01$) (Figure).

Intraobserver agreement for Fuhrman nuclear grading was 75% for grade 1, 73.3% for grade 2, 100% for grade 3, and 100% for grade 4. Interobserver agreement for Fuhrman nuclear grading was 75% for grade 1, 70% for grade 2, 54.5% for grade 3, and 100% for grade 4. The $x^2$ test for homogeneity was performed to determine if proportions of the pathologists’ grades revealed a similar pattern. The null hypothesis of similarity was rejected for both grading rounds ($P < 0.001$).

SD: Standard deviation. Tumor size: Represents the single greatest dimension.
Table 3. Interobserver reproducibility of Fuhrman nuclear grades.

<table>
<thead>
<tr>
<th>Fuhrman nuclear grade</th>
<th>Second pathologist</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>First pathologist</td>
<td>1</td>
</tr>
<tr>
<td>(initial assessment)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Number of cases</td>
<td>11</td>
</tr>
</tbody>
</table>

$\kappa = 0.42$

Discussion

The Fuhrman nuclear grading system is currently the most widely used grading system for RCC worldwide (4,7,8). It is applicable to any size, pattern, and cell-type of RCC (7). It has been shown that this nuclear grading system has prognostic value, especially for conventional and papillary RCCs (9). The Fuhrman grading system is widely accepted because of its simplicity and proven correlation with other pathologic variables. In the present study pathologic stage was significantly correlated with the Fuhrman nuclear grades of both pathologists. Several studies have reported differences in RCC-specific survival after separating cases with grade 1-2 tumors from those with grade 3-4 tumors. Additionally, the prognostic value of the Fuhrman nuclear grading system has been confirmed by most studies; patients with grade 3 or 4 tumors had less favorable prognoses than those with grade 1 or 2 tumors (10-16). Problems with Fuhrman nuclear grading include intraobserver and interobserver variability and reproducibility among pathologists, and this variability has been explored in several published studies.

In the present study intraobserver agreement was substantial ($\kappa = 0.66$, 80.4% concordance rate) and higher than that reported by Al-Aynati et al. ($\kappa = 0.29$ to 0.62, mean: 0.45) (17); however, interobserver agreement was moderate ($\kappa = 0.42$, 67.4% concordance rate), which is consistent with the results obtained by Ficarra et al. ($\kappa = 0.44$) (17), Lanigan et al. ($\kappa = 0.33$) (18), Al-Aynati et al. ($\kappa = 0.19$ to 0.44, mean: 0.29) (19), and Lang et al. ($\kappa = 0.09$ to 0.36, mean: 0.22) (20). The interobserver concordance rate (67.4%) obtained in the present study is lower than that reported by Bretheau et al. (95%) (21). In the present study intraobserver agreement for Fuhrman nuclear grades 3 (100%) and 4 (100%) was high, as was interobserver agreement for grades 1 (75%) and 4 (100%); however, intraobserver agreement for Fuhrman nuclear grade 2 (73.3%) and interobserver agreement for grade 3 were lower (54.5%).

Intraobserver and interobserver variability for Fuhrman nuclear grading can be due to several factors. First, RCC is a heterogeneous tumor that is usually composed of cells of different grades. In the presented series coexistence of 2 grades in the same tumor was observed in 47.8% of the cases. Fuhrman nuclear grade depends on the outlines of cell nuclei, the presence of nucleoli, and nuclear size, and must therefore be determined according to the least differentiated cell component (7). Searching for the least differentiated component is a time consuming procedure and requires investigating all tumor slides carefully, which may contribute to the variability in grading among pathologists. Sampling an inadequate number of tumors and suboptimal tissue fixation may be other challenging factors in the identification of the area with the highest grade. A homogeneous appearance in grade 1 RCCs and bizarre, multilobed nuclei in grade 4 tumors might contribute to higher grading reproducibility. On the other hand, pathologist subjectivity in assessing nuclear shape, cell outlines, and the presence of nucleoli in grade 2 and 3 tumors might be the cause of lower reproducibility. Several variants of the Fuhrman grading system have recently been described (22). It has been suggested that collapsing the Fuhrman grading system into a 2- or 3-tiered scheme can improve intraobserver and interobserver reproducibility (17). The use of nuclear morphometry in conjunction with histopathological grading may ensure more objective assessment of RCCs (23,24).
Fuhrman nuclear grading of RCC has shown a significant prognostic correlation with metastatic potential and patient survival in several studies (10-16); however, studies measuring intraobserver and interobserver agreement of Fuhrman grade show poor results in terms of reproducibility. In the current study, intraobserver and interobserver agreement were substantial and moderate, respectively. These variations can be explained, to a certain degree, by pathologist subjectivity in application of nuclear grades. There remains a need for improved standardization of nuclear criteria.

References


