The chromatin pattern of cell nuclei is of prognostic value for renal cell carcinomas

Christine François[a], Myriam Remmelink[b], Michel Petein[b], Roland van Velthoven[c],
André Danguy[a], Eric Wespes[d], Isabelle Salmon[b], Robert Kiss[a,*,**] and
Christine Decaestecker[a,***]

[a]*Laboratory of Histology, Faculty of Medicine, Université Libre de Bruxelles, Brussels, Belgium*
[b]*Department of Pathology,* [d]*Department of Urology, Erasmus University Hospital, Brussels, Belgium*
[c]*Division of Urology, Department of Surgery, J. Bordet Institute, Brussels, Belgium*

Abstract. Using a series of 105 renal cell carcinomas (RCCs) we investigated whether features quantitatively describing the appearance of Feulgen-stained nuclei and, more particularly, of their chromatin (on the basis of computer-assisted microscopy) can contribute any significant prognostic information. Thirty morphonuclear and 8 nuclear DNA content-related variables were thus generated. The actual prognostic values of this set of cytometric variables was compared (by means of discriminant statistical analysis) to conventional diagnostic and/or prognostic markers including histopathological grades, tumour invasion levels and the presence or absence of metastases. We obtained complete clinical follow-ups for 49 of the 105 RCC patients under study, making it possible to define a subset of patients with a bad prognosis (i.e., who died in the 12 months following nephrectomy) and a subset of patients with a good prognosis (i.e., who survived at least 24 months following nephrectomy). An original method of data analysis related to artificial intelligence (decision tree induction) enabled a strong prognostic model to be set up. In the case of 10 new patients, this model identified all the dead patients as having a bad survival status, with a total of 8 correct predictions. Another prognostic model similarly generated enabled the correct predictions to be confirmed.

Keywords: Renal cell carcinoma, prognosis, Feulgen staining, image cytometry, chromatin pattern, DNA ploidy, artificial intelligence

1. Introduction

Renal cell carcinomas (RCCs) constitute a highly unpredictable neoplasm with a tendency to recur or progress, and to cause death many years after initial treatment [27]. In comparison with stage and grade, which remain the most useful tools in the case of RCCs, neither DNA quantification nor oncogene expression are useful as indicators of behavior at clinical level [17]. The prognosis of RCCs

---

is also determined by two other variables, i.e., the tumour burden (expressed in the size or weight of the primary tumour), and the extension of the disease (lymph node invasion, distant metastases or renal vein invasion) [31].

Although many RCC grading systems have been described [1,9,29,30], none can be said to have achieved widespread acceptance [17,21].

Of the different RCC grading systems, Fuhrman et al. [9] suggest retaining exclusively nuclear features. Using these features, Tosi et al. [31] report that survival is found to be a function of nuclear grade. However, these authors [31] also report that although this nuclear grading is relatively simple when compared to other grading systems, it is still highly subjective and therefore does not guarantee, either high consistency or reproducibility in its results.

Most of the terms used by Fuhrman et al. [9] to describe cell nucleus appearance and, particularly, chromatin pattern, can be quantified by means of the computer-assisted microscope analysis of Feulgen-stained nuclei. This has been demonstrated in a recent investigation on RCCs [7], as for other tumours of the urological system including the prostate [22,23] and the bladder [6,33].

The aim of the present study is to investigate on a series of 105 RCCs whether features quantitatively describing the appearance of Feulgen-stained nuclei and, more particularly, their chromatin can contribute any significant prognostic information. Recent work shows the relevance of using chromatin features as prognostic factors in the case of patients with prostate [15,35], oesophageal [11], renal [32] and brain [4] cancers. In the present study, a set of 30 quantitative morphonuclear variables was determined by means of computer-assisted microscopy. Their actual prognostic values were evaluated by means of discriminant analysis (a multifactorial statistical method of data analysis). We further investigated whether the quantification of the nuclear deoxyribonucleic acid (DNA) content (the DNA ploidy level) of the RCC can improve the accuracy of the prognostic value of quantitative chromatin pattern description. As previously described, the DNA ploidy level was determined by means of 8 distinct quantitative variables [3].

As a final step, a logical model was established by means of the decision tree technique as the means of making practical use (as information additional to conventional histological grading or clinical staging) of the prognostic information contributed by cytometry-generated features. In contrast to conventional discriminant analysis, this technique produces explicit logical classification rules which, as has previously been shown, can be easily used as an aid to tumour diagnosis and prognosis in the bladder [3], the soft tissues [12], and the brain [4].

2. Materials and methods

2.1. Patients' characteristics and clinical data

One hundred and five renal cell carcinomas were obtained from patients who had undergone partial or radical nephrectomy at the Erasmus Hospital and the J. Bordet Institute between 1987 and 1996. Histologically, 87 of the tumours were of clear cell type, 4 were granular, 7 chromophobe, 3 sarcomatoid and 4 had papillary patterns. The series included 45 women and 65 men aged from 27 to 85 years (median age: 60 years). The length of clinical follow-up ranged from 2 to 74 months. The survival status was determined as "good" for those patients who were alive at least 24 months after surgery, and "bad" for those who died of cancer in the 12 months following nephrectomy. On the basis of these cut-off values, 49 cases were available to define the two groups of patients, i.e., the Good group, which contained 35 patients who did not die from their disease during the time of the

analysis and whose post-nephrectomy clinical follow-up ranged between 24 and 74 months (mean: 45.7 months), and the Bad group, which contained 14 patients who died from their disease between 1 and 12 months (mean: 10.9 months) following nephrectomy. (None of those patients had received any treatment in addition to nephrectomy.)

The tumour samples were diagnosed and graded independently by three pathologists. The histopathological grading was performed according to the system proposed by Furhman et al. [9].

The series of 105 tumours included 8 grade I, 61 grade II, 31 grade III and 5 grade IV. In order to obtain more homogeneous groups in terms of malignancy levels we merged the grade I and grade II specimens into a low-grade group ($n = 69$), and the grade III and grade IV specimens into high-grade group ($n = 36$).

The clinical stage was determined according to the TNM classification [14]. The series of 105 RCCs thus included 4 T1, 73 T2, 18 T3A, 6 T3B and 4 T4. We also merged these five groups into one low-invasive (T1, T2; $n = 77$) and one highly-invasive (T3A, T3B, T4; $n = 28$) groups.

Since the presence or absence of metastasis is of high prognostic value, we also took this into consideration. We obtained evidence of the presence of metastases (M^+ group) in 18 patients and of the absence (M^- group) in 87.

An independent set of 10 patients with similar clinical characteristics and complete follow-ups was used in a validation phase.

2.2. Cytological samples

The tumour cases came from archive materials, i.e., formalin-fixed paraffin-embedded tissues. Two paraffin blocks were available for each case. Five sections were cut from each block. The first, third and fifth, were 5 μm thick, stained with hematoxylin–eosin and subjected to histopathological diagnosis and grading. The second and fourth were 50 μm thick and were submitted to a method described in detail elsewhere [16]. Briefly, this method enables single-cell nuclei suspensions to be obtained (after pronase digestion) that are centrifuged onto glass slides (cytospins) and then stained by means of the Feulgen reaction. This process depends on a series of previously analyzed factors in the determination of the appropriate laboratory routine protocols [16].

2.3. Computer-assisted microscopy

Feulgen-stained nuclear images of between 350 and 700 cells were digitised for each case. Computer-assisted microscopy was performed by means of a SAMBA 200 system (Alcatel-TITN, Grenoble, France). The way in which the cell nuclei were selected for analysis and then digitised is detailed elsewhere [3,6,22,33].

The variables quantitatively describing the DNA ploidy level included the DNA index (which was calculated according to the method proposed by Wheeless et al. [34]) and 7 variables determining the proportion of the different DNA ploidy-related stem cell lines in each of the 105 RCCs under study. These proportions were calculated as detailed elsewhere [3]. Briefly, optical density was computed on 256 densitometric values (gray levels) on each pixel of the digitized image of a Feulgen-stained nucleus. The nuclear DNA content of each nucleus was calculated from its integrated optical density value (IOD), i.e., the sum of the optical density values of each pixel of a nucleus. This procedure was carried out on the 350–700 cell nuclei analyzed per case and enabled a DNA histogram to be computed for each one. This histogram was characterised by 7 variables determining the percentages of the different ploidy-related cell sub-populations in each histogram. These variables are the percentage of

Table 1
Cytometrical variables

DNA ploidy level	Morphonuclear and chromatin texture	
	Mean	Standard deviation
DNA index (DI)	P1 = nuclear area	SDP1 (anisonucleosis level)
% diploid cell nuclei (%2C)	P2 = integrated optical density	SDP2
% hyperdiploid cell nuclei (%H2C)	P3 = mean optical density	SDP3
% triploid cell nuclei (%3C)	P4 = skewness index	SDP4
% hypertriploid cell nuclei (%H3C)	P5 = variance of optical density	SDP5
% tetraploid cell nuclei (%4C)	P6 = kurtosis index	SDP6
% hypertetraploid cell nuclei (%H4C)	P7 = short run length frequency	SDP7
% pentaploid cell nuclei (%5C)	P8 = long run length frequency	SDP8
	P9 = grey level distribution	SDP9
	P10 = relative distribution frequency	SDP10
	P11 = relative distribution percentage	SDP11
	P12 = local mean	SDP12
	P13 = energy	SDP13
	P14 = matrix coefficient variance	SDP14
	P15 = contrast	SDP15

diploid (%2C variable), hyperdiploid (%H2C variable), triploid (%3C variable), hypertriploid (%H3C variable), tetraploid (%4C variable), hypertetraploid (%H4C variable) and pentaploid (%5C variable) cell nuclei in each DNA histogram.

The 30 morphonuclear variables included the mean and standard deviation values calculated for 15 different morphonuclear features quantitatively describing the morphology of the cell nuclei and their chromatin patterns. These 15 morphonuclear features were calculated using four different algorithms [2,10,13].

The first algorithm calculates one geometric feature, i.e., the nuclear area (P1 feature), which determines nuclear size. The second algorithm calculates 5 features describing the cell nuclei densitometrically. These 5 features are the integrated optical density (P2 feature), the mean optical density (P3 feature), the kurtosis (P6 feature) and skewness (P4 feature) indices, and the variance of optical density (P5 feature). The third algorithm calculates 5 features describing chromatin texture by means of the run length matrix. These 5 features are the short (P7 feature) and long (P8 feature) run length frequencies, their relative distributions (P10 feature) and percentages (P11 feature), and the grey level distribution (P9 feature). Finally, the fourth algorithm also calculates features describing chromatin texture, but using a co-occurrence matrix. The last 4 features include the local mean (P12 feature), energy (P13 feature), the matrix coefficient variance (P14 feature) of the co-occurrence matrix, and contrast (P15 feature).

The calculation of the standard deviation values relating to each of these 15 features (SDP1–SDP15 features) enabled the level of tumour heterogeneity to be characterised in each tumour group under study. More particularly, the standard deviation (SDP1) of the first morphonuclear feature (P1, nuclear size) determines the anisonucleosis level.

All these variables are summarised in Table 1.

2.4. Data analysis

2.4.1. Discriminant analysis

Discriminant analysis, a multivariate statistical analysis (the mathematical principle of which is detailed elsewhere [19]), was used to bring out the real diagnostic and/or prognostic values contributed by the morphonuclear and/or DNA ploidy-related variables.

Discriminant analysis enables data to be described in terms of "separability" between different *a priori* defined groups, also named "classes" (e.g., histopathologically- and clinically-related groups). Discriminant analysis is therefore able to generate a description of data by means of discriminant factors which give prominence to the differential features of cases from distinct classes. In the present study, only 2-class problems were considered (i.e., low-grade versus high-grade, low-invasive versus highly-invasive, M^- versus M^+, and the Good versus Bad groups), for which only one discriminant factor can be generated by discriminant analysis [19]. A measure of the separation between the classes can be evaluated by calculating the distances between each class-centroid. A centroid represents the "mean position" on the discriminant factor of the cluster of cases defining a given group. The distances between centroids are usually measured by means of *squared mahalanobis distances* (SMDs), with take into account both variances and covariances in the space of the variables selected in the discriminant model. A level of statistical significance in the shape of a *p*-value can be associated with each SMD. A *p*-value smaller than 0.05 thus indicates that the two centroids can be considered as significantly distinct. In the present case, this means that the two tumour groups thus analysed (and represented by their centroids) exhibit significantly distinct morphonuclear (and/or DNA ploidy-related) profiles. This is a necessary but insufficient condition for ensuring the good classification of individual cases (see the classification principles below).

As discriminant factors give the best representations for separating data into different classes, discriminant analysis uses these factors to allocate cases to one or another of these classes. Briefly, the allocation rule consists in attributing a case to the nearest class-centroid in the discriminant factor space. This rule produces linear separating surfaces which are the median hyperplanes (in the factorial space) between the centroids of the different classes. This rule makes the implicit hypothesis that the *a priori* probabilities of the classes are equal, and is easily adapted if this is not the case. In the present study, we imposed equal *a priori* probabilities because of markedly unbalanced case distribution between the classes. In fact, in generating allocation rules, discriminant analysis bases itself on a statistical model which hypothesizes that the data are normally distributed into each class, and that the class variance–covariance matrices are equal.

The descriptive accuracy of the allocation rules (for the data under study) is generally detailed in a confusion matrix showing the number of misclassifications across the groups, and the total classification rate.

A stepwise procedure was used in the present work to progressively insert the variables into the discriminant model and thus to be able to identify the variables most useful in describing specific characteristics across the different RCC groups under study [19]. In the present study, the stepwise discriminant analysis program of the *Statistica* (Statsoft, Tulsa, OK, USA) package was used together with the default options (the entering of variables in the model with *F*-values larger than 1). However, the insertion of variables was stopped when the classification rate associated with the model stopped improving.

2.4.2. Decision tree (*DT*)

The DT technique is a supervised learning method which produces logical classification rules [24]. These rules express relations, detected in the data, between some feature values and pre-assigned classes. These techniques "naturally" break down a classification problem into a hierarchical set of questions concerning the value of the selected features. These techniques thus construct classification rules in the form of trees, where each terminal node (leaf) is labelled as a class and each intermediate node tests a feature of a case, the outcome of which specifies the branch to be followed in the tree (see

Table 3, for example). DT algorithms generally use a two-phase scheme, namely: (a) the growing phase for the construction of the classification rules, and (b) the pruning phase, which simplifies the trees to avoid the overfitting of the training data. A specific algorithm, i.e., C4.5 [25], was used in the present work. This algorithm uses an entropy criterion (based on Shannon's information theory [28]) to select a test at each node of the tree (growing phase), and then prunes the tree (pruning phase) using an error-based criterion with a binomial error approximation (as detailed in [25]). In this study, we used a pruning confidence level of 10% (in place of the default value of 25%) to increase the pruning ability of the algorithm [25]. Its purpose is to avoid the overfitting of small data series.

Originally, decision tree algorithms such as C4.5 (and also other supervised learning algorithms) were designed to handle data with *discrete* and symbolic feature values. In order to deal with continuously valued features, C4.5 must treat them as *discrete*, with many possible values (all the values observed). In the present study, a *discretisation* process was used which transforms each numerical feature into an ordinal one with a small number of values (prior to the learning process). This *discretisation* process is detailed and justified elsewhere [5]. Briefly, it consists of a sequence of k-means clusterings (for successive k-values: 2, 3, 4, etc., fixing the number of clusters) applied to each feature. The successive k-means results are then inspected for cluster stability detection. This aims to detect clusters which have a similar aspect in two or more successive clusterings. The clustering that gives the largest number of stable clusters is selected for each feature, and the cluster bounds are interpreted as discretisation intervals.

As for discriminant analysis, a decision-tree-adapted feature selection strategy was used to reduce the set of features submitted to these two methods. The selection strategy was based on a stability study for the different features according to a process detailed elsewhere [5]. Briefly, the stability of the features was measured by the selection frequency of each feature over the decision trees generated by a traditional cross-validation process (see below). The features which appeared to have a sufficient degree of stability (or selection frequency) were selected [5].

A traditional cross-validation technique was performed (also available with the C4.5 package) to finally select the best decision tree model (with or without data discretisation, and with or without feature selection) with respect to the data under study. This technique enabled the best model to be chosen through an evaluation of tree performance on new data [5]. Briefly, a cross-validation is performed by dividing the data into N blocks of cases that have similar sizes and class distributions ($N = 20$ in the present study). A decision tree is constructed from the remaining $N - 1$ blocks for each block in turn and tested on the unseen cases in the hold-out block. The classification rate averaged on the test blocks can be used as an estimate of the decision tree's accuracy on new data.

3. Results

3.1. Morphometrical and textural parameters

3.1.1. Tumour grade and invasion level

The top of Table 2 lists the 4 discriminant variables (in the order of their selection) selected by the stepwise process to distinguish between the low- and high-grade groups. Only these 4 variables (described in Table 1) were considered in the discriminant model. This enabled a significant distinction to be made between the two group-centroids, as is shown by the p-value ($p = 0.012$) associated with the inter-centroid distance. However, when the discriminant model was used to classify each tumour

Table 2
Discriminant analysis performed on the few morphonuclear variables selected during the stepwise process

	Variables in the model	*p*-levels (centroid distance)	Confusion matrices (equal *a priori* probab.)		
				L	H	% correct
Low/high histological grade	P9, P14, SDP8, SDP14	0.012	Low	52	17	75
			High	15	21	58
			Total rate			70
				L	H	% correct
Low/high invasion level (T variable in the TNM system)	SDP5, P5, SDP12	0.0027	Low	53	24	69
			High	10	18	64
			Total rate			68
				M^-	M^+	% correct
M^- (absence)/ M^+ (presence) of metastasis	SDP2, P12, SDP6	$< 10^{-4}$	M^-	78	9	90
			M^+	7	11	61
			Total rate			85
				G	B	% correct
Good/bad survival status	SDP2, P6	0.0006	Good	31	4	89
			Bad	6	8	57
			Total rate			80

In the confusion matrices the lines relate to actual groups and the columns to classifier decisions.

case, the classification rate did not exceed 70% and showed a high error rate for the High group (42% of errors, see the confusion matrix in Table 2).

Three variables were selected with respect to the tumour invasion level (i.e., low versus high, see Table 2) and these enabled a significant ($p = 0.005$) degree of discrimination to be made between the two group-centroids. A classification rate (68%) was obtained that was similar to the one noted above for the grading classification, but which had a more balanced classification rate in each group.

The relatively low classification rates obtained above indicate that on the basis of their morphonuclear characteristics those tumours characterised by a high grade or invasion level did not differ markedly from the tumours with a low grade or invasion level.

3.1.2. Presence versus absence of metastases

Figures 1(A) and (B) give a breakdown of the patients who either did (M^+ group) or did not (M^- group) show metastases as a function of tumour grade and invasion level (characterised by the T variable of the TNM classification). As will be seen, no efficient information concerning the presence or absence of metastases was provided by these two diagnostic features. While the χ^2-tests carried out on the two case distributions illustrated in Figs 1(A) and (B) were significant ($p = 0.037$ and 0.002 for tumour grade and invasion level, respectively), it can be seen that the patients with metastases (M^+ group) are equally distributed between low and high grades or T-levels. In fact, only the extreme grades (I and IV) or T-levels (T1 and T4) had any real prognostic value but these tumours concerned only a few cases in our series (data not shown).

In contrast to the invasion level and tumour grade, the morphonuclear variables provided useful and significant prognostic information, as shown in Table 2. Only 3 of the 30 quantitative morphonuclear variables were used in the discriminant model (Table 2). The level of statistical significance was as high as $p < 0.0001$ (Table 2). A total classification rate of 85% was obtained which is illustrated in Fig. 1(C). When comparing Fig. 1(C) with (A) and (B), the contribution of morphonuclear variables
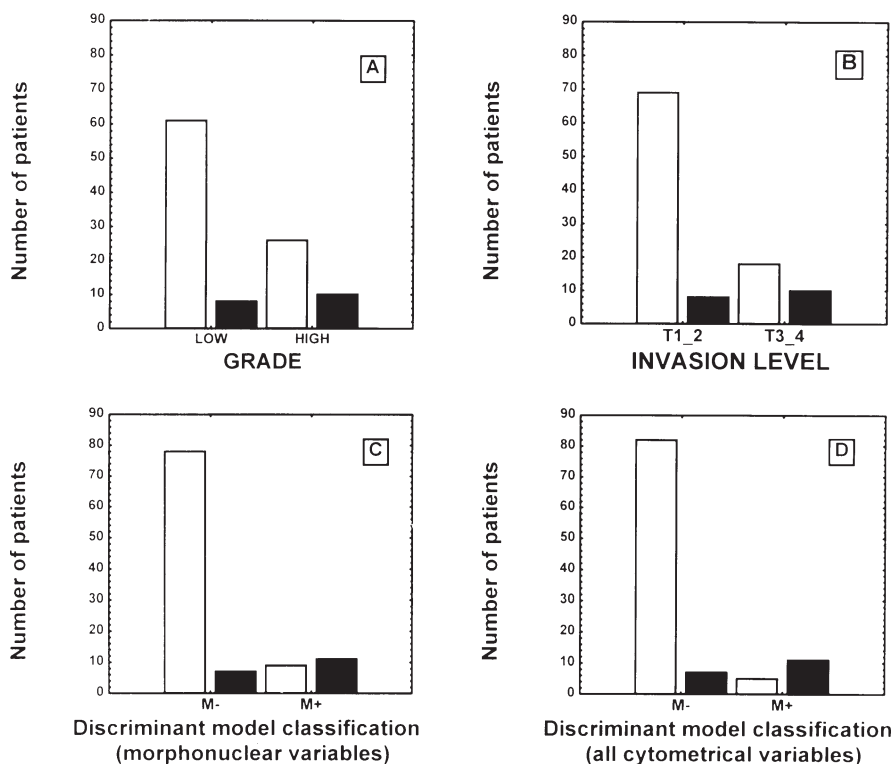
METASTASIS STATUS

Fig. 1. This figure shows the breakdown of patients who did (M^+ group, black column) or did not (M^- group, white column) show metastases (by the time of their nephrectomy) in function of tumour grade (A) and invasion level (B). Figures (C) and (D) show the classifications results obtained by means of discriminant analysis performed on morphonuclear variables only (C) and on both DNA ploidy-related and morphonuclear variables (D).

appears clearly. Indeed, of the 85 cases classified by discriminant analysis as M^-, 78 (i.e., 92%) related correctly to patients without metastases.

3.1.3. Survival status

Figures 2(A)–(C) show the breakdown of the patients in terms of good or bad survival status according to tumour grade, invasion level and the presence or absence of metastases. A low grade and invasion level were generally associated with a good survival status (Figs 2(A) and (B)). In contrast, the groups with a high grade or a high invasion level split almost equally into good and bad survival-status-related groups. In each case, the χ^2-test gave only a weak level of statistical significance ($p = 0.02$). This is in accordance with the data reported above on the presence/absence of metastases. A significant association ($p < 0.001$) did, of course, appear between survival status and the presence or absence of metastases (Fig. 2(C)).

The discriminant model generated from the morphonuclear features was based on two variables only (see Table 2). This model made possible a highly significant level of separability ($p = 0.0005$) between the two group-centroids and gave a total classification rate of 80%, with a confusion matrix similar to the one obtained for the problem concerning the absence or presence of metastasis (see Table 2).
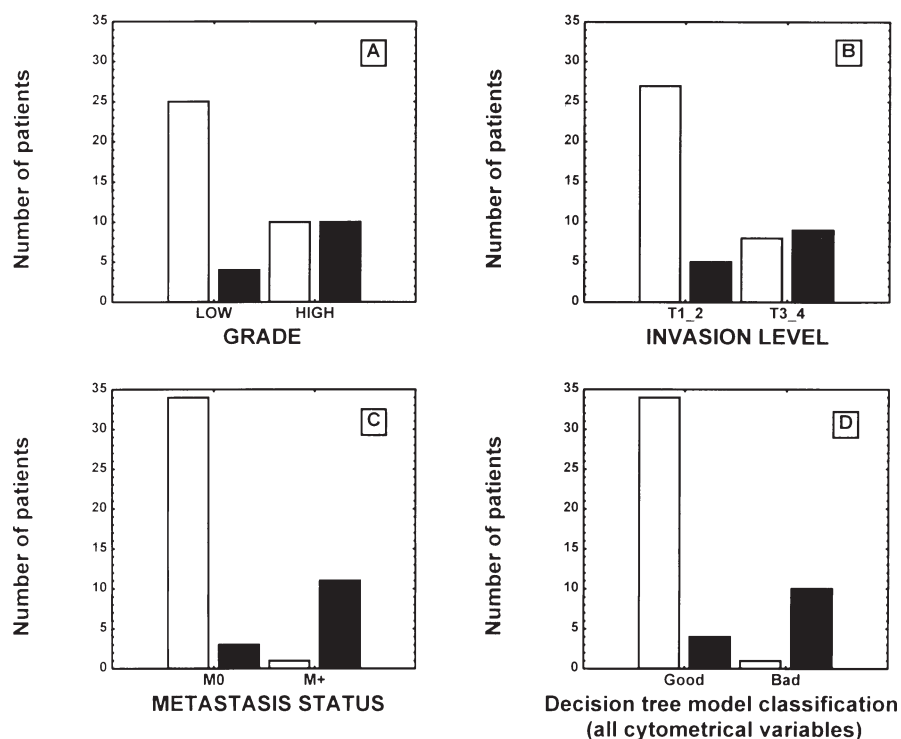
SURVIVAL STATUS

Fig. 2. This figure shows the breakdown of the patients in terms of good (white column) and bad (black column) survival status in function of tumour grade (A) and invasion level (B), and the presence or absence of metastases (C). Figure (D) shows the classification results obtained by means of the decision tree model performed on cytometry-generated variables.

3.2. Adding the 8 DNA ploidy-related variables to the 30 morphonuclear ones

The information contributed by only the 8 DNA ploidy-related variables yielded discriminant analysis results no better than those given above and concerning the morphonuclear variables (data not shown). Combining the 8 DNA ploidy-related variables with the 30 morphonuclear ones was no beneficial in distinguishing the different histological grades and invasion levels (data not shown). In contrast, as for as the two prognostic-related questions (presence/absence of metastases and the survival status) are concerned, a significant improvement was brought about by the combination of the two types of cytometry-related variables, as is illustrated in Table 3.

Three variables were selected for the metastasis-related groups (Table 3). The first is related to ploidy level and measures the percentage of highly aneuploid cells (%5C). The other two are morphonuclear and have already been selected in the model described in Table 2. The %5C variable was also the first selected to distinguish between the good and bad survival status (Table 3).

A comparison of the results in Tables 2 and 3 shows that the information contributed by the combination of ploidy-related and morphonuclear variables improved the results obtained with the morphonuclear variables only. In fact, these improvements concern the groups of patients with a good prognostic value (M^- and good status which obtained 94% of correct classifications), so enabling the total classification rate to be increased by about 4% for the two models.

Table 3
Discriminant analysis performed on the variables selected among DNA ploidy- and morphonuclear-related variables

	Variables in the model	*p*-levels (centroid distance)	Confusion matrices and classification rates			
				M^-	M^+	% correct
M^- (absence)/ M^+ (presence) of metastasis	%5C, SDP6, SDP12	$< 10^{-4}$	M^-	82	5	94
			M^+	7	11	61
			Total rate			89
				G	B	% correct
Good/bad survival status	%5C, P6, %H3C, SDP4	0.0003	Good	33	2	94
			Bad	6	8	57
			Total rate			84

In the confusion matrix the lines relate to actual groups and the columns to classifier decisions.

3.3. Decision tree models as RCC prognosticators

3.3.1. Contribution of cytometry-generated and/or clinical variables (*histopathological grade and tumour invasion level*)

In this section, three decision tree models are proposed to characterise patients' survival status. The first has been generated from the cytometry-generated variables only (labelled 'Model 1'), the second combines the information constituted by the cytometry- and invasion level-related variables (Model 2), and the third (Model 3) is based on invasion level and discretized data (see Section 2). The variables considered for each of them had previously been submitted to a stability study explained in Section 2. Only the most stable variables were selected and submitted to the decision tree algorithm (including a reinforced pruning phase). Table 4 shows the models finally obtained.

The left-hand column in Table 4 shows Model 1 which is based on cytometry-generated variables only. As can be seen, this model uses only 3 quantitative variables, i.e., 2 morphonuclear- (P4 and P6, see Table 1) and 1 DNA ploidy- (%5C) related one. This can be read as follows. If the P6 variable assumes a value larger than 304 for a case, it must be considered as having a bad prognosis. The numbers 5/0 in brackets indicate that 5 cases in our series satisfied this P6 variable-related condition and that each of them had a bad prognosis (0 classification error). If this first condition is not satisfied, the %5C variable-related condition must be considered. Thus, if the %5C variable assumes a value larger than 6, a bad prognosis must once again be considered. Of the four cases in our series satisfying the sequence of variable-related conditions leading to this branch, 1 had a good prognosis and was thus misclassified (see the second number in brackets). The last part of the tree can be read similarly. The confusion matrix below the tree shows that only 5 classification errors (10%) occurred in the set of 49 patients submitted to the model described here. These classification results are illustrated in Fig. 2(D). While the decision tree model used fewer variables than the discriminant analysis model at the bottom of Table 3, this tree gave a better discriminatory description of the data.

Model 2 is described in the right-hand column of Table 4. It proposes a good descriptive model with few variables (the T invasion level and 3 cytometry-related variables and once again the %5C variable). It can be read as the previous one described in the left-hand column.

In the bottom of Table 4, Model 3 combined the invasion level-related feature and *discretized* cytometry-generated variables. This combination was revealed as the best by cross-validation (data not shown). This time, the threshold values appearing in the tree correspond to discretized values (0, 1, 2, etc.). For example, discretized threshold 4 appearing in the first rule on the SDP4 variable

Table 4
Decision trees

Model 1: Cytometry-related features	Model 2: Invasion level and cytometry-related features
(After pruning)	(After pruning)
P6 > 304: BAD (5/0)	INV_LEV = T3_4:
P6 ⩽ 304:	| SDP9 > 1045: BAD (5/0)
| %5C > 6: BAD (4/1)	| SDP9 ⩽ 1045: GOOD (12/4)
| %5C ⩽ 6:	INV_LEV = T1_2:
| | P4 ⩽ −27.6: BAD (2/0)	| %5C ⩽ 2: GOOD (24/0)
| | P4 > −27.6: GOOD (38/4)	| %5C > 2:
	| | %3C ⩽ 5: BAD (5/0)
	| | %3C > 5: GOOD (3/0)

	G	B	% correct		G	B	% correct
Good	34	1	97	Good	35	0	100
Bad	4	10	71	Bad	4	10	71
Total rate			90	Total rate			92

Model 3: Stable and discretized features

(After pruning)			
SDP4 > 4: BAD (4/0)			
SDP4 ⩽ 4:			
| INV_LEV = T1_2: GOOD (30/3)			
| INV_LEV = T3_4:			
| | P12 > 3: BAD (3/0)	G	B	% correct
| | P12 ⩽ 3:	Good	35	0	100
| | | DI > 0: GOOD (6/0)	Bad	4	10	71
| | | DI ⩽ 0:	Total rate		92
| | | | %H3C ⩽ 0: BAD (3/0)			
| | | | %H3C > 0: GOOD (3/1)			

In the confusion matrices the lines relate to actual groups and the columns to classifier decisions.

corresponds to a numerical value of 50. As observed for Model 2, Model 3 also misclassified only 4 cases (8% of error) in the set of 49 patients submitted to the decision tree analysis.

3.3.2. Using decision tree models to predict patient status

The three models described above, i.e., Models 1, 2 and 3, were further used to classify 10 new patients who had not been included in the data series previously used to generate the different classification models. Table 5 describes the survival status of these patients as well as the predictions of the 3 decision tree-related models; disagreements are indicated in bold characters. It appeared that Models 1 and 3 gave the best prediction results (80% of correct predictions). It will be noted that only Model 3 identified all the dead patients as having a bad status. However, this model was inclined to overestimate tumour aggressiveness by classifying two living patients in the bad group. Inversely, Model 2 was more inclined to underestimate tumour aggressiveness. It will be noted that the patient 2 classified as bad by Models 2 and 3 is characterised by clinical bad prognostic factors: a high T-level (T3B), a high grade (III) and a sarcomatoid type. However, this patient was alive at the time of the analysis and had a follow-up period of 50 months (more than 4 years). This tumour was accordingly associated with a good prognosis from a "pure" cytometrical point of view, as the Model 1 results show.

If the results of the best two models are combined (i.e., Models 1 and 3), it will be seen that when these two models were in accordance, the associated predictions were always correct.

Table 5
Patients' status predicted by three decision tree-related models

Patient	Status	Surv. period[a]	Grade T-level	Model 1	Model 2	Model 3
1	Alive	49	III/T2	Good	Good	Good
2	Alive	50	III/T3B	Good	**Bad**	**Bad**
3	Alive	52	I/T2	Good	Good	Good
4	Alive	52	II/T2	Good	Good	Good
5	Dead	8	II/T1	**Good**[b]	**Good**	Bad
6	Alive	40	III/T3A	Good	Good	Good
7	Alive	38	II/T2	Good	Good	Good
8	Alive	23	III/T2	**Bad**	Good	Good
9	Alive	36	II/T3A	Good	Good	**Bad**
10	Dead	3	III/T4	Bad	**Good**	Bad

[a]Survival period in months.
[b]Decision tree predictions in disagreement with patient survival status are indicated in bold characters.

4. Discussion

RCC prognosis mainly relates to TNM clinical staging, which takes into account the level of invasion, i.e., local (the T variable) versus distal (the M variable) levels, and histopathological grade. However, even when TNM staging and histopathological grade are clearly identified, the course of patients with RCC may differ considerably, and approximately 50% with proven localised disease suffer metastasis after nephrectomy [26]. Ljungberg et al. [18] state that it is of great clinical importance to identify patients who will suffer from progressive disease later on because adjuvant therapy might be indicated for them.

The problem of reproducibility in the histopathological grading of RCCs is a real one. Twenty years ago, Syrjänen and Hjelt [30] tried to select the histopathological features to use to establish a reproducible grading of RCCs. They [30] tested four different grading systems, i.e., (1) a classification based on the histological type of tumour growth, (2) the grading of malignancy according to Arner et al. [1], (3) a grading based on the nuclear structure of carcinoma cells (this grading system has been developed further by Fuhrman et al. [9]), and (4) a combined grading system based on nuclear structure and the demarcation of a tumour from its surrounding tissue. As detailed below, this combination of diagnostic and/or prognostic information contributed by different systems was also exploited in the present study. Of the four systems described by Syrjänen and Hjelt [30], the histological type of growth appeared to be the least appropriate for grading RCCs, while the grading system based on nuclear tumour cell structure appeared to be an accurate measure of their intrinsic malignancy [30]. This latter grading, which is based on nuclear structures, was shown to give an excellent correlation between tumour grade and patients' 5-year survival rate [30]. This is the reason why we retained this grading system in the present study.

Syrjänen and Hjelt [30] further report that whereas the grading system of Arner et al. [1] is an appropriate method for malignancy grading, it is associated with low reproducibility.

More recently, Lanigan et al. [17] performed a study similar to Syrjänen and Hjelt [30] and assessed the degree of interobserver variation in four experienced pathologists' use of four commonly applied grading systems. Lanigan et al. [17] report that Syrjänen and Hjelt's grading system [30] is subject to less interobserver variability than other commonly used classifications, including the ones proposed by Skinner et al. [29] and Arner et al. [1].

Lanigan et al. [17] state that it is difficult to explain the basis of the differences observed between the various grading systems. One possibility in connection with the problem of low reproducibility in RCC grading systems could be due to the fact that RCCs may be composed of a variety of different cell types and may also exhibit a variety of histological patterns [20]. This problem of cell heterogeneity inside a tumour can be reduced by identifying the different cell populations constituting it. This identification can be carried out relatively easily by means of computer-assisted microscopy applied to Feulgen-stained nuclei. This technique enables cell nuclei and their chromatin patterns to be described by means of 30 quantitative variables, and their nuclear DNA content by 8 others [3,4,6,7,12,22,23,33].

As demonstrated in our recent study [7], this technique is thus fully compatible with the grading system proposed by Fuhrman et al. [9].

The data from our study show that discriminant analysis performed on quantitative morphonuclear variables did not enable the 105 cases to be clearly identified in terms of histopathological grades. However, we recently proved that more sophisticated techniques such as decision tree algorithms combined with additional processes (such as discretisation and stable features selection) were able to effect grade identification [7].

We obtained similar discriminant analysis results on the distinction that can be made between the different tumour invasion levels.

Better results were obtained by simple discriminant models distinguishing between metastatic status (i.e., the M^- or the M^+ status from the TNM system) and the bad or good survival status groups of patients. Hence, very significant p-levels were obtained across the group-centroids. However, the discriminant models generated on morphonuclear variables were not sufficiently efficient to avoid misclassifications between the two (metastasis or survival status-related) groups. These results were enhanced by taking into consideration the additional information contributed by the DNA ploidy-related variables. The selection of the %5C variable (i.e., the percentage of highly aneuploid cells with more than 5c DNA) to contribute prognostic information was in accordance with the recent results that we obtained on the prognostic ability of DNA ploidy level in RCCs [8].

In accordance with what we had previously observed in the case of RCC diagnosis [7], the decision tree technique combined with stable feature selection was more able to produce efficient prognostic models on the basis of morphonuclear and ploidy-related variables. Without including more variables than discriminant analysis, the decision tree model was characterised by a classification rate of 90% (5 misclassifications), i.e., 6% more than discriminant analysis. These different performances of the two classification techniques are due to the relative "flexibility" of the two algorithms in distinguishing between groups. Hence, discriminant analysis is a linear model which uses relatively strong hypotheses about data (Gaussian distributions and the equality of variance–covariance matrices between groups) while the decision tree is more "data-driven" (no hypothesis is formulated about the data) and approximates class boundaries by piecewise-linear surfaces which are defined by the monovariate tests generated by the algorithm. As shown in our results, different decision tree models can be generated, either with or without clinical variables such as the T invasion level. This is another advantage of decision tree algorithms which are able to treat quantitative and qualitative data simultaneously.

To validate our approach and illustrate a possible practical application we applied our models to an independent data set of 10 patients. The model generated on stable and discretized variables (Model 3) identified the 2 dead cases (bad status) included in the test series. In the context of a practical application, our results suggest that the results of Models 3 and 1 could be combined to associate a degree of confidence with the predictions. Hence, if the two models agree, the shared

prediction (good or bad) could be considered as expressing a good degree of confidence, while they disagree, the case could be considered as being suspect and prone to unfavorable development. Such a combination would enable patient 5 in Table 5 to be considered as "at risk" in spite of the low grade (II) and invasion level (T1) of her tumour.

In conclusion, it appears from the present work that quantitative cytometry-generated variables are able to provide a level of accuracy in terms of RCC diagnosis and/or prognosis which is as high as, if not higher than, the information contributed by conventional diagnostic features such as histopathological grade and TNM staging. These quantitative cytometry-generated variables also have the major advantage over conventional markers in that they are reproducible.

Acknowledgements

References

[1] O. Arner, C. Blanck and T. von Schreeb, Malignancy grading of renal adenocarcinoma, *Acta Chir. Scand.* **346**(Suppl.) (1965), 1–50.

[2] G. Brugal, C. Garbay, F. Giroud and D. Adhel, A double scanning microphotometer for image analysis: Hardware, software and biomedical applications, *J. Histochem. Cytochem.* **27** (1979), 144–152.

[3] C. Decaestecker, R. van Velthoven, M. Petein, T. Janssen, I. Salmon, J.L. Pasteels, P. Van Ham, C. Schulman and R. Kiss, The use of the decision tree technique and image cytometry to characterize aggressiveness in World Health Organization (WHO) grade II superficial transitional cell carcinoma of the bladder, *J. Pathol.* **178** (1996), 274–283.

[4] C. Decaestecker, I. Camby, N. Nagy, J. Brotchi, R. Kiss and I. Salmon, Improving morphology-based malignancy grading schemes in astrocytic tumors by means of computer-assisted techniques, *Brain Pathol.* **8** (1998), 29–38.

[5] C. Decaestecker, M. Remmelink, I. Salmon, I. Camby, D. Goldshmidt, M. Petein, P. Van Ham, J.L. Pasteels and R. Kiss, Methodological aspects of using Decision Trees to characterize leiomyomatous tumors, *Cytometry* **24** (1996), 83–92.

[6] C. DePrez, Y. de Launoit, R. Kiss, M. Petein, J.L. Pasteels, A. Verhest and R. van Velthoven, Computerized morpho-nuclear cell image analysis of malignant disease in bladder tissues, *J. Urol.* **143** (1990), 694–699.

[7] C. François, C. Decaestecker, M. Petein, P. Van Ham, J.-L. Pasteels, A. Danguy, I. Salmon, R. van Velthoven and R. Kiss, Classification strategies for the grading of renal cell carcinomas, based on nuclear morphometry and densitometry, *J. Pathol.* **183** (1997), 141–150.

[8] C. François, C. Decaestecker, I. Salmon, M. Petein, M. Remmelink, T. Janssen, A. Peltier, E. Wespes, C. Schulman, R. van Velthoven and R. Kiss, Stem cell line identification is of prognostic value for renal cell carcinomas, *Anal. Quant. Cytol. Histol.* (1998, in press).

[9] S.A. Fuhrman, L.C. Lasky and C. Limas, Prognostic significance of morphologic parameters in renal cell carcinoma, *Am. J. Surg. Pathol.* **6** (1982), 655–663.

[10] M.M. Galloway, Texture analysis using grayrun length, *Comput. Graph. Image Process.* **4** (1975), 172–179.

[11] F. Gao, U. Jutting, K. Rodenacker, P. Gais and P.Z. Lin, Relevance of chromatin features in the progression of esophageal epithelial sever dysplasia, *Anal. Cell. Pathol.* **13** (1997), 17–28.

[12] D. Goldschmidt, C. Decaestecker, J.V. Berthe, L. Gordower, M. Remmelink, A. Danguy, J.L. Pasteels, I. Salmon and R. Kiss, The contribution of computer-assisted methods for histopathological classification of adipose tumors, *Lab. Invest.* **75** (1996), 295–306.

[13] R.M. Haralick, T. Shanmugan and I. Dinstein, Textural features for image classification, *IEE Trans. Syst. Man Cybern.* **3** (1973), 610–621.

[14] UICC (International Union Against Cancer), *TNM Classification of Malignant Tumors*, P. Hermanek and L.H. Sobin, eds, 4th edn, 2nd rev., Springer, Berlin, 1992, pp. 148–150.

[15] T. Jorgensen, Y. Kanagasingam, O. Kaalhus, K.J. Tveter, M. Bryne, F. Skjorten, A. Berner and H.E. Danielsen, Prognostic factors in patients with metastatic (stage D2) prostate cancer: experience from Scandinavian Prostatic Cancer Group Study-2, *J. Urol.* **158** (1997), 164–170.

[16] R. Kiss, I. Salmon, I. Camby, S. Gras and J.L. Pasteels, Characterization of factors in routine laboratory protocols that significantly influence the Feulgen reaction, *J. Histochem. Cytochem.* **41** (1993), 935–945.

[17] D. Lanigan, R. Conroy, C. Barry-Walsh, B. Loftus, D. Royston and M.A. Leader, Comparative analysis of grading systems in renal adenocarcinoma, *Histopathology* **24** (1994), 473–476.

[18] B. Ljungberg, P. Larsson and R. Stenling, Flow cytometric deoxyribonucleic acid analysis in stage I renal cell carcinoma, *J. Urol.* **146** (1991), 697–699.

[19] G.J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, 1992.

[20] L.J. Medeiros, S.A. Michie, D.E. Johnson, R.A. Waenke and L.M. Weiss, An immunoperoxidase study of renal cell carcinomas. Correlation with nuclear grade, cell type, and histologic pattern, *Hum. Pathol.* **19** (1988), 980–987.

[21] W.M. Murphy, J.B. Beckwith and G.M. Farrow, Tumors of the kidney, in: *Tumors of the Kidney, Bladder, and Related Urinary Structures*, Atlas of Tumor Pathology, Armed Forces Institute of Pathology, Washington, 1994, pp. 92–178.

[22] M. Petein, Y. de Launoit, R. Kiss, C. DePrez, K. Crols, J.L. Pasteels, A. Verhest and R. van Velthoven, Characterization of the morphonuclear features and DNA ploidy of prostatic disease, *Prostate* **16** (1990), 199–208.

[23] M. Petein, P. Michel, R. van Velthoven, J.L. Pasteels, M. Brawer, J.R. Davis, R. Nagle and R. Kiss, Morphonuclear relationship between prostatic intraepithelial neoplasia and cancers as assessed by digital cell image analysis, *Am. J. Clin. Pathol.* **96** (1991), 628–634.

[24] J.R. Quinlan, Induction of Decision Trees, *Machine Learning* **1** (1986), 81–106.

[25] J.R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[26] G. Raviv, I. Leibovich, Y. Mor, D. Nass, O. Medalia, B. Goldwasser and O. Nativ, Localized renal cell carcinoma treated by radical nephrectomy. Influence of pathologic data and the importance of DNA ploidy pattern on disease outcome, *Cancer* **72** (1993), 2207–2212.

[27] J.L. Ruiz-Cerda, M. Hernandez, F. Gomis, C.D. Vera, B.F. Kimler, J.E. O'Connor and F. Jimenez-Cruz, Value of deoxyribonucleic acid ploidy and nuclear morphometry for prediction of disease progression in renal cell carcinoma, *J. Urol.* **155** (1996), 459–465.

[28] C.E. Shannon, A mathematical theory of communication, *Bell Sys. Tech. J.* **27** (1948), 379–423, 623–656.

[29] D.G. Skinner, R.B. Colvin, C.D. Vermillion, R.C. Pfister and W.F. Leadbetter, Diagnosis and management of renal cell carcinoma, *Cancer* **28** (1971), 1165–1177.

[30] K. Syrjänen and L. Hjelt, Grading of human renal adenocarcinoma, *Scand. J. Urol. Nephrol.* **12** (1978), 49–55.

[31] P. Tosi, P. Luzi, J.P.A. Baak, C. Miracco, R. Santopietro, C. Vindgni, F.M. Mattei, A. Acconcia and M.R. Massai, Nuclear morphology as an important prognostic factor in stage I renal cell carcinoma, *Cancer* **58** (1986), 2512–2518.

[32] H.G. van der Poel, P.F. Mulders, G.O. Oosterhof, H.E. Schaafsma, J.C. Hendriks, J.A. Schalken and F.M. Debruyne, Prognostic value of karyometric and clinical characteristics in renal cell carcinoma. Quantitative assessment of tumor heterogeneity, *Cancer* **72** (1993), 2667–2674.

[33] R. van Velthoven, M. Petein, A. Zlotta, W.J. Oosterlinck, A. Vandermeijden, C. Zandona, H. Roels, J.L. Pasteels, C. Schulman and R. Kiss, Computer-assisted chromatin texture characterization of Feulgen-stained nuclei in a series of 331 transitional bladder cell carcinomas, *J. Pathol.* **173** (1994), 235–242.

[34] L.L. Wheeless, R.A. Badalament, R.W. de Vere White, Y. Fradet and B. Tribukait, Consensus review of clinical utility of DNA cytometry in bladder cancer, *Cytometry* **14** (1993) 478–481.

[35] K. Yogesan, T. Jorgensen, F. Albregtsen, K.J. Tveter and H.E. Danielsen, Entropy-based texture analysis of chromatin structure in advanced prostate cancer, *Cytometry* **24** (1996), 268–276.