Comparative Assessment of Urinary Prostate Cancer Antigen 3 and *TMPRSS2:ERG* Gene Fusion with the Serum [−2]Proprostate-Specific Antigen–Based Prostate Health Index for Detection of Prostate Cancer

Carsten Stephan,[1,2*] Klaus Jung,[1,2] Axel Semjonow,[3] Kai Schulze-Forster,[4] Henning Cammann,[5] Xinhai Hu,[1] Hellmuth-A. Meyer,[1] Martin Bögemann,[3] Kurt Miller,[1] and Frank Friedersdorff[1]

BACKGROUND: We compared urinary prostate cancer antigen 3 (PCA3), transmembrane protease, serine 2 (*TMPRSS2*):v-ets erythroblastosis virus E26 oncogene homolog (avian) (*ERG*) gene fusion (T2:ERG), and the serum [−2]proprostate-specific antigen ([−2]proPSA)-based prostate health index (Phi) for predicting biopsy outcome.

METHODS: Serum samples and first-catch urine samples were collected after digital rectal examination (DRE) from consented outpatients with PSA 0.5–20 μg/L who were scheduled for prostate biopsy. The PCA3 score (PROGENSA PCA3, Hologic Gen-Probe) and T2:ERG score (Hologic Gen-Probe) were determined. Measurements of serum PSA, free PSA, and [−2]proPSA (Beckman Coulter) were performed, and the percentages of free PSA (%fPSA) and Phi ([−2]proPSA/ fPSA × √PSA) were determined.

RESULTS: Of 246 enrolled men, prostate cancer (PCa) was diagnosed in 110 (45%) and there was no evidence of malignancy (NEM) in 136 (55%). A first set of biopsies was performed in 136 (55%) of all men, and 110 (45%) had ≥1 repeat biopsies. PCA3, Phi, and T2:ERG differed significantly between men with PCa and NEM, and these markers showed the largest areas under the ROC curve (AUCs) (0.74, 0.68, and 0.63, respectively). PCA3 had the largest AUC of all parameters, albeit not statistically different from Phi. Phi showed somewhat lower specificities than PCA3 at 90% sensitivity. Combination of both markers enhanced diagnostic power with modest AUC gains of 0.01–0.04. Although PCA3 had the highest AUC in the repeat-biopsy cohort, the highest AUC for Phi was observed in DRE-negative patients with PSA in the 2–10 μg/L range.

CONCLUSIONS: PCA3 and Phi were superior to the other evaluated parameters but their combination gave only moderate enhancements in diagnostic accuracy for PCa at first or repeat prostate biopsy.

© 2012 American Association for Clinical Chemistry

The detection of prostate cancer (PCa)[6] is closely related to prostate-specific antigen (PSA) but is limited by the low specificity of this biomarker (*1*). Additional parameters such as the percentage of free PSA (%fPSA) (*2*) and other PSA subforms (*3*) or clinical parameters like prostate volume, age, and status of digital rectal examination (DRE) have demonstrated added value within multivariate models (*4*). However, biomarkers are still needed that can be used to detect aggressive PCa or that confer additional diagnostic value in patients with persistently increased PSA values and several previous negative biopsies. To address these needs, researchers have been focused on fPSA subforms [e.g., benign PSA, intact PSA, proprostate-specific antigen (proPSA)] (*5*). Of biomarkers in the latter category, [−2]proPSA has yielded promising results (*6*), and its automated assay has been available since 2008 (*7, 8*). When [−2]proPSA was combined with PSA and %fPSA to compute the prostate health index (Phi) using the formula: Phi = [−2]proPSA/fPSA × √PSA,

PCa detection was improved compared with the use of %fPSA or [−2]proPSA alone *(9, 10)*. In 2012, [−2]proPSA was approved by the US Food and Drug Administration (FDA) to be used for initial biopsy decisions in men with PSA concentrations in the range of 4–10 μg/L and negative DRE.

The urine marker prostate cancer antigen 3 (PCA3) has been shown to increase the specificity of PSA, especially for patients with repeat biopsies *(11–13)*. This finding has resulted in the recent FDA approval in 2012 of the use of PCA3 for men scheduled for repeat biopsy. A correlation between PCA3 and biopsy outcome has been demonstrated *(11, 13)*, but within multivariate models PCA3 increased accuracy only modestly *(13, 14)*.

The detection of gene fusions of transmembrane protease, serine 2 (*TMPRSS2*)[7] to v-ets erythroblastosis virus E26 oncogene homolog (avian) (*ERG*) in PCa tissue *(15)* and its prevalence in approximately 50% of all PCa patients *(16, 17)* has been demonstrated. Based on these findings, a urinary assay for *TMPRSS2:ERG* fusion (T2:ERG) is currently in development using the same platform and technology as the PCA3 assay *(18)*. T2:ERG showed its clinical value for PCa detection in small cohorts *(18, 19)* and in a more recent multicenter study *(20)*.

Reviewed data have shown increased diagnostic accuracy with the use of PCA3 or T2:ERG *(21)*. Investigations with data on Phi and PCA3 *(22)* or PCA3 and T2:ERG have been performed *(20, 23)*, but a comparison of all 3 biomarkers in the same population has not been conducted so far. Therefore the aims of this study were to compare the clinical validity of Phi, PCA3, and T2:ERG in general and in different subpopulations of men who had undergone initial and repeated biopsies and to evaluate the diagnostic power of these biomarkers within a multivariate model. It can be assumed that the urinary- and serum-based markers, although they reflect different tissue situations and therefore differ in their diagnostic validity, might have substantially improved diagnostic power if used together.

Materials and Methods

STUDY POPULATION

We evaluated 320 samples from 281 men from 2 prostate cancer centers (center 1, Berlin, n = 194; center 2, Muenster, n = 126) from January 2009 to March 2012 (see flowchart in Fig. 1). The study was approved by

[7] Human genes: *TMPRSS2*, transmembrane protease, serine 2; *ERG*, v-ets erythroblastosis virus E26 oncogene homolog (avian).

both hospital ethics committees and reported in accordance with the Standards for the Reporting of Diagnostic Accuracy *(24)*. Urine and serum samples were collected from men scheduled for prostate biopsy (10–22 cores) owing to suspicious DRE, suspicious transrectal ultrasonography findings, or increased PSA concentration or PSA velocity. Study exclusion criteria included urinary infections, medications (androgen or 5-α-reductase inhibitors), or interventions that could alter PSA concentrations. Of the 281 patients included in this study, 35 were excluded for reasons of PSA >20 μg/L (n = 20), not giving consent for biopsy or not being biopsied within an appropriate timeframe (n = 10), insufficient cells in urine for PCA3 measurement (n = 2), and no fPSA measured (n = 3). In addition, for 22 men with a total of 61 measurements, 39 repeated measurements (second, n = 22; third, n = 12; fourth, n = 4; and fifth; n = 1) were performed within the study period owing to repeat biopsies or active surveillance. Thus, these 39 samples after the first measurement were also excluded for the evaluation of data.

The final study cohort of 246 men included 110 PCa patients (45%). Of the 110 PCa patients, 37 had a suspicious DRE (34%). An initial biopsy was performed in 70 (64%) of the patients and 1–6 repeat biopsies in 40 (36%). One repeat biopsy was performed in 21 men, 2 in 8 men, 3 in 7 men, 4 in 3 men, and 6 repeat biopsies in 1 man. The Gleason scores at biopsy were distributed as follows: Gleason 5, n = 2; Gleason 6, n = 70; Gleason 7, n = 29; Gleason 8, n = 4; Gleason 9, n = 4; and Gleason 10, n = 1.

Of the 136 men with no evidence of malignancy (NEM), 109 had benign biopsy results (suspicious DRE, 14%) and 27 had the diagnosis of high-grade prostatic intraepithelial neoplasia (suspicious DRE, 11%), of whom 2 of 3 had at least 1 repeat biopsy. The 136 NEM patients included 66 (49%) with initial biopsy and 70 (51%) with 1–7 repeat biopsies, of whom 37 had 1, 17 had 2, 9 had 3, 4 had 4, 1 had 5, 1 had 6, and 1 had 7 repeat biopsies.

METHODS

Blood and urine sampling were done within 14 days before biopsy. Sample collection and storage were performed according to a recommended standard operating procedure for [−2]proPSA *(25)*. Thus, blood samples were centrifuged within 3 h and if analyzed within 48 h, serum was stored at 4 °C. All other analyzed samples were stored at −70 or −80 °C for a maximum of 3 years. Of the 126 serum samples from center 2, 123 were measured prospectively with PSA, fPSA, and [−2]proPSA. The remaining 3 samples from center 2 and all 194 samples from center 1 were analyzed retrospectively. All se-

Fig. 1. Flowchart for all analyzed samples and all final included patients.
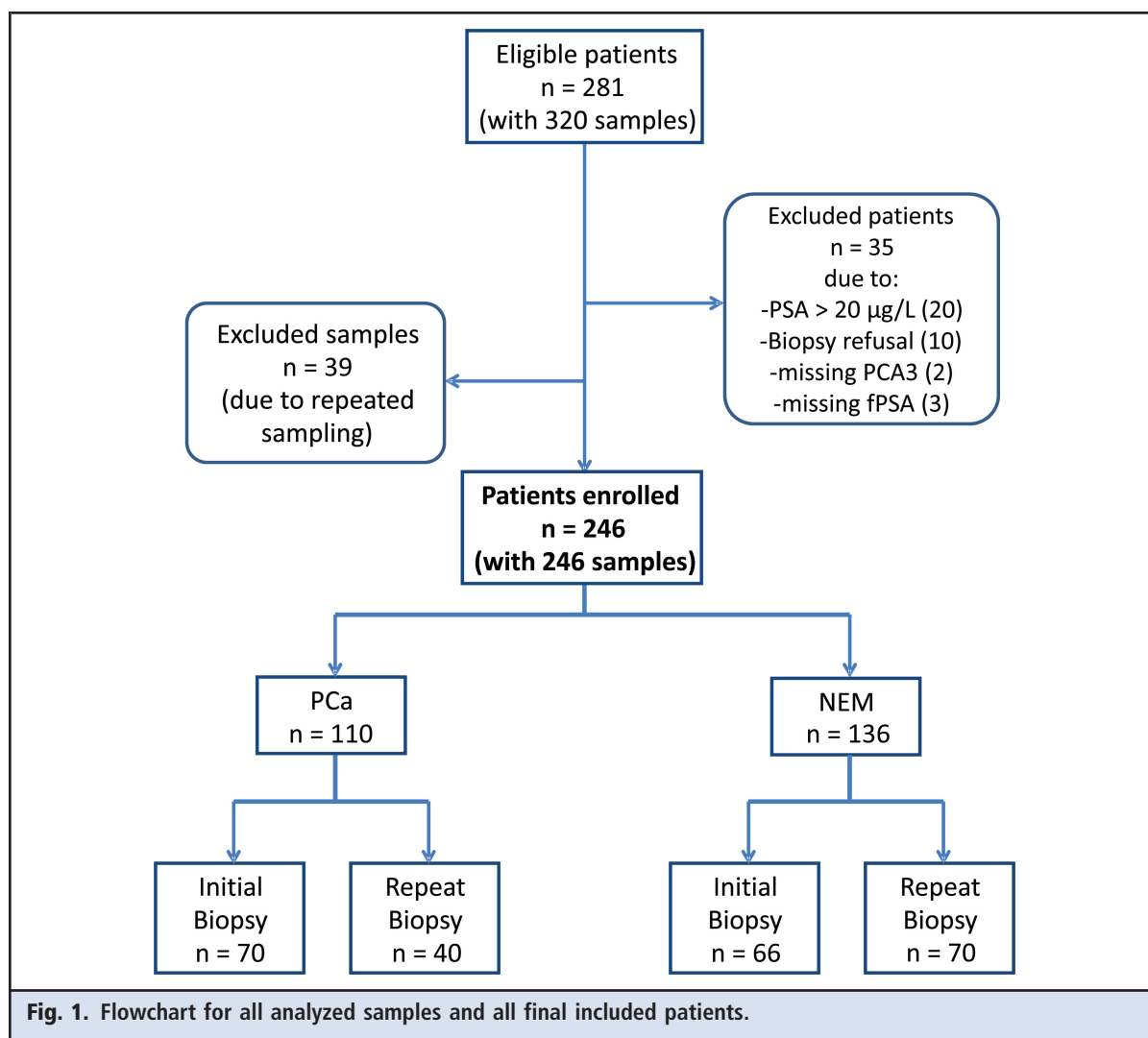
rum samples were processed by the Access 2 Immunoassay System analyzer (Beckman Coulter), calibrated against the WHO standard for PSA and fPSA, with approximately 20%–25% lower values compared with traditional Hybritech calibration *(26)*. The analytical performance of the measurements assessed with control materials (Beckman Coulter) showed values within the allowed recommended limits.

After a DRE with 3 strokes per lobe as described earlier *(27)*, urine samples were collected (PROGENSA PCA3 urine sample collection kit, Hologic Gen-Probe), and the PROGENSA PCA3 assay and T2:ERG research test were performed retrospectively in all samples. The PCA3 score was calculated as: (mRNA PCA3)/(mRNA PSA) × 1000. For T2:ERG the urine samples were processed similarly and the T2:ERG score was calculated as: (mRNA T2:ERG)/(mRNA PSA) × 100000 *(20)*.

Transrectal ultrasonography was used to determine prostate volume. The Gleason scores were estimated according to the 2005 consensus conference of the International Society of Urological Pathology *(28)*.

STATISTICAL ANALYSIS

A sample size of 182 was needed assuming a 45% PCa positive biopsy rate and 2.5% nonevaluable study participants to demonstrate a sensitivity of 80% and a specificity of 50% within a CI of ±10% at the 0.05 significance level (2-sided).

Statistical analyses were performed with SPSS version 19.0 (SPSS Inc) and MedCalc version 12.2.1 (MedCalc Software). Several tests [Mann–Whitney

Table 1. Comparisons between patients with NEM and those with PCa.[a]				
Characteristics	**All**	**PCa**	**NEM**	***P***
No. of patients	246	110	136	
Age, years	65 (59–70) [41–81]	67 (61–71) [49–80]	64.5 (58–69) [41–81]	0.004
PSA, μg/L	6.05 (4.06–8.12) [0.50–19.77]	6.25 (4.35–8.45) [0.59–19.30]	5.68 (3.81–7.78) [0.50–19.77]	0.196
Volume, mL[b]	45 (32–63) [14–166]	40 (30–56) [14–150]	50 (35–67) [16–166]	0.002
%fPSA	14.4 (10.7–19.2) [4.58–47.4]	13.4 (9.55–18.3) [5.39–33.5]	16.0 (11.6–19.3) [4.58–47.4]	0.007
[−2]proPSA, ng/L	13.6 (9.7–23.1) [1.9–60.9]	16.1 (10.4–27.4) [2.3–60.9]	12.4 (9.15–19.7) [1.9–58.1]	0.017
%[−2]proPSA $\times 10^3$	18.5 (13.5–24.8) [3.7–80.4]	21.6 (16.4–27) [5.7–80.4]	16.1 (12.2–22.7) [3.7–41.4]	<0.0001
Phi	42.3 (31.5–60.2) [9.4–210.7]	50.6 (36.9–69.5) [15.3–210.7]	37.2 (28.2–49.1) [9.4–145.2]	<0.0001
T2:ERG	29.3 (8.8–108) [0–85064]	51.8 (14.3–229) [0.22–85064]	23.6 (6.3–54.8) [0–1371]	0.0003
PCA3	30 (16–60) [1–333]	51 (25–88) [7–231]	22.5 (11–41.5) [1–333]	<0.0001

[a] All data are given as median (interquartile range 25%–75%) including the [range].
[b] Prostate volume data are missing from one PCa and 3 NEM patients.

U-test, Kruskal Wallis test, McNemar test, and Spearman rank correlation coefficient (r_s)] were performed and are indicated in the corresponding passage of the text. Area under the curve (AUC) was estimated from ROC curves according to the method of DeLong et al. *(29)*. ROC curves were used to compare specificities at given sensitivities. *P* values <0.05 (2-sided) were considered significant.

To test the ability of Phi, PCA3, and T2 to improve specificity in detecting PCa at biopsy, these variables were used together in a multivariate artificial neural network (ANN) and binary logistic regression models with age, PSA, %fPSA, prostate volume, and DRE status as described previously in detail *(30)*. These models comprised only 1, 2, or all 3 new biomarkers without partial or full addition of the 5 traditional parameters PSA, %fPSA, prostate volume, age, and DRE status. Each of the ANN models had 3 layers: 1 input layer with 2–8 neurons, 1 hidden layer with 2 or 3 neurons, and 1 output neuron that ranged from 0 (low) to 1 (high PCa risk). ROC curve analyses for single parameters and ANN modeling using Bayesian regularization were performed with the whole data group or subgroups, respectively, and the leave-one-out approach was used for internal validation of the ANN models.

The models, including the calibration plots and the decision analysis *(31)* for models and single parameters, were constructed with MATLAB-software, especially the Neural Network Toolbox (Mathworks).

Results

CHARACTERISTICS AND SIGNIFICANCES OF THE STUDY GROUP
The 169 patients included from center 1 (77 PCa, 51 with initial biopsies; 92 NEM, 41 with initial biopsies)

were first compared with 77 patients included from center 2 (33 PCa, 19 with initial biopsies; 44 NEM, 25 with initial biopsies). No significant differences were found between the 2 centers for age, prostate volume, PSA, %fPSA, Phi, PCA3, and T2:ERG (*P* 0.13–0.99). Only the DRE status (center 1, 40% DRE-positive PCa; center 2, 18% DRE-positive PCa) differed significantly between PCa patients (*P* = 0.024). No difference was found between NEM patients from both centers (*P* = 0.744). The difference in PCa patients was negligible because DRE was not an independent predictor in any model nor did the AUC for DRE differ between both centers (0.602 vs 0.594; *P* = 0.87). AUCs for all other parameters did not differ significantly between centers 1 and 2 (*P* from 0.21–0.94). Thus, we evaluated the merged data of both centers.

PSA was not found to differ between PCa and NEM patients, but all other parameters were significantly different in univariate analysis, with the highest significance levels for the new markers PCA3, Phi, and T2:ERG (Table 1). Almost all significance levels remained unchanged when we subdivided the study cohort in those with an initial biopsy or 1–7 repeat biopsies and in those with a PSA concentration of 1.6–8 μg/L (WHO calibration) or 2–10 μg/L (traditional Hybritech calibration) regardless of whether DRE was negative or not.

ASSOCIATION BETWEEN CLINICOPATHOLOGICAL VARIABLES AND PCA3, Phi, AND T2:ERG
Correlation analyses were performed in all patients and in PCa patients separately (see Tables 1 and 2 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol59/issue1). In brief, PCA3 and Phi were significantly correlated [r_s = 0.17; *P* = 0.009] as were PCA3

and T2:ERG (r_s = 0.33; P < 0.0001) but not Phi and T2:ERG (r_s = 0.02; P = 0.72). PCA3 and T2:ERG did not show any association with PSA or %fPSA (P from 0.14–0.78), but Phi was strongly correlated with PSA (r_s = 0.4; P < 0.0001) and inversely with %fPSA (r_s = −0.44; P < 0.0001), as expected. In PCa patients, only PSA (r_s = 0.44; P < 0.0001) and Phi (r_s = 0.3; P = 0.002) were strongly correlated with Gleason score, whereas PCA3, %fPSA, and T2:ERG were not (P - 0.2–0.91).

ROC ANALYSES

In Table 2 and Fig 2, the AUCs of most parameters and the best ANN model are shown for the total group and the subcohorts according to the PSA value and DRE status or initial vs repeat biopsy. Data on age (AUCs, 0.58–0.63), prostate volume (AUCs, 0.59–0.64), fPSA (AUCs, 0.52–0.56), and [−2]proPSA (AUCs, 0.56–0.62) are omitted to simplify analysis. Among the single parameters, PCA3 consistently had the largest and Phi the second largest AUC, although none of these AUC differences were significant (P - 0.21–0.78). With the exception of the repeat biopsy group and %[−2]proPSA, which is very similar to Phi, the T2:ERG was always the third best parameter, and differences in AUCs vs Phi or %fPSA were never significant (P - 0.08–0.8). However, it should be pointed out that PCA3 showed higher specificity values (some of which were significantly higher) than all the other single markers from the clinically relevant threshold of 90% sensitivity upward. With a 90% sensitivity cutoff, 6%, 15%, or 20% of biopsies could have been avoided in the whole population with the use of Phi, PCA3, or the best model compared with PSA.

All ANN models with the largest AUC in the respective groups were based only on Phi and PCA3, with the exception of the repeat biopsy group, for which the further addition of T2:ERG yielded in the best AUC (Table 2, Fig. 2B). There was a modest but not statistically significant improvement over PCA3 alone when PCA3 and Phi (and in 1 case, T2:ERG) were added to the ANN model (AUC gains ranged from 0.01 to 0.04, P from 0.05 to 0.97). In the 3 groups with initial and repeat biopsies and PSA concentrations of 2–10 μg/L with negative DRE, the ANN was not significantly different from Phi.

DECISION CURVE ANALYSIS

To compare the clinical usefulness of each single marker (32) with PSA and the classic ANN model with the variables age, PSA, %fPSA, prostate volume, and DRE status (model 1) as well as with the expanded ANN model in which PCA3 and Phi were included (model 2) or a model based only on PCA3 and Phi (model 3), decision curve analyses (DCA) were per-

Table 2. ROC curve analysis with AUC and specificities at 90% sensitivity for all parameters and the respective best model.[a]

| Criterion/ parameter	All patients with PSA 0.5–20 μg/L (n = 246)		Initial biopsy (n = 136)		Repeat biopsy (n = 110)		Patients with PSA 2–10 μg/L[b] (n = 197)		Patients with PSA 2–10 μg/L[b] and negative DRE (n = 154)	
	AUC	Specificity	AUC	Specificity	AUC	Specificity	AUC	Specificity	AUC	Specificity
PSA	0.55[c,d] (0.48–0.61)	15.4[d] (9.8–22.6)	0.6 (0.51–0.68)	18.2 (9.8–29.6)	0.55[d] (0.46–0.65)	14.3[d] (7.1–24.7)	0.51[c,d] (0.44–0.58)	13.5[d] (7.8–21.3)	0.52[c,d] (0.43–0.6)	12.6[d] (6.7–21)
%fPSA	0.6[c,d] (0.54–0.66)	18.4[d] (12.3–25.9)	0.6 (0.51–0.68)	16.7 (8.6–27.9)	0.61[d] (0.51–0.7)	21.4 (12.5–32.9)	0.58[c,d] (0.51–0.65)	14.4[d] (8.5–22.4)	0.62[d] (0.54–0.7)	16.8 (9.9–25.9)
T2:ERG	0.63[d] (0.57–0.69)	10.3[c,d] (5.7–16.7)	0.68 (0.59–0.75)	21.1 (12.1–33)	0.53[d] (0.44–0.63)	4.3[d] (0.9–12)	0.66[d] (0.59–0.73)	10.8[d] (5.7–18.1)	0.63[c] (0.54–0.7)	10.5[d] (5.2–18.5)
%[−2]proPSA	0.66 (0.6–0.72)	21.3[d] (14.8–29.2)	0.64 (0.55–0.72)	27.3 (17–39.6)	0.66 (0.57–0.75)	21.4 (12.5–32.9)	0.67[d] (0.6–0.74)	19.8 (12.9–28.5)	0.7 (0.62–0.77)	30.5 (21.5–40.8)
Phi	0.68 (0.62–0.74)	21.3[d] (14.8–29.2)	0.68 (0.6–0.76)	28.8 (18.3–41.3)	0.69 (0.6–0.78)	18.6[d] (10.3–29.7)	0.68 (0.61–0.75)	18.9 (12.1–27.5)	0.71 (0.63–0.78)	22.1 (14.2–31.8)
PCA3	0.74 (0.68–0.79)	33.8[c] (25.9–42.4)	0.7 (0.62–0.78)	31.8 (20.9–44.4)	0.77 (0.69–0.85)	35.7[c] (24.6–48.1)	0.75 (0.68–0.81)	31.5 (23.6–48.5)	0.74 (0.67–0.81)	33.7 (0.67–0.81)
ANN[e]	0.76[c] (0.70–0.81)	38.2[c] (30–47)	0.73 (0.65–0.81)	30.3 (19.6–42.9)	0.78 (0.69–0.85)	55.7[c] (43.3–67.6)	0.79[c] (0.73–0.85)	38.7[c] (29.6–48.5)	0.77 (0.7–0.83)	39[c] (29.1–49.5)

[a] AUC and specificity at 90% sensitivity with 95% CIs (in parentheses).
[b] The traditional 2–10 μg/L PSA range was also used for the WHO calibration.
[c] Significantly different from Phi (P <0.05) regarding AUC [DeLong et al. (29)] or specificity at 90% sensitivity (McNemar test).
[d] Significantly different from PCA3 (P <0.05) regarding AUC [DeLong et al. (29)] or specificity at 90% sensitivity (McNemar test).
[e] The respective ANN model was built only with Phi and PCA3 except the repeat biopsy group, for which T2:ERG was the third input value.
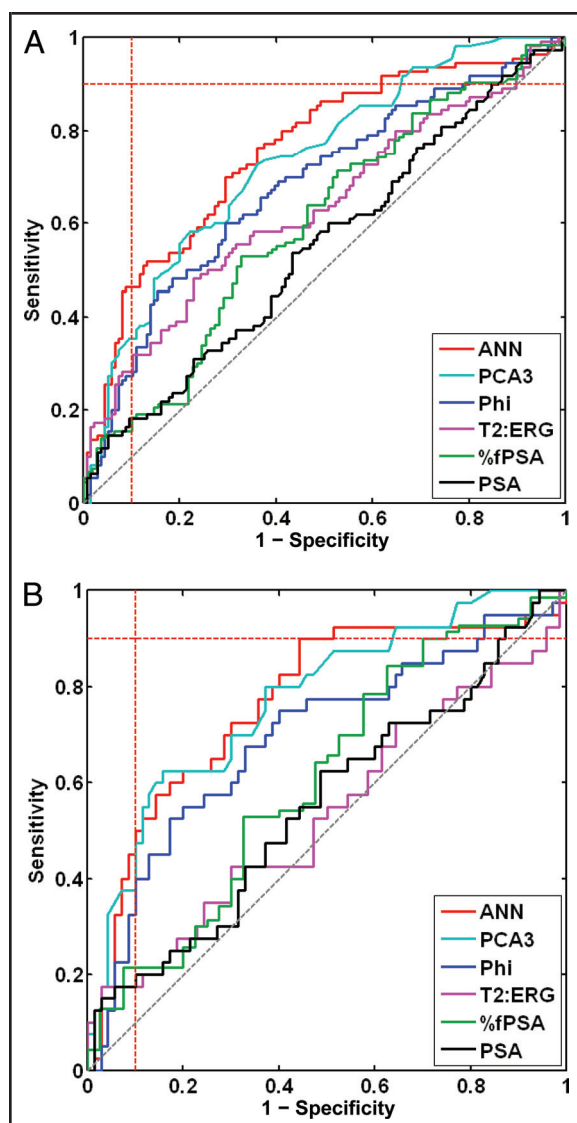
Fig. 2. (A), ROC curves for PCA3, Phi, T2:ERG, %fPSA or PSA, and for the ANN model based on PCA3 and Phi for the whole cohort of 246 men within the PSA range of 0.5–20 μg/L.

(B), ROC curves for the ANN model based on PCA3, Phi, T2:ERG, and the single parameters PCA3, Phi, T2:ERG, %fPSA, and PSA for the repeat biopsy cohort of 110 men with PSA within the range of 0.5–20 μg/L.

formed. The DCA of PCA3 and Phi (T2:ERG is not shown) showed a higher net benefit and larger useful probability range than PSA and %fPSA (see online Supplemental Fig. 1), whereas both ANN models with PCA3 and Phi had an increased net benefit against the classic ANN model (see online Supplemental Fig. 2). The PCA3- and Phi-related ANN models also showed a

better calibration than model 1 (see online Supplemental Fig. 3).

Discussion

There is an unquestioned need to supplement the information from PSA analysis to improve PCa detection and reduce unnecessary biopsy and overtreatment. PCA3 is already accepted as useful biomarker for patients undergoing repeat biopsies *(11, 14, 33)* as well as in initial biopsy cohorts *(13, 34)*. The combination of PSA, fPSA, and [−2]proPSA via the formula for Phi has been proven to enhance the diagnostic accuracy of PSA and %fPSA for pathological staging *(35)* and for early PCa diagnosis especially at initial biopsy *(9, 36)*. Recently, %[−2]proPSA and Phi have also been successfully tested in repeat biopsy patients, but in contrast with our study, data on %[−2]proPSA were better than on Phi *(37)*. During the preparation of this manuscript, a report of a study with combined data on Phi and PCA3 in 151 patients was published *(22)*. Data on urinary PCA3 and T2:ERG were available from a small cohort of 78 men *(19)*. The quantitative assay for T2:ERG was introduced later *(18)*, and data have been published on PCA3 and T2:ERG used together with PSA *(20, 38)*. To our knowledge, data from studies that have combined data on all 3 markers Phi, PCA3, and T2:ERG have not been available.

In our study, Phi, PCA3, and T2:ERG showed significantly higher values in PCa patients compared with NEM patients, although PSA was equal (Table 1). Our AUC data suggested that PCA3 had slightly improved diagnostic accuracy compared with Phi, although the AUCs were not statistically different (Table 2). In the recently published study by Ferro et al. *(22)* conducted with 151 initially biopsied patients, PCA3 and Phi also were noted to perform comparably. However, in our study PCA3 had the largest AUCs, and significant differences in specificity compared with Phi were reached at 90% sensitivity in the whole (34% vs 21%) and repeat biopsy (36% vs 19%) cohorts. This finding opens discussion about whether PCA3 should eventually be used as the best single parameter instead of PSA as the first line diagnostic test for PCa detection, as already discussed by Roobol et al. *(39)*. Another option could be to reserve PCA3 for those patients with repeat biopsies, for whom the advantage is the greatest over Phi in terms of AUC (0.08) and at 90% sensitivity (17% higher specificity). On the other hand, the combination of PSA, fPSA, and proPSA in the form of Phi should be preferred, owing to its high specificity at 90% sensitivity in the initial biopsy group, its economic advantage, and its lower discomfort for the patients. In-

terestingly, although PCA3 was better than all other parameters except Phi and the ANN, Phi itself was significantly superior to %fPSA in only 2 cohorts. Comparison data of Phi to %fPSA were not provided by Ferro et al. *(22)*, but in other studies a significant improvement of Phi over %fPSA was also found *(9, 36)*. In the cohort with initial biopsies, Phi had the smallest AUC difference (0.02) compared with PCA3. Compared with PCA3, Phi performed best in the initial biopsy group, yielding an AUC of 0.68 compared to an AUC for PCA3 of 0.7 in this group and 0.71 in those patients with a negative DRE in the PSA range of 2–10 μg/L. However, the small size of some subgroups was a limitation of our study. In 3 subgroups the calculated sample size of 182 could not be reached, so further comparisons in larger cohorts will be useful.

The usefulness of multivariable models was limited in this first study comparing PCA3, Phi, and T2:ERG. Ferro et al. *(22)* stated that multivariable analysis produced no significant model to improve the performance of their single biomarkers. Inclusion of all factors such as age, DRE status, prostate volume, PSA, %fPSA, or T2:ERG did not improve the AUC over the simple model with PCA3 and Phi only. About 10 years ago, there was a need to include rather inaccurate parameters such as prostate volume or the DRE status *(40, 41)*. It will be advantageous if these clinical parameters are no longer needed. A study by Chun et al. *(14)* showed a significant AUC gain ($P = 0.04$) from 0.68 (for PCA3) to 0.73 (model) in a large study of 809 men with PCA3. A similar AUC gain from 0.8 (model without [−2]proPSA) to 0.85 was seen in one of the first studies using the automated [−2]proPSA assay *(8)*.

Regarding new single biomarkers, it should be emphasized that this study was the first to provide data on the quantitative urinary assay T2:ERG that included a comparison with the best available PCa biomarkers. T2:ERG demonstrated a slightly better performance compared with PSA and %fPSA, but no statistical significance was found. In this study, T2:ERG accuracy for predicting repeat biopsy outcome was lower than that observed in prior studies that used the same T2:ERG research assay *(20, 42)*. The overall performance in multivariable models was not improved by T2:ERG. However, the advantage of T2:ERG might be seen in subgroups of aggressive PCa. These investigations were not part of the present study. Our study was performed to show the performance of the quantitative T2:ERG assay and to compare this biomarker with those having the best diagnostic accuracy, such as PCA3 and Phi. Interestingly, correlations with the Gleason score were not seen with the PCa-specific urine markers but only with Phi and even more strongly with PSA.

To generate cutoff values, absolute values for PCA3, Phi, and T2:ERG are necessary but there are different recommendations, especially for PCA3. A sensitivity of 90% can be reached with an absolute PCA3 level of 15 on the basis of our data. In this study, the highest efficacy for a PCA3 value was found at 28 with 73% sensitivity and 64% specificity, which is very close to the FDA-approved cutoff of 25. A 90% sensitivity cutoff for Phi corresponded with a value of 27.5. At 90% sensitivity Phi improved specificity over that of %fPSA and PSA by 3% and 6%, whereas PCA3 prevented 15.4% and 18.4% and the ANN 20% and 23% of unnecessary biopsies compared with %fPSA and PSA, respectively. For the research assay of T2:ERG there are no existing cutoff recommendations so far. However, a 90% sensitivity cutoff would correspond to an absolute value of 2.3.

To summarize, PCA3 and Phi outperformed other analyzed biomarkers, whereas T2:ERG failed to significantly improve the ability to diagnose PCa. PCA3 reached the largest AUC and showed higher-specificity values than all the other single markers from the clinically relevant threshold of 90% sensitivity upward. Although PCA3 showed the largest advantage in the repeat biopsy cohort, Phi and PCA3 performed comparably in the initial biopsy cohort and in the 2–10 μg/L PSA range cohort with negative DRE. The combination of both markers further enhanced the diagnostic power and the clinical utility as shown by DCA.

References

1. Lilja H, Ulmert D, Vickers AJ. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. Nat Rev Cancer 2008;8:268–78.

2. Catalona WJ, Partin AW, Slawin KM, Brawer MK, Flanigan RC, Patel A, et al. 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;279:1542–7.

3. Roddam AW, Duffy MJ, Hamdy FC, Ward AM, Patnick J, Price CP, et al. Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2–10 ng/ml: systematic review and meta-analysis. Eur Urol 2005;48:386–99.

4. Schroder F, Kattan MW. The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review. Eur Urol 2008;54:274–90.

5. Mikolajczyk SD, Marks LS, Partin AW, Rittenhouse HG. Free prostate-specific antigen in serum is becoming more complex. Urology 2002;59:797–802.

6. Catalona WJ, Bartsch G, Rittenhouse HG, Evans CL, Linton HJ, Amirkhan A, et al. Serum pro prostate specific antigen improves cancer detection compared to free and complexed prostate specific antigen in men with prostate specific antigen 2 to 4 ng/ml. J Urol 2003;170:2181–5.

7. Sokoll LJ, Wang Y, Feng Z, Kagan J, Partin AW, Sanda MG, et al. [-2]proenzyme prostate specific antigen for prostate cancer detection: a national cancer institute early detection research network validation study. J Urol 2008;180:539–43.

8. Stephan C, Kahrs AM, Cammann H, Lein M, Schrader M, Deger S, et al. A [-2]proPSA-based artificial neural network significantly improves differentiation between prostate cancer and benign prostatic diseases. Prostate 2009;69:198–207.

9. Guazzoni G, Nava L, Lazzeri M, Scattoni V, Lughezzani G, Maccagnano C, et al. Prostate-specific antigen (PSA) isoform p2PSA significantly improves the prediction of prostate cancer at initial extended prostate biopsies in patients with total PSA between 2.0 and 10 ng/ml: results of a prospective study in a clinical setting. Eur Urol 2011;60:214–22.

10. Jansen FH, van Schaik RH, Kurstjens J, Horninger W, Klocker H, Bektic J, et al. Prostate-specific antigen (PSA) isoform p2PSA in combination with total PSA and free PSA improves diagnostic accuracy in prostate cancer detection. Eur Urol 2010;57:921–7.

11. Haese A, de la Taille A, Van Poppel H, Marberger M, Stenzl A, Mulders PF, et al. Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy. Eur Urol 2008;54:1081–8.

12. Perdona S, Cavadas V, Di Lorenzo G, Damiano R, Chiappetta G, Del Prete P, et al. Prostate cancer detection in the "grey area" of prostate-specific antigen below 10 ng/ml: head-to-head comparison of the updated PCPT calculator and Chun's nomogram, two risk estimators incorporating prostate cancer antigen 3. Eur Urol 2011;59:81–7.

13. de la Taille A, Irani J, Graefen M, Chun F, de Reijke T, Kil P et al. Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions. J Urol 2011;185:2119–25.

14. Chun FK, de la Taille A, Van Poppel H, Marberger M, Stenzl A, Mulders PF, et al. Prostate cancer gene 3 (PCA3): development and internal validation of a novel biopsy nomogram. Eur Urol 2009;56:659–68.

15. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;310:644–8.

16. Demichelis F, Fall K, Perner S, Andren O, Schmidt F, Setlur SR, et al. TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene 2007;26:4596–9.

17. Esgueva R, Perner S, LaFargue J, Scheble V, Stephan C, Lein M, et al. Prevalence of TMPRSS2-ERG and SLC45A3-ERG gene fusions in a large prostatectomy cohort. Mod Pathol 2010;23:539–46.

18. Groskopf J, Siddiqui J, Aubin SMJ, Sefton-Miller L, Day J, Blase A, et al. Feasibility and clinical utility of a TMPRSS2:ERG gene fusion urine test [Abstract]. Eur Urol Suppl 2009;8:195.

19. Hessels D, Smit FP, Verhaegh GW, Witjes JA, Cornel EB, Schalken JA. Detection of TMPRSS2-ERG fusion transcripts and prostate cancer antigen 3 in urinary sediments may improve diagnosis of prostate cancer. Clin Cancer Res 2007;13:5103–8.

20. Tomlins SA, Aubin SM, Siddiqui J, Lonigro RJ, Sefton-Miller L, Miick S, et al. Urine TMPRSS2:ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA. Sci Transl Med 2011;3:94ra72.

21. Salagierski M, Schalken JA. Molecular diagnosis of prostate cancer: PCA3 and TMPRSS2:ERG gene fusion. J Urol 2012;187:795–801.

22. Ferro M, Bruzzese D, Perdona S, Mazzarella C, Marino A, Sorrentino A, et al. Predicting prostate biopsy outcome: prostate health index (phi) and prostate cancer antigen 3 (PCA3) are useful biomarkers. Clin Chim Acta 2012;413:1274–8.

23. Aubin SM, Miick S, Williamsen S, Hodge P, Meinke J, Blase A, et al. Improved prediction of prostate biopsy outcome using PCA3, TMPRSS2:ERG gene fusion and serum PSA. J Urol Suppl 2008;179:725.

24. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem 2003;49:1–6.

25. Semjonow A, Kopke T, Eltze E, Pepping-Schefers B, Burgel H, Darte C. Pre-analytical in-vitro stability of [-2]proPSA in blood and serum. Clin Biochem 2010;43:926–8.

26. Stephan C, Bangma C, Vignati G, Bartsch G, Lein M, Jung K, et al. 20–25% lower concentrations of total and free prostate-specific antigen (PSA) after calibration of PSA assays to the WHO reference materials–analysis of 1098 patients in four centers. Int J Biol Markers 2009;24:65–9.

27. Groskopf J, Aubin SM, Deras IL, Blase A, Bodrug S, Clark C, et al. APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. Clin Chem 2006;52:1089–95.

28. Epstein JI, Allsbrook WC Jr, Amin MB, Egevad LL. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma. Am J Surg Pathol 2005;29:1228–42.

29. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.

30. Stephan C, Jung K, Soosaipillai A, Yousef GM, Cammann H, Meyer H, et al. Clinical utility of human glandular kallikrein 2 within a neural network for prostate cancer detection. BJU Int 2005;96:521–7.

31. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.

32. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak 2008;8:53.

33. Auprich M, Augustin H, Budaus L, Kluth L, Mannweiler S, Shariat SF, et al. A comparative performance analysis of total prostate-specific antigen, percentage free prostate-specific antigen, prostate-specific antigen velocity and urinary prostate cancer gene 3 in the first, second and third repeat prostate biopsy. BJU Int 2012;109:1627–35.

34. Goode RR, Marshall SJ, Duff M, Chevli E, Chevli KK. Use of PCA3 in detecting prostate cancer in initial and repeat prostate biopsy patients. Prostate 2013;73:48–53.

35. Guazzoni G, Lazzeri M, Nava L, Lughezzani G, Larcher A, Scattoni V, et al. Preoperative prostate-specific antigen isoform p2PSA and its derivatives, %p2PSA and prostate health index, predict pathologic outcomes in patients undergoing radical prostatectomy for prostate cancer. Eur Urol 2012;61:455–66.

36. Catalona WJ, Partin AW, Sanda MG, Wei JT, Klee GG, Bangma CH, et al. A multicenter study of [-2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range. J Urol 2011;185:1650–5.

37. Lazzeri M, Briganti A, Scattoni V, Lughezzani G, Larcher A, Gadda GM, et al. Serum index test %[-2]proPSA and prostate health index are more accurate than prostate specific antigen and %fPSA in predicting a positive repeat prostate biopsy. J Urol 2012;188:1137–43.

38. Salami SS, Schmidt F, Laxman B, Regan MM, Rickman DS, Scherr D, et al. Combining urinary detection of TMPRSS2:ERG and PCA3 with serum PSA to predict diagnosis of prostate cancer. Urol Oncol [Epub ahead of print 2011 May 19].

39. Roobol MJ, Schroder FH, van Leeuwen P, Wolters T, van den Bergh RC, van Leenders GJ, Hessels D. Performance of the prostate cancer antigen 3 (PCA3) gene and prostate-specific antigen in pre-

screened men: exploring the value of PCA3 for a first-line diagnostic test. Eur Urol 2010;58:475–81.

40. Stephan C, Cammann H, Semjonow A, Diamandis EP, Wymenga LFA, Lein M, et al. Multicenter evaluation of an artificial neural network to increase prostate cancer detection rate and reduce

unnecessary biopsies. Clin Chem 2002;48:1279–87.

41. Stephan C, Cammann H, Meyer HA, Muller C, Deger S, Lein M, Jung K. An artificial neural network for five different assay systems of prostate-specific antigen in prostate cancer diagnostics. BJU Int 2008;102:799–805.

42. Cornu JN, Cancel-Tassin G, Egrot C, Gaffory C, Haab F, Cussenot O. Urine TMPRSS2:ERG fusion transcript integrated with PCA3 score, genotyping, and biological features are correlated to the results of prostatic biopsies in men at risk of prostate cancer. Prostate [Epub ahead of print 2012 Jul 20].