Three New Serum Markers for Prostate Cancer Detection Within a Percent Free PSA-Based Artificial Neural Network

Carsten Stephan,[1]* Chuanliang Xu,[1,2] David A. Brown,[3] Samuel N. Breit,[3] Anja Michael,[1] Terukazu Nakamura,[4,5] Eleftherios P. Diamandis,[4,6] Hellmuth Meyer,[1] Henning Cammann,[7] and Klaus Jung[1]

[1]*Department of Urology, University Hospital Charité Berlin, Berlin, Germany*
[2]*Department of Urology, Changhai Hospital, Second Military Medical University, Shanghai, China*
[3]*Centre for Immunology, St Vincent's Hospital, University of New South Wales, Sydney, Australia*
[4]*Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Canada*
[5]*Department of Urology, Kyoto Prefectural University of Medicine, Kamigyo-ku, Kyoto, Japan*
[6]*Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada*
[7]*Institute for Medical Biometry, University Hospital Charité Berlin, Berlin, Germany*

BACKGROUND. We aimed to evaluate the value of macrophage inhibitory cytokine 1 (MIC-1), human kallikrein 11 (hK11) migration inhibitor factor (MIF) in comparison to prostate-specific antigen (PSA) and %fPSA and also to develop a %fPSA-based ANN with the new input factors to determine whether these additional markers can further eliminate unnecessary prostate biopsies.

METHODS. Serum samples from 371 patients with prostate cancer (PCa, n = 135) or benign prostate hyperplasia (BPH, n = 236) within the PSA range 0.5–20 μg/L were analyzed for total PSA, free PSA, MIC-1, hK11, and MIF. 'Leave one out' ANN models with these variables and prostate volume were constructed and compared to logistic regression (LR) and all single parameters.

RESULTS. The discriminatory power of MIC-1, hK11, and MIF was less than that for PSA despite significant differences in BPH compared to PCa patients. At 90% and 95% sensitivity, the artificial neural networks (ANNs) were only significantly better than %fPSA if prostate volume was included.

CONCLUSIONS. ANNs with the novel input factors of MIC-1, MIF, and/or hK11 and additional use of prostate volume demonstrated significant advantage compared with %fPSA and tPSA and may lead to a reduction in unnecessary prostate biopsies. *Prostate*

KEY WORDS: **prostate cancer; prostate-specific antigen; macrophage inhibitory cytokine-1; macrophage migration inhibitory factor; human kallikrein 11; artificial neural network**

Abbreviations: ANN, artificial neural network; AUC, area under receiver operating characteristic curve; BPH, benign prostatic hyperplasia; DRE, digital rectal examination; hK11, human kallikrein 11; MIC-1, macrophage inhibitory cytokine-1; MIF, macrophage migration inhibitory factor; PCa, prostate cancer; PSA, prostate-specific antigen; %fPSA, percent free PSA; tPSA, total PSA.

Grant sponsor: Mildred-Scheel-Foundation; Grant number: 70-3295-ST1; Grant sponsor: National Health & Medical Research Council, Australia; Grant sponsor: New South Wales Health Research and Development infrastructure grant.

*Correspondence to: Carsten Stephan, Department of Urology, Charité Hospital Berlin, CCM, Schumannstrasse 20/21, D-10098 Berlin, Germany. E-mail: carsten.stephan@charite.de
Received 11 July 2005; Accepted 17 October 2005
DOI 10.1002/pros.20381
Published online in Wiley InterScience (www.interscience.wiley.com).

INTRODUCTION

Measurement of prostate-specific antigen (PSA) improves the early detection of prostate cancer (PCa), which is one of the most commonly diagnosed cancer in men [1]. Elevated PSA levels are not unique to patients with PCa and occur also in benign prostatic conditions such as benign prostatic hyperplasia (BPH) or chronic prostatitis [2]. The detection of molecular forms of PSA improves the specificity of PSA testing [3,4], in particular the use of percent free PSA (%fPSA) [5]. In order to reduce the large number of unnecessary prostate biopsies (up to 75%), new markers and other methods of individual risk estimation are urgently needed. This study focused on new markers where overexpression in PCa tissue has been proven and serum assays have already been developed [6].

Macrophage inhibitory cytokine-1 (MIC-1) is a member of the transforming growth factor-β (TGF-β) superfamily, originally cloned from macrophages using cDNA subtraction methodology [7]. MIC-1 is synthesized as a 60 kDa dimer and is cleaved from its propeptide by furin like proconvertases to release the mature 25 kDa protein [8]. However in tumors and tumor cell lines, MIC-1 is frequently secreted from cells in an unprocessed, propeptide containing form. This remains localized in tissues due to strong matrix binding mediated by its propeptide [8]. MIC-1 is also known by numerous synonyms including prostate-derived factor (PDF) [9]. Elevation of tumor and/or of serum MIC-1 levels has been documented in many cancers. MIC-1 mRNA has been shown to be upregulated in PCa compared to BPH and high compared to low Gleason score tumors [10]. Recently, using a sensitive sandwich ELISA assay [11], in a study of 1,000 patients, the use of serum MIC-1 in combination with PSA and %fPSA determination significantly improved the specificity of PSA for PCa detection [12].

Another new marker is the cytokine macrophage migration inhibitory factor (MIF). MIF was first described approximately 40 years ago, originally as a product of T-lymphocytes, which is known to activate macrophages and prevent their random migration. Subsequently, MIF was recognized to be ubiquitously expressed and have varied functions including regulation of inflammatory and immune response, induction of cell proliferation, angiogenesis, and inhibition of tumor suppressor genes [13]. Upregulation of MIF mRNA has been reported in PCa epithelial cells compared to normal [14]. Additionally, using a sandwich ELISA [15], serum concentrations of MIF are elevated in PCa patients, compared to BPH and control patients [14].

The third new marker, human kallikrein 11 (hK11) is a serine protease of the human kallikrein family, which has been recently expanded to include 15 members [16]. All kallikreins including PSA (hK3) share important similarities, including mapping at the same chromosomal locus (19q13.4) or significant homology at both the nucleotide and protein level [17]. Initially known as trypsin-like serine protease [18], tissue expression of the hK11 protein showed the highest levels in prostatic tissue extracts and seminal plasma [19]. A study of 150 PCa and BPH patients demonstrated significantly lower hK11 and hK11/tPSA ratio levels in PCa patients than in BPH patients suggesting potential diagnostic utility [20].

Other approaches to improve the PCa detection rate are the use of different models of logistic regression (LR) [21,22] and artificial neural networks (ANNs) using %fPSA together with PSA [23–26]. Especially ANNs have been increasingly used not only for detection but also for staging and prognosis of PCa [27–30]. A key advantage of ANNs compared with conventional methods like LR is their ability to resolve complex non-linear relations among variables, without the need for any prior assumptions about these relations. However, a true advantage of ANNs in comparison to LR is only partially seen [31,32]. In a comparison of 28 studies, the ANN tied with LR in 50% of all cases and in the 8 largest studies with a sample size >5,000, LR, and ANN tied in 7 cases [31].

The aims of this study were threefold. Initially, we wished to evaluate MIC-1, MIF, and hK11 in comparison to PSA and %fPSA for the diagnosis of PCa as single markers of disease. We then sought to develop a %fPSA-based ANN using MIC-1, MIF, and/or hK11 and evaluated it for diagnostic sensitivity and specificity. We also wished to determine the effect of additional clinical information like prostate volume on the diagnostic characteristics of our ANN.

MATERIALS AND METHODS

Study Population

The study population consisted of 371 men seen at the University Hospital Charité Berlin, Germany who were initially selected between 1998 and 2002 for studies on the diagnostic usefulness of ANNs in PCa diagnostics, and have been described in detail previously [25,33,34]. The study included 136 patients (mean age + SD, 62.7 + 5.9 years) with PCa stratified according to the TNM classification and WHO grading scale. From 79 PCa patients treated with radical prostatectomy, the pathological stages were: pT2a, pT2b, and pT2c (n = 61), pT3a and pT3b (n = 17), pT4 (n = 1). The grading of these operated patients were as following: G1a and 1b (n = 3), G2a and G2b (n = 44), G3a and G3b (n = 32). The diagnosis of two PCa

patients was obtained by TURP with stage T1a and T1b (both G2). The remaining 55 PCa patients were clinically staged as T1c (n = 10), T2a and T2b (n = 24), T3a and T3b (n = 21) and treated with radiation therapy, hormonal therapy, a combination of both or watchful waiting. The grading of biopsies from these 55 PCa patients were: G1a and 1b (n = 5), G2a and G2b (n = 41), G3a and G3b (n = 9).

The 235 patients with histopathology proven BPH (67.7 + 7.6 years) were untreated and had disease confirmed by transrectal ultrasound-guided sextant or octant prostate biopsy, transurethral resection or open adenomectomy. Since all patients were urologically referred the population of this study was not a screening population. The digital rectal examination (DRE) status was suspect in 53% of the PCa patients and 9.3% of the BPH patients. Previously, archived (at −80°C) and unthawed serum samples were used for serum marker determination. Selection criteria were availability (at least 1.5 ml in 4 vials) since samples were used also for other studies. All serum samples were taken before any diagnostic procedure or surgery of the prostate, and at least 3 weeks after DRE, prostate biopsy or transrectal ultrasound. The study was performed in accordance with the ethical standards of the Helsinki Declaration and was approved by the local ethical board of the hospital. Prostate volumes determined by transrectal ultrasound using the prolate ellipse formula (height × width × length × 0.52) were taken from the records and were available from 288 of 371 men.

METHODS

Total PSA (tPSA) and free PSA were assayed in Berlin, Germany using the IMMULITE PSA and IMMULITE Free PSA assays (DPC, Los Angeles, CA). Serum MIF was quantified in 200 μl duplicate samples using an enzyme linked immunosorbent assay (DuoSet, Cat. no. DY289; R&D Systems, Minneapolis, MN) as described by Meyer-Siegler et al. [15] and our group [35]. MIC-1 serum concentration was determined with 50 μl triplicates using an in house ELISA assay, previously described in detail [36,37]. Estimation of hK11 concentrations was performed in Canada by immunofluorometric serum assay [19]. All measurements were performed in duplicate as described [20].

Data were analyzed in the tPSA ranges 0.5–20 μg/L (n = 371, group I) and 2–10 μg/L (n = 226, group II) because the ''gray zone'' of tPSA is becoming more important especially for valuing of new markers for PCa detection. For those patients with prostate volume available, the tPSA ranges 0.5–20 μg/L (n = 288, group III) and 2–10 μg/L (n = 174, group IV) were also analyzed. Because of a possible loss of ~20% of all patients, we did not exclude the first two groups without available prostate volume. For all four groups, ANN models were constructed with the MATLAB Neural Network Toolbox (The Mathworks, Natick, MA). A feed forwarded back-propagation network was applied in which the input layer consisted of the variables: tPSA, %fPSA, age, MIF, MIC-1, hK11 and, if available, prostate volume. Various ANN models with the typical structure of three layers (input, hidden, and output layer) with inclusion of tPSA, %fPSA, age and one, two or all three new markers and also with inclusion of prostate volume (4, 5, 6 or 7 input neurons) and a various number of 2–5 neurons in the hidden layers were evaluated. The best performance of the ANNs were achieved with three neurons as hidden layers so that finally three neurons were used. Each ANN model was evaluated by the leave-one-out method, which has been previously described in detail [38]. Briefly, in a data set of N patients N separate times, the ANN is trained on all data except for one patient and a prediction is made for this patient. By using the leave-one-out method for the ANN there is no need for a validation within the training run. Each patient is ones the tested person, while all others were the training population. To get the best generalization of the ANN (i.e., to avoid overfitting), we used a routine that automatically sets the optimal performance function (Bayesian regularization) within the training run as also used by Finne et al. [23]. The evaluation of the model is made by computing the average error of all N predictions. The ANN output ranged from 0 (low PCa risk) to 1 (high PCa risk). The output values of the training run were then used to build the receiver-operating characteristic (ROC) curve. Thus, further testing is not necessary by using the leave one-out method.

Statistical Analysis

Statistical analyses were performed using the statistical software package, SPSS 11.5 for Windows (SPSS, Chicago, IL). Differences between groups were assessed with the Kruskal–Wallis test of variance or the Mann–Whitney *U*-test. Correlation between serum markers and clinical information were assessed using the Spearman rank correlation coefficient. LR without forward or backward stepwise analysis was estimated by using the same variables as for ANN analysis. The diagnostic validity of tPSA, %fPSA, MIF, MIC-1, and hK11 as well as the ANN and LR output values were evaluated by ROC curve analysis. The areas under the ROC curves (AUC) were compared using the GraphROC 2.1 for Windows [39] and MedCalc 8.0 (MedCalc Software, Mariakerke, Belgium). Significance was defined as $P < 0.05$ or $P < 0.01$ for correlation analyses.

RESULTS

Differences Between PCa and BPH Patients and Correlations

Median values for tPSA, %fPSA, MIC-1, MIF, hK11, age, prostate volume (if available), the LR output, and the ANN output for the four analyzed groups I–IV in the two tPSA ranges, with and without available prostate volume, are given in Table I. TPSA, %fPSA, age, and MIC-1 were significantly different between PCa and BPH in all four groups (hK11 in three groups) whereas MIF could only reach borderline significance level ($P = 0.045$) in group I. Prostate volume, the LR and ANN outputs were also significantly different ($P < 0.0001$) between PCa and BPH in all groups.

Whereas MIC-1 ($r_S = 0.49$ each, P always <0.01, if not indicated) and hK11 ($r_S = 0.26$ and 0.29) correlated only with age and to themselves ($r_S = 0.3$ each) in both analyzed tPSA ranges, MIF did not correlate at all with tPSA, %fPSA, MIC-1, hK11, age or prostate volume in both analyzed tPSA ranges and also if separately examining PCa and BPH patients. This is in agreement with our own data, which were published recently [35]. Age was correlated to prostate volume in both analyzed tPSA ranges ($r_S = 0.24$) and to %fPSA ($r_S = 0.28$ and 0.36). PSA correlated to prostate volume ($r_S = 0.25$) which was mostly an effect of BPH patients where the correlation was much stronger ($r_S = 0.56$) than in the PCa patients ($r_S = 0.19$, $P = 0.045$). The %fPSA correlated significantly to prostate volume ($r_S = 0.26$ and 0.45). Lastly, the only negative correlation was observed between PSA and %fPSA in both tPSA ranges ($r_S = -0.4$ and -0.25).

ANN and ROC Analyses

ROC comparisons were performed in all analyzed groups I–IV, for all available parameters, as well as for the respective ANN and LR models using their output values (Table II). Prostate volume as a single parameter reached an AUC of 0.67 (tPSA range 0.5–20 μg/L, group III) and an AUC of 0.795 (group IV). As seen in Table II, none of the new markers MIC-1, hK11 or MIF could reach the discriminatory power of tPSA nor %fPSA ($P < 0.0001$) with exception of MIC-1 compared to tPSA ($P = 0.42$) in group IV. If analyzing data for all patients in group I, MIC-1 and hK11 did not differ ($P = 0.9$), but were significantly better than MIF ($P = 0.022$ and 0.009). With the exception between LR and ANN ($P = 0.385$), all other comparisons reached the significance level of $P = 0.001$. Thus, the LR ($P = 0.001$) and ANN model ($P = 0.0003$) could significantly increase the performance of %fPSA.

For the LR and ANN calculation, different combinations of input factors were tested (e.g., exclusion of age, tPSA, %fPSA, volume, one, two or all new markers). In group I, the ANN model with all available input factors (tPSA, %fPSA, age, MIC-1, hK11, and MIF) reached an AUC of 0.843, whereas exclusion of MIC-1, hK11, MIF or even PSA gave similar AUCs of 0.858, 0.844, 0.842, and 0.855. The exclusion of %fPSA significantly deceased the AUC to 0.8, but the ANN model with exclusion of age performed best and reached an AUC of 0.862. However, the LR model with inclusion of all

TABLE I. Patients Within the Groups and Median Values for tPSA, %fPSA, MIC-1, hK11, MIF, Age, Volume, LR, and ANN Outputs

Parameter	BPH				PCa			
tPSA range group	0.5–20 μg/L, group I	2–10 μg/L, group II	0.5–20 μg/L, group III	2–10 μg/L, group IV	0.5–20 μg/L, group I	2–10 μg/L, group II	0.5–20 μg/L, group III	2–10 μg/L, group IV
number	235	143	173	102	136	83	115	72
tPSA (μg/L)	4.0*	4.7*	4.1*	5.25**	8.05	5.9	7.8	5.85
%fPSA (%)	16.9*	16.0*	16.9*	16.05*	8.4	9.0	8.3	8.5
MIC-1 (ng/L)	971**	905**	932**	915**	798	781	806	747
hK11 (ng/L)	0.167**	0.174**	0.162**	0.157	0.141	0.139	0.137	0.143
MIF (μg/L)	0.204**	0.212	0.187	0.199	0.159	0.158	0.152	0.156
Age in years	68*	67*	67*	67*	63	63	63	62
Gland volume	n.a.	n.a.	43	49	n.a.	n.a.	30	28
LR output	0.16*	0.173*	0.156*	0.132*	0.625	0.5	0.713	0.617
ANN output	0.183*	0.151*	0.056*	0.121*	0.686	0.582	0.859	0.714

n.a., not available.
*$P < 0.0001$, when compared to PCa.
**$P < 0.05$, when compared to PCa.

TABLE II. Areas Under ROC Curves ± Standard Error for tPSA, %fPSA, MIC-1, hKII, MIF, LR, and ANN in the Respective tPSA Ranges, With and Without Prostate Volume (Groups I–IV)

	Areas under the curves (AUCs) for the respective parameter						
tPSA range, group	tPSA	%fPSA	MIC-1	hK11	MIF	LR	ANN
0.5–20 μg/L, group I	0.769 ± 0.025*	0.807 ± 0.023**	0.600 ± 0.03*	0.602 ± 0.03 *	0.562 ± 0.031*	0.853 ± 0.02	0.862 ± 0.02
2-10 μg/L, group II	0.660 ± 0.037*	0.788 ± 0.03**	0.593 ± 0.04*	0.614 ± 0.038*	0.556 ± 0.041*	0.827 ± 0.027	0.837 ± 0.028
0.5-20 μg/L, group III	0.636 ± 0.03*	0.691 ± 0.028*	0.55 ± 0.031*	0.557 ± 0.031*	0.537 ± 0.032*	0.872 ± 0.021**	0.907 ± 0.02
2-10 μg/L, group IV	0.623 ± 0.043*	0.799 ± 0.033*	0.603 ± 0.044*	0.579 ± 0.044*	0.551 ± 0.045*	0.871 ± 0.026	0.883 ± 0.027

*$P < 0.0001$, when compared with the ANN.
**$P < 0.01$, when compared with the ANN.

parameters had the largest AUC of 0.853. The wald coefficients in the LR model revealed the largest impact on the power of the LR model for %fPSA (wald coefficient: 30.2) and tPSA (28.6) followed by age (14.7), hK11 (2.3), MIF (1.5) whereas MIC-1 had a value of 0. To simplify the comparisons, only those ANN and LR models were taken and compared which achieved the largest AUC. For group II, the ANN model with exclusion of age and the LR model with all input parameters performed best, and in group III the ANN and the LR model with exclusion of MIC-1 performed best. The wald coefficients in the LR model for group III revealed the largest values for tPSA (24.9), %fPSA (18.1), and prostate volume (16.5) followed by age (9.0), hK11 (2.0), MIF (1.7), and MIC-1 (0.24). The largest AUCs in group IV were obtained in the ANN model with exclusion of age and in the LR model with exclusion of MIC-1. Data for group II (Table II) indicate again that none of the new markers reached the AUC of tPSA and that MIC-1, hK11, and MIF showed no difference to each other but significance between MIC-1 and MIF ($P = 0.006$). The %fPSA performed significantly better than tPSA ($P < 0.0001$) but again the %fPSA had a significantly smaller AUC in comparison to LR ($P = 0.003$) and ANN ($P = 0.004$), whereas LR and ANN performed equally to each other ($P = 0.92$). Analyses in all patients with available prostate volume (group III) showed that MIC-1, hK11, and MIF did not differ from each other (P from 0.2 to 0.63) and that %fPSA was significantly better than tPSA. Interestingly, the ANN model could reach significance to the LR model ($P = 0.0013$) and reached an AUC of >0.9 (see Fig 1). In group IV, similar results were calculated compared to group III. Again the three new markers did not differ (P from 0.055 to 0.43) and the LR and ANN model performed significantly better than %fPSA but were equal to each other ($P = 0.86$).

Cutoff Analyses

The specificities for tPSA, %fPSA, MIC-1, hK11, MIF, the LR, and the ANN models for the given sensitivities at 90% and 95% are summarized in Table III. In all comparisons, the respective LR and ANN models were significantly better than MIC-1, hK11 or MIF ($P < 0.0001$). Compared to tPSA, the ANN models always reached significance (P from 0.02 to <0.0001). However, compared to %fPSA, neither the ANN (P from 0.23 to 1) nor the LR models (P from 0.15 to 0.94) could demonstrate an improvement in specificity at 90% sensitivity, and at 95% sensitivity, if analyzing the groups I and II without the inclusion of prostate volume. When analyzing the specificities at 90% sensitivity in group III (Table III), the ANN model was significantly better than all other parameters

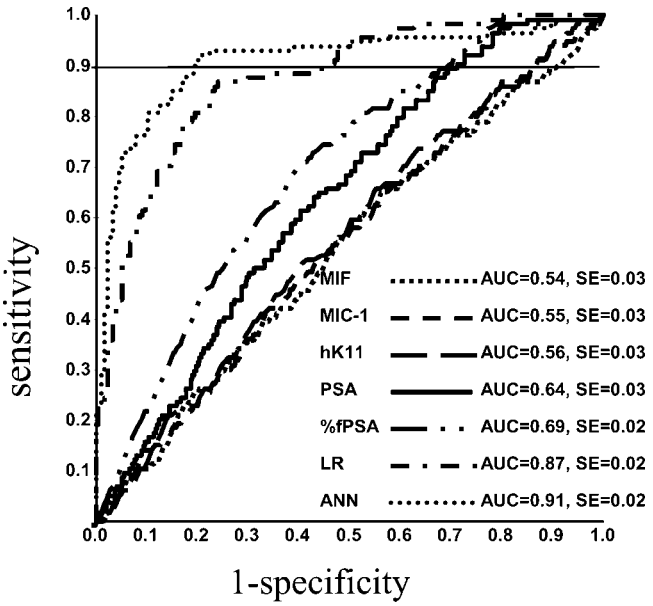

Fig. 1. Receiver-operating characteristic (ROC curve analysis) of tPSA, %fPSA, MIC-1, hK11, MIF, LR, and ANN output for those 288 patients with prostate volume available in the tPSA range 0.5–20 μg/L (group III), SE: standard error, line at 90% sensitivity.

including the LR model with a specificity of 80% compared to 55% (LR) and 30% (%fPSA) or 29% for tPSA. As one of the key results of this study is also shown in Figure 1. However, at 95% sensitivity the difference in specificity between the ANN model (40%) and tPSA (22%) or %fPSA (24%) was smaller but still significant. Alternatively, the LR model showed a non-significant, but higher specificity (48%, $P = 0.48$) compared to the ANN model. Results from group IV revealed similar significance levels compared to group III. Again, at 90% sensitivity the ANN model performed significantly better than all other variables (P from 0.045 to <0.0001). However, at 95% sensitivity the ANN performed only equally to %fPSA ($P = 0.45$) and moreover, the LR model was significantly better than the ANN model ($P = 0.0009$) and also better than %fPSA ($P = 0.044$). Data at 90% and 95% specificity are not included because only high sensitivities are of interest in this clinical situation.

DISCUSSION

PSA is currently viewed as the most useful tumor markers for PCa detection, staging, and prognosis [1]. For early detection and improvement of specificity compared to tPSA, %fPSA has demonstrated its value [2]. Nevertheless, neither PSA nor %fPSA can accurately predict the aggressiveness of PCa or the rate of postoperative biochemical failure. New improved

TABLE III. Specificities at 90% and 95% Sensitivity, Respectively for tPSA, %fPSA, MIC-1, hKII, MIF, LR, and the ANN

		Specificities (%) and 95% confidence intervals in parentheses						
tPSA, group	Sensitivity cutoffs (%)	tPSA	%fPSA	MIC-1	hK11	MIF	LR	ANN
0.5–20 μg/L	90	47 (41.3–52.4)*	45.1 (39.6–50.7)	17.9 (13.9–22.5)***	18.3 (14.3–23)***	8.51 (5.73–12.2)***	54.5 (48.9–60)	52.3 (46.8–57.9)
Group I	95	34.5 (29.3–40)**	38.7 (33.4–44.3)	14 (10.5–18.4)***	12.8 (9.36–17)***	5.11 (3–8.2)***	37.4 (32.2–43)	38.7 (33.4–44.3)
2–10 μg/L	90	28 (21.9–34.8)*	42 (35–49.2)	16.1 (11.3–22.1)***	21 (15.6–27.4)***	8.39 (4.9–13.3)***	48.9 (41.8–56.1)	46.2 (39.1–53.4)
Group II	95	13.3 (8.9–19)***	38.5 (31.7–45.7)	13.3 (8.9–19)***	12.6 (8.3–18.2)***	3.5 (1.41–7.33)***	37.1 (30.3–44.2)	39.2 (32.3–46.4)
0.5–20 μg/L	90	28.8 (24.4–33.5)***	29.9 (25.4–34.6)***	13.5 (10.4–17.3)***	13.5 (10.4–17.3)***	8.68 (6.1–12)***	55.2 (48.6–61.6)***	80.5 (74.8–85.2)
Group III	95	21.9 (17.9–26.3)***	24.3 (20.2–28.8)***	9.03 (6.4–12.3)***	8 (5.54–11.2)***	4.51 (2.71–7.1)***	47.7 (41.2–54.2)	40.2 (34–46.7)
2–10 μg/L	90	21.6 (15.1–29.4)***	42.2 (33.9–50.8)*	14.7 (9.3–21.8)***	13.7 (8.5–20.7)***	6.86 (3.3–12.6)***	61.8 (53.2–69.8)**	67.6 (59.2–75.3)
Group IV	95	12.7 (7.75–19.6)***	40.2 (32–48.8)	10.8 (6.2–17.3)***	8.82 (4.72–15)***	3.92 (1.37–8.9)***	58.8 (50.2–67)†	46.1 (37.6–54.7)

*$P < 0.01$, when compared with the ANN.
**$P < 0.05$, when compared with the ANN.
***$P < 0.0001$, when compared with the ANN.
†Significantly better than the ANN with $P < 0.05$.

biomarkers are urgently needed especially for the identification of Gleason 4/5 tumors [40].

The results from this study indicate that serum measurements of MIC-1, hK11 or MIF, used as individual markers, could not improve PCa detection compared with tPSA and %fPSA. The %fPSA-based ANNs we developed using MIC-1, MIF, and/or hK11 performed significantly better compared to %fPSA and tPSA. Additionally, inclusion of prostate volume as a clinical parameter further improved the diagnostic capacity of the ANN model (80% specificity) compared to %fPSA (30% specificity) at the 90% sensitivity cutoff, but this improvement was significantly diminished at the 95% sensitivity cut-off. Consequently, %fPSA-based ANNs with MIC-1, hK11, MIF, and prostate volume may be helpful to reduce unnecessary biopsies. Thus, approaches like ANN and LR models with their capability to significantly reduce prostate biopsies are promising tools to overcome the dilemma of over-diagnosis and subsequently over-treatment of PCa in the PSA era because it has not been proven that PCa mortality is significantly decreased due to screening with PSA alone.

The reason that some of the individual markers may perform poorly when used as single markers is probably a reflection of the complex biology of tumor cytokine production. In the case of MIC-1, our previous results indicate that MIC-1 serum concentrations were significantly lower in PCa patients [12]. A likely reason for this unusual finding is that incompletely processed MIC-1, containing the propeptide region is tightly bound to the extracellular matrix of PCa and cannot easily diffuse into the circulation [8]. Additionally, as the Gleason sum increases, extracellular matrix binding of MIC-1 generally decreases with a concomitant rise in serum MIC-1 levels. Consequently, using a single cut-off, MIC-1 would perform poorly as a single marker, as higher grade tumors would tend to be missed. However, in combination with markers that behave in a more linear unidirectional fashion such as %fPSA, sensitivity and specificity may well be enhanced. For example, in BPH and PCa patients with a marginally depressed %fPSA, consideration of the MIC-1 level might more accurately classify them into benign versus malignant disease. Conversely, tumors with a normal %fPSA might be distinguished from BPH on the basis of a depressed MIC-1. It is likely that similarly complex regulatory factors may apply also to MIF.

Like MIC-1, the cytokine MIF was also been found to be overexpressed in PCa epithelial cells compared to normal prostate cells [14]. Additionally, MIF serum concentrations are elevated in the sera of PCa patients compared with BPH and controls [14]. In the study by Meyer-Siegler et al. [14] MIF strongly correlated to tPSA ($r_S = 0.61$) in 509 analyzed patients, but unfortunately a histological diagnosis of PCa was only available in 152 patients making interpretation of this data difficult. In the current study, MIF did not correlate to any other parameter including tPSA ($r_S = -0.04$, $P = 0.43$) and furthermore, MIF values were decreased and not significantly higher. These results are in agreement with a recent study by Michael et al. [35] and seemingly at odds with the small study by Meyer-Siegler et al. [14]. This apparent contradiction may be due to differing stages of disease of patient enrolled in the individual studies. Despite MIF showing the weakest performance of all the new markers with a smaller AUC and lower specificities at the given cutoffs compared to tPSA and %fPSA, this parameter had a substantial value within an ANN. It is noteworthy that a parameter without correlation to tPSA or to %fPSA might be more suitable for an ANN compared to another marker with a correlation to tPSA, like hK2, similar to the situation encountered with MIC-1.

As one of the 15 kallikreins, hK11 was expected to confirm the recently obtained and promising results from a study by Nakamura et al. [20]. Our recent data confirmed this publication [20] by measuring significant lower results in PCa patients. However, the ratio hK11/tPSA (AUC: 0.753 group I and 0.659 group II) could only reach the AUC of PSA but not %fPSA (see Table II). This is in contrast to data obtained by Nakamura et al. [20] where the authors found a significantly larger AUC for hK11/tPSA and %fPSA (each 0.83) compared to tPSA (0.69). However, as seen with the other markers MIC-1 and MIF, the hK11 values contributed to the ANN which was significantly better than %fPSA. The hK11 assay has also been reported to be a useful prognostic factor in ovarian cancer [41].

Overall it is clear that ANNs are superior to %fPSA and the addition of MIC-1, MIF, and hK11 further enhance this diagnostic superiority. In all four ROC comparisons, the ANN performed significantly better than %fPSA. This is reflected in the increase in the AUC from 0.05 (group I and II) to 0.08 (group IV) and 0.22 (group III) when comparing %fPSA with the ANN output. The greatest improvement in diagnostic capacity occurred when prostate volume was included in the ANN for analysis of the patient group at the tPSA range 0.5–20 μg/L (Table II). This improvement was beyond the individual discriminatory capacity of prostate volume, which had an AUC of only 0.67 and is likely to be due to its capacity to further classify whether serum marker production is normal or abnormal. For example, with MIC-1, while MIC-1 serum level is significantly lower in localized PCa, the quotient of serum MIC-1 per unit of prostate volume is significantly higher than in BPH [12]. Some of us have

previously shown that ANNs significantly decrease unnecessary biopsies with the inclusion of five ANN input factors (tPSA, %fPSA, age, prostate volume, and the status of the DRE) in a multi-center study of 1,188 patients [33]. However, DRE, while being an important urological investigation is notoriously subjective depending on the experience of the clinician. Consequently, the exclusion of DRE and inclusion of the relatively objective clinical parameter prostate volume with significant improved diagnostic accuracy is a notable finding.

It is clear that there are limitations in ANNs when examining the cutoffs of 90% and 95% sensitivity in a number of groups despite at 90% sensitivity the specificity for the ANN model reached 80% (group III) and 67.6% (group IV) whereas %fPSA had a specificity of only 30% and 42%. This relatively large increase in specificity by 26% to 50% is comparable with own data [33]. The comparison of ANN and LR regarding AUC (Table II) and cutoff analysis at 90% and 95% sensitivity (Table III) showed in three of four (AUC) and five of eight (cutoff) calculations no differences between ANN and LR models indicating a relatively equal performance of both methods which is in concordance to other comparisons [31]. On the other side, the ANN and LR were significantly better than the other in the remaining four comparisons. Thus, ANN and LR methods showed both advantage and disadvantage to each other.

On the other hand, there was no advantage for using the ANN or the LR models compared to %fPSA in five of eight cutoff comparisons (Table III). However, a number of ANN and LR models provided significant advantage over the use of %fPSA alone showing that all three new markers can contribute to the improved performance of ANN. These results indicate that the improved performance of an ANN is likely to be achieved if the new input factors are not or weakly correlated to tPSA or %fPSA. Using these criteria, MIC-1, hK11, and especially MIF were suitable candidates to improve the ANN diagnostic accuracy, even in the face of poor performance as single markers when compared to the benchmarks of tPSA and %fPSA.

In conclusion, these data show that the ANNs with the new input factors MIC-1, MIF, hK11, and especially the clinical parameter prostate volume were significantly better compared with %fPSA and may be helpful in reducing unnecessary prostate biopsies.

ACKNOWLEDGMENTS

We gratefully thank Paul E.C. Sibley for the helpful corrections and comments.

REFERENCES

1. Polascik TJ, Oesterling JE, Partin AW. Prostate specific antigen: A decade of discovery-what we have learned and where we are going. J Urol 1999;162:293–306.
2. Stephan C, Jung K, Lein M, Sinha P, Schnorr D, Loening SA. Molecular forms of prostate-specific antigen and human kallikrein 2 as promising tools for early diagnosis of prostate cancer. Cancer Epidemiol Biomarkers Prev 2000;9: 1133–1147.
3. Lilja H, Christensson A, Dahlen U, Matikainen MT, Nilsson O, Pettersson K, Lövgren T. Prostate-specific antigen in serum occurs predominantly in complex with alpha 1-antichymotrypsin. Clin Chem 1991;37:1618–1625.
4. Stenman UH, Leinonen J, Alfthan H, Rannikko S, Tuhkanen K, Alfthan O. A complex between prostate-specific antigen and alpha 1-antichymotrypsin is the major form of prostate-specific antigen in serum of patients with prostatic cancer: Assay of the complex improves clinical sensitivity for cancer. Cancer Res 1991; 51:222–226.
5. Catalona WJ, Partin AW, Slawin KM, Brawer MK, Flanigan RC, Patel A, Richie JP, deKernion JB, Walsh PC, Scardino PT, Lange PH, Subong EN, Parson RE, Gasior GH, Loveland KG, Southwick PC. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: A prospective multicenter clinical trial. JAMA 1998; 279:1542–1547.
6. Sullivan PM, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93:1054–1061.
7. Bootcov MR, Bauskin AR, Valenzuela SM, Moore AG, Bansal M, He XY, Zhang HP, Donnellan M, Mahler S, Pryor K, Walsh BJ, Nicholson RC, Fairlie WD, Por SB, Robbins JM, Breit SN. MIC-1, a novel macrophage inhibitory cytokine, is a divergent member of the TGF-beta superfamily. Proc Natl Acad Sci USA 1997; 94:11514–11519.
8. Bauskin AR, Brown DA, Junankar S, Rasiah KK, Eggleton S, Hunter M, Liu T, Smith D, Kuffner T, Pankhurst GJ, Johnen H, Russell PJ, Barret W, Stricker PD, Grygiel JJ, Kench JG, REFAU>Henshall SM, Sutherland RL, Breit SN. The propeptide mediates formation of stromal stores of PROMIC-1: Role in determining prostate cancer outcome. Cancer Res 2005;65:2330–2336.
9. Paralkar VM, Vail AL, Grasser WA, Brown TA, Xu H, Vukicevic S, Ke HZ, Qi H, Owen TA, Thompson DD. Cloning and characterization of a novel member of the transforming growth factor-beta/bone morphogenetic protein family. J Biol Chem 1998;273:13760–13767.
10. Nakamura T, Scorilas A, Stephan C, Yousef GM, Kristiansen G, Jung K, Diamandis EP. Quantitative analysis of macrophage inhibitory cytokine-1 (MIC-1) gene expression in human prostatic tissues. Br J Cancer 2003;88:1101–1104.
11. Moore AG, Brown DA, Fairlie WD, Bauskin AR, Brown PK, Munier ML, Russell PK, Salamonsen LA, Wallace EM, Breit SN. The transforming growth factor-ss superfamily cytokine macrophage inhibitory cytokine-1 is present in high concentrations in the serum of pregnant women. J Clin Endocrinol Metab 2000; 85:4781–4788.
12. Brown DA, Stephan C, Ward RL, Law M, Hunter M, Bauskin AR, Amin J, Jung K, Diamandis EP, Hampton GM, Russell PJ, Giles GG, Breit SN. Serum macrophage inhibitory cytokine-1 (MIC-1) levels for the diagnosis and tumor grading of prostate cancer. Clin Cancer Res 2005; in press.

13. Mitchell RA, Bucala R. Tumor growth-promoting properties of macrophage migration inhibitory factor (MIF). Semin Cancer Biol 2000;10:359–366.
14. Meyer-Siegler KL, Bellino MA, Tannenbaum M. Macrophage migration inhibitory factor evaluation compared with prostate specific antigen as a biomarker in patients with prostate carcinoma. Cancer 2002;94:1449–1456.
15. Meyer-Siegler K, Fattor RA, Hudson PB. Expression of macrophage migration inhibitory factor in the human prostate. Diagn Mol Pathol 1998;7:44–50.
16. Diamandis EP, Yousef GM, Clements J, Ashworth LK, Yoshida S, Egelrud T, Nelson PS, Shiosaka S, Little S, Lilja H, Stenman UH, Rittenhouse HG, Wain H. New nomenclature for the human tissue kallikrein gene family. Clin Chem 2000;46:1855–1858.
17. Yousef GM, Diamandis EP. The new human tissue kallikrein gene family: structure, function, and association to disease. Endocr Rev 2001;22:184–204.
18. Yoshida S, Taniguchi M, Hirata A, Shiosaka S. Sequence analysis and expression of human neuropsin cDNA and gene. Gene 1998;213:9–16.
19. Diamandis EP, Okui A, Mitsui S, Luo LY, Soosaipillai A, Grass L, Nakamura T, Howarth DJ, Yamaguchi N. Human Kallikrein 11: A new biomarker of prostate and ovarian carcinoma. Cancer Res 2002;62:295–300.
20. Nakamura T, Scorilas A, Stephan C, Jung K, Soosaipillai AR, Diamandis EP. The usefulness of serum human kallikrein 11 for discriminating between prostate cancer and benign prostatic hyperplasia. Cancer Res 2003;63:6543–6546.
21. Carlson GD, Calvanese CB, Partin AW. An algorithm combining age, total prostate-specific antigen (PSA), and percent free PSA to predict prostate cancer: Results on 4298 cases. Urology 1998;52: 455–461.
22. Virtanen A, Gomari M, Kranse R, Stenman UH. Estimation of prostate cancer probability by logistic regression: Free and total prostate-specific antigen, digital rectal examination, and heredity are significant variables. Clin Chem 1999;45:987–994.
23. Finne P, Finne R, Auvinen A, Juusela H, Aro J, Maattanen L, Hakama M, Rannikko S, Tammela TL, Stenman U. Predicting the outcome of prostate biopsy in screen-positive men by a multilayer perceptron network. Urology 2000;56:418–422.
24. Babaian RJ, Fritsche H, Ayala A, Bhadkamkar V, Johnston DA, Naccarato W, Zhang Z. Performance of a neural network in detecting prostate cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/mL. Urology 2000;56:1000–1006.
25. Stephan C, Jung K, Cammann H, Vogel B, Brux B, Kristiansen G, Rudolph B, Hauptmann S, Lein M, Schnorr D, Sinha P, Loening SA. An artificial neural network considerably improves the diagnostic power of percent free prostate-specific antigen in prostate cancer diagnosis: Results of a 5-year investigation. Int J Cancer 2002;99:466–473.
26. Djavan B, Remzi M, Zlotta A, Seitz C, Snow P, Marberger M. Novel artificial neural network for early detection of prostate cancer. J Clin Oncol 2002;20:921–929.
27. Haese A, Chaudhari M, Miller MC, Epstein JI, Huland H, Palisaar J, Graefen M, Hammerer P, Poole EC, O'Dowd GJ, Partin AW, Veltri RW. Quantitative biopsy pathology for the prediction of pathologically organ-confined prostate carcinoma: A multi-institutional validation study. Cancer 2003;97:969–978.
28. Reckwitz T, Potter SR, Snow PB, Zhang Z, Veltri RW, Partin AW. Artificial neural networks in urology: Update 2000. Prostate Cancer Prostatic Dis 1999;2:222–226.
29. Stephan C, Cammann H, Jung K. Viewpoint: Artificial neural networks: Has the time come for their use in prostate cancer patients? Nature Clin Pract Urol 2005;2:262–263.
30. Veltri RW, Chaudhari M, Miller MC, Poole EC, O'Dowd GJ, Partin AW. Comparison of logistic regression and neural net modeling for prediction of prostate cancer pathologic stage. Clin Chem 2002;48:1828–1834.
31. Sargent DJ. Comparison of artificial neural networks with other statistical approaches. Cancer 2001;91:1636–1642.
32. Schwarzer G, Schumacher M. Artificial neural networks for diagnosis and prognosis in prostate cancer. Semin Urol Oncol 2002;20:89–95.
33. Stephan C, Cammann H, Semjonow A, Diamandis EP, Wymenga LFA, Lein M, Sinha P, Loening SA, Jung K. Multicenter evaluation of an artificial neural network to increase prostate cancer detection rate and reduce unnecessary biopsies. Clin Chem 2002;48:1279–1287.
34. Stephan C, Vogel B, Cammann H, Lein M, Klevecka V, Sinha P, Kristiansen G, Schnorr D, Jung K, Loening SA. Nutzung von artifiziellen neuronalen Netzwerken zur Risikoabschätzung eines Prostatakarzinoms: Biopsieindikationen im PSA-Bereich 2–20 μg/L. Urologe A 2003;42:1221–1229.
35. Michael A, Stephan C, Kristiansen G, Burckhardt M, Loening SA, Schnorr D, Jung K. Diagnostic validity of macrophage migration inhibitory factor in serum of patients with prostate cancer: A re-evaluation. Prostate 2005;62:34–39.
36. Brown DA, Bauskin AR, Fairlie WD, Smith MD, Liu T, Xu N, Breit SN. Antibody-based approach to high-volume genotyping for MIC-1 polymorphism. Biotechniques 2002;33:118–120, 122, 124.
37. Brown DA, Breit SN, Buring J, Fairlie WD, Bauskin AR, Liu T, Ridker PM. Concentration in plasma of macrophage inhibitory cytokine-1 and risk of cardiovascular events in women: A nested case-control study. Lancet 2002;359:2159–2163.
38. Stephan C, Jung K, Soosaipillai A, Yousef GM, Cammann H, Meyer H, Xu C, Diamandis EP. Clinical usefulness of human glandular kallikrein 2 (hK2) within a neural network for prostate cancer detection. BJU Int 2005;96:521–527.
39. Kairisto V, Poola A. Software for illustrative presentation of basic clinical characteristics of laboratory tests—GraphROC for Windows. Scand J Clin Lab Invest 1995;55(Suppl 222):43–60.
40. Stamey TA. Preoperative serum prostate-specific antigen (PSA) below 10 μg/L predicts neither the presence of prostate cancer nor the rate of postoperative PSA failure. Clin Chem 2001;47: 631–634.
41. Diamandis EP, Borgono CA, Scorilas A, Harbeck N, Dorn J, Schmitt M. Human kallikrein 11: An indicator of favorable prognosis in ovarian cancer patients. Clin Biochem 2004;37:823–829.