SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
PROSTATE CANCER
DIAGNOSTIC MARKERS

Received
5 March 2014

Accepted
1 May 2014

Published
23 May 2014

Correspondence and requests for materials should be addressed to Y.T. (83432350@qq.com) or J.F.X. (jxu@wakehealth.edu)

* These authors contributed equally to this work.

Diagnostic ability of %p2PSA and prostate health index for aggressive prostate cancer: a meta-analysis

Wenying Wang[1,2]*, Meilin Wang[2]*, Li Wang[2], Tamara S. Adams[2], Ye Tian[1] & Jianfeng Xu[2]

[1]Department of Urology, Beijing Friendship Hospital, Capital Medical University, Beijing, 100050, China, [2]Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA.

The role of [-2]proPSA (p2PSA) based diagnostic tests for the detection of aggressive prostate cancer (PCa) has not been fully evaluated. We conducted a meta-analysis to evaluate the diagnostic performance of p2PSA/free PSA (%p2PSA) and prostate health index (Phi) tests for PCa and to evaluate their ability in discriminating between aggressive and non-aggressive PCa. A total of 16 articles were included in this meta-analysis. For the detection of PCa, the pooled sensitivity, specificity, and AUC were 0.86 (95% CI, 0.84–0.87), 0.40 (95% CI, 0.39–042) and 0.72 (95% CI, 0.67–0.77) for %p2PSA respectively, and were 0.85 (95% CI, 0.83–0.86), 0.45 (95% CI, 0.44–0.47) and 0.70 (95% CI = 0.65–0.74) for Phi, respectively. In addition, the sensitivity for discriminating PCa between higher Gleason score (≥7) and lower Gleason score (<7) was 0.96 (95% CI, 0.93–0.98) and 0.90 (95% CI, 0.87–0.92) for %p2PSA and Phi respectively, and the specificity was low, only 0.09 (95% CI, 0.06–0.12) and 0.17 (95% CI, 0.14–0.19) for %p2PSA and Phi, respectively. Phi and %p2PSA have a high diagnostic accuracy rates and can be used in PCa diagnosis. Phi and %p2PSA may be useful as tumor markers in predicating patients harboring more aggressive disease and guiding biopsy decisions.

Prostate Cancer (PCa) is one of the most commonly diagnosed malignant genitourinary tumors among men in United States and other Western countries[1]. To date, prostate-specific antigen (PSA) is the most widely used and successful serum marker to detect PCa, however, the clinical value of PSA for early detection of PCa has been questioned because of its poor specificity, especially in men with serum PSA levels in the 2–10 ng/ml range[2]. It is well-known that PSA testing leads to overdiagnosis and overtreatment of indolent PCa that does not evolve into aggressive life-threatening cancers[3]. Thus, significant effort has been placed on the development of new serum markers that could complement PSA for early aggressive PCa diagnosis.

It has been suggested that measurement of the precursor PSA isoform [−2] proPSA (p2PSA) and its derivatives might offer improvement of PCa detection in men with a total PSA (tPSA) level between 2 and 10 ng/ml[4]. The isoform p2PSA is a promising marker for PCa detection, and compared to other isoforms, measurement of the p2PSA to free PSA (%p2PSA) and Prostate Health Index (Phi), which is calculated using the formula (p2PSA/fPSA $\times \sqrt{}$tPSA) can better distinguish between PCa and non-PCa than tPSA and fPSA/tPSA (%fPSA)[5–7]. Values for %p2PSA and Phi were significantly higher in patients with PCa than patients with benign prostatic hyperplasia (BPH) and chronic histologic prostatic inflammation[8,9].

Previous studies have also shown that %p2PSA and Phi were significantly associated with an increased probability of detecting aggressive PCa, as indicated by a Gleason score ≥7[5,8]. Both %p2PSA and Phi gained attention as potentially very promising markers, that is, to be able to help predict aggressive PCa within a tPSA range of 2–10 ng/ml[10–12]. But the ability of these two markers to distinguish between aggressive PCa and non-aggressive PCa remains unclear. Here, a systematic review was performed to evaluate the emerging role of %p2PSA and Phi in diagnosing PCa, and the ability of these markers to distinguish patients with and without aggressive PCa.

Results

Description of included studies. As shown in Figure 1, we identified 323 papers from a literature search and excluded 299 irrelevant articles. The remaining 24 potentially relevant studies were further assessed. Eight studies were further excluded because they contained duplicated populations[6,8,13–18]. Although the study by Lazzeri et al.[6] was excluded due to partially duplicated data with the study of Lughezzani et al.[19], the data on aggressive and non-aggressive PCa were available and extracted.

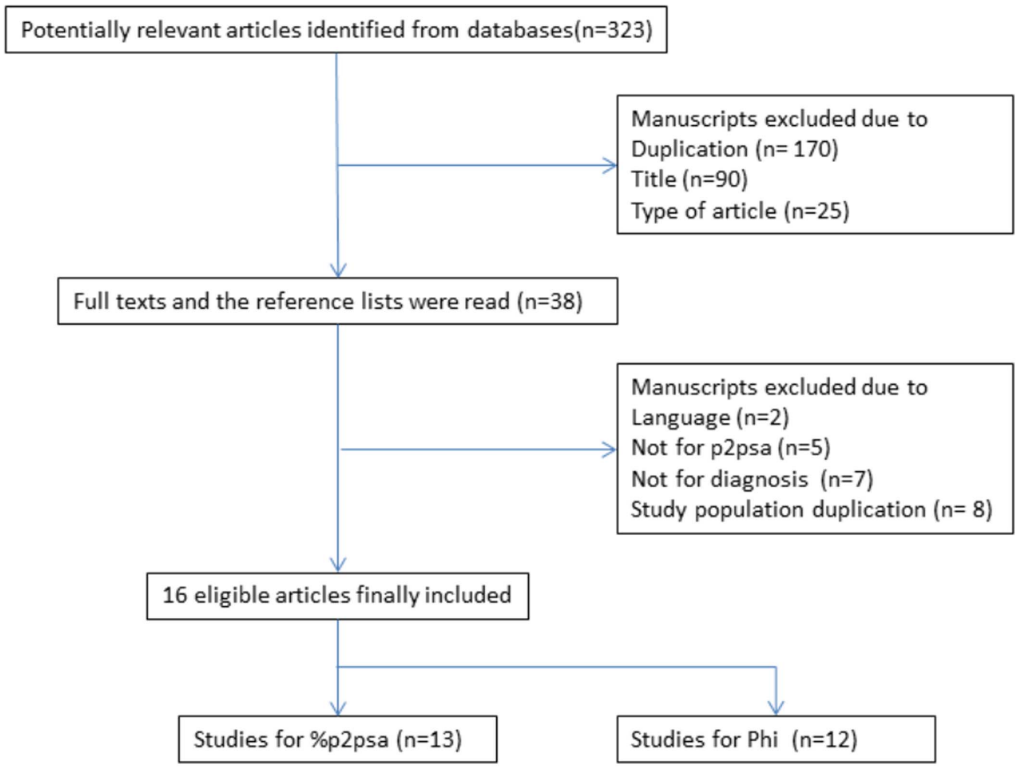

Figure 1 | Flowchart of selecting process for meta-analysis.

A total of 16 papers were included in the meta-analysis[5,7,10,11,19–30], 13 papers reported on %P2PSA, and 12 papers reported on Phi. In addition, the study by Jansen[7] included two different populations which were treated as two data sets, and the study by Scattoni[20] contained initial and repeat biopsy groups which were treated as two separate studies. Finally, 14 studies were reported on %P2PSA and Phi. The detailed characteristics of each study are presented in Table 1. All studies presented the results of sensitivity and specificity, and most studies presented the results of ROC curve. The cutoff points were reported in seven studies for %p2PSA and nine studies for Phi. The data on the patients with and without aggressive PCa were extracted from four and five studies for %p2PSA and Phi, respectively.

Overall meta-analyses for %p2PSA and Phi. A total of 6,279 patients (2,722 PCa and 3,557 non-PCa) and 5,856 patients (2,474 PCa and 3,382 non-PCa) were included in the pooled analyses for %p2PSA (Supplementary table 1) and Phi (Supplementary table 2), respectively. The pooled sensitivity and specificity were 0.86 (95% CI, 0.84–0.87) and 0.40 (95% CI, 0.39–042) for %p2PSA respectively, and were 0.85 (95% CI, 0.83–0.86) and 0.45 (95% CI, 0.44–0.47) for Phi, respectively (Figure 2).

The ROC space showed a curvilinear trend of points and Spearman's correlation coefficient was 0.705 ($P = 0.005$) for %p2PSA and 0.744 ($P = 0.002$) for Phi. It was suggested that there was the existence of a threshold. Due to the presence of significant heterogeneity for sensitivity ($I^2 = 90.4\%$, $\chi 2 = 135.15$, $P < 0.001$), specificity ($I^2 = 96.1\%$, $\chi 2 = 329.39$, $P < 0.001$) for %p2PSA; and sensitivity ($I^2 = 91.5\%$, $\chi 2 = 152.79$, $P < 0.001$), specificity ($I^2 = 93.1\%$, $\chi 2 = 188.66$, $P < 0.001$), and diagnostic OR (Cochran-Q = 24.25, $P = 0.0289$) for Phi; the diagnostic indices were calculated using a random effect model. We used a summary SROC curve to aggregate data and obtained a symmetrical curve with an AUC of 0.72 (95% CI = 0.67–0.77) and 0.70 (95% CI = 0.65–0.74) for %p2PSA and Phi, respectively (Supplementary Figure 1).

Subgroup analysis. Significant heterogeneity for Phi ($I^2 = 46.4\%$, $P = 0.0289$) and low heterogeneity ($I^2 = 29.1\%$, $p = 0.1448$) were observed. Therefore, further subgroup analyses were performed to explore the heterogeneity on data stratified by ethnicity group, study design and QUADAS score. Subgroup analyses showed that a retrospective study design ($I^2 = 61.6\%$, p = 0.0231) and low QUADAS score ($I^2 = 63.7\%$, p = 0.0171) were responsible for the heterogeneity of %p2PSA (Table 2); and the prospective study design ($I^2 = 51.3\%$, $p = 0.0368$), Caucasian group ($I^2 = 52.5\%$, $p = 0.0168$) and low QUADAS score ($I^2 = 55.9\%$, $p = 0.0201$) were responsible for the heterogeneity of Phi (Table 3).

The ability of %p2PSA and Phi to discriminate between aggressive and non-aggressive PCa. A Total of 728 cases of PCa (292 with Gleason score ≥ 7 versus 436 with Gleason score < 7) and 1,157 cases of PCa (431 with Gleason score ≥ 7 versus 726 with Gleason score < 7) were included in the pooled analysis of %p2PSA (Supplementary table 3) and Phi (Supplementary table 4).

The pooled sensitivity for the detection of PCa with higher Gleason score (≥ 7) was 0.96 (95% CI, 0.93–0.98) and 0.90 (95% CI, 0.87–0.92), while the specificity was low, only 0.09 (95% CI, 0.06–0.12) and 0.17 (95% CI, 0.14–0.19) for %p2PSA and Phi respectively (Figure 3). The pooled AUC of SROC was 0.54 (95% CI,−0.52–1.61) and 0.67 (95% CI, 0.57–0.77) for %p2PSA and Phi respectively (Supplementary Figure 1). The accuracy of %p2PSA and Phi in discriminating aggressive PCa with non-aggressive PCa is summarized in Table 4.

Assessment of publication bias. We analyzed possible publication bias by generating funnel plots, and the shape of the funnel plots was symmetrical in the diagnosis of PCa suggesting the absence of publication bias. We also used the Deeks' funnel plot asymmetry tests (t = 1.86, $P = 0.09$ for %p2PSA and t = 1.37, $P = 0.20$ for Phi), results suggested that there was no publication bias in the meta-analysis (Figure 4).

Table 1 | Characteristics of studies included in this meta-analysis for %p2PSA and Phi

First Author	Year	Sampling frame	Years of recruitment	Country	Ethnicities	Inclusion criteria	Patients age(y)	Assay method
Scattoni	2013	Prospective	2011.12–2012.5	Italy	Caucasian	PSA2–15 ng/ml, and/or positive DRE	67.5 ± 7.5	Beckman Coulter
Scattoni	2013	Prospective	2011.12–2012.5	Italy	Caucasian	PSA2–15 ng/ml, and/or positive DRE	67.5 ± 7.5	Beckman Coulter
Stephan	2013	Prospective, retrospective	2003–2012	Germany, France	Caucasian	PSA1.6–8 ng/ml,	64 (63–65)	Beckman Coulter
Perdonà	2013	Prospective	2010.3–2010.12	Italy	Caucasian	PSA<20 ng/ml, normal DRE	64.91 ± 7.37	Beckman Coulter
Ferro	2013	Prospective	N/A	Italy, Germany	Caucasian	Age ≥ 50,PSA2–10 ng/ml	60.4 ± 6.0	Beckman Coulter
Lazzeri	2013	Prospective	2011.3–2012.7	Italy, Germany, France, UK, Spain	Caucasian	Age ≥ 45,PSA2-10 ng/ml,and/or positive DRE	64.2 ± 7.5	Beckman Coulter
Ito	2013	Prospective	2004.4–2007.12	Japan	Asian	PSA2–10 ng/ml	66 (37–82)	Beckman Coulter
Ng	2013	Rertospective	2008.3–2013.3	China	Asian	Age ≥ 50,PSA4–10 ng/ml,normal DRE	65.9 (50–79)	Beckman Coulter
Lughezzani	2012	Prospective	2010.7–2011.7	Italy	Caucasian	PSA0.5–20 ng/ml, and/or abnormal DRE	64.3 ± 7.8	Beckman Coulter
Guazzoni	2011	Prospective	2010.3–2010.8	Italy	Caucasian	PSA2–10 ng/ml, normal DRE	63.3 ± 8.2	Beckman Coulter
Catalona	2011	Prospective,retrospective	2003.10–2009.6	USA	Caucasian	Age ≥ 50, PSA2–10 ng/ml, normal DRE	62.8 ± 7.0	Beckman Coulter
Jansen	2010	Rertospective	1994–1997.2	Austria	Caucasian	Age ≥ 50, PSA2–10 ng/ml	66(55–75)	Beckman Coulter
Jansen	2010	Rertospective	1993	Netherlands	Caucasian	Age ≥ 50, PSA2–10 ng/ml	69(50–77)	Beckman Coulter
Le	2010	Prospective	2007.4	USA	Caucasian	PSA2.5–10 ng/ml, normal DRE	65	Beckman Coulter
Sokoll	2010	Prospective	N/A	USA	Caucasian	PSA0.29–310.6 ng/ml	61.7 ± 8.6	Beckman Coulter
Stephan	2009	Rertospective	2002–2006	Germany	Caucasian	PSA0.26–28.4 ng/ml, normal DRE	62.1 ± 5.63	Beckman Coulter
Sokoll	2008	Rertospective	N/A	USA	Caucasian	PSA0.48–33.18 ng/ml	62.2 ± 8.2	Beckman Coulter
Catalona	2003	Rertospective	1995–2002	USA and Austria	Caucasian	PSA2–10 ng/ml	N/A	Research assay

First Author	Biopsy cores	Patients number	PCa number	AUC(PSA)	AUC(%fPSA)	AUC (%p2PSA)	Cutoff (%p2PSA)	AUC (Phi)	Cutoff (Phi)
Scattoni	14–24	116	40	0.54 (0.42–0.65)	0.67 (0.56–0.78)	N/A	No	0.69 (0.59–0.79)	Yes
Scattoni	14–24	95	30	0.60 (0.46–0.73)	0.52 (0.38–0.65)	N/A	No	0.72 (0.59–0.84)	Yes
Stephan	10–22	1362	668	0.56 (0.53–0.59)	0.61 (0.59–0.64)	0.72 (0.70–0.75)	No	0.74 (0.71–0.76)	No
Perdonà	18	160	47	0.51 (0.41–0.61)	0.61 (0.51–0.71)	0.68 (0.58–0.78)	Yes	0.71 (0.61–0.80)	Yes
Ferro	≥16	300	108	0.52 (0.45–0.59)	0.62 (0.55–0.69)	0.76 (0.71 0.82)	Yes	0.77 (0.72–0.83)	Yes
Lazzeri	≥12	646	264	0.50 (0.46–0.54)	0.64 (0.61–0.68)	0.67 (0.64–0.71)	Yes	0.67 (0.64–0.71)	Yes
Ito	8–22	239	53	N/A	N/A	N/A	Yes	N/A	Yes
Ng	10	230	21	0.547 (0.421–0.674)	0.538 (0.413–0.663)	0.768 (0.660–0.876)	Yes	0.781 (0.675–0.887)	Yes
Lughezzani	≥18	729	280	0.51 (0.48–0.55)	0.62 (0.58–0.65)	N/A	No	0.70 (0.66–0.73)	No
Guazzoni	18–22	268	107	0.53(0.47–0.59)	0.58(0.52–0.64)	0.76(0.71–0.81)	Yes	0.76 (0.70–0.81)	Yes
Catalona	≥6	892	430	0.525	0.648	N/A	No	0.703	Yes
Jansen	≥6	405	226	0.585(0.535–0.634)	0.675(0.627–0.721)	0.716(0.669–0.759)	No	0.750(0.704–0.791)	No
Jansen	≥6	351	174	0.534(0.473–0.594)	0.576(0.523–0.629)	0.695(0.644–0.743)	No	0.709(0.658–0.756)	No
Le	N/A	63	26	0.5	0.68	0.76	No	0.77	No
Sokoll	≥10	566	245	0.58(0.53–0.64)	0.66(0.61–0.71)	0.70(0.65–0.75)	No	N/A	No
Stephan	8–12	475	264	0.56 (0.51–0.61)	0.77 (0.73–0.81)	0.78 (0.74–0.82)	No	N/A	No
Sokoll	≥10	123	63	0.52(0.42–0.63)	0.61(0.51–0.71)	0.69(0.60–0.79)	Yes	N/A	No
Catalona	6–10	1091	456	N/A	0.602	0.638	No	N/A	No

Abbreviation: DRE, digital rectal examination; N/A, not available.

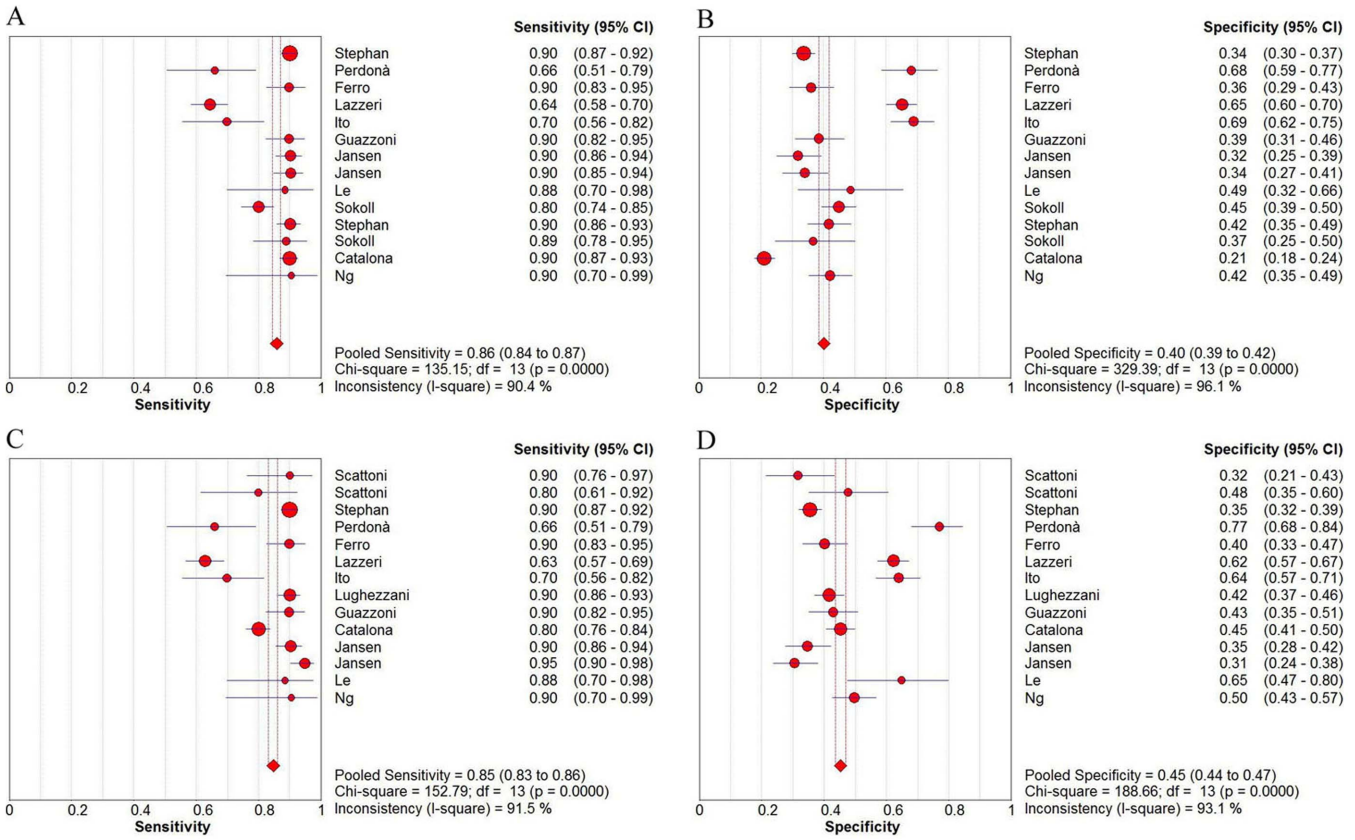

Figure 2 | Forest plot of sensitivities and specificities of %p2PSA and Phi for the diagnosis of PCa. (A) sensitivity for %p2PSA; (B) specificity for %p2PSA; (C) sensitivity for Phi; (D) specificity for Phi.

Discussion

Despite the wide and successful use of PSA blood testing, its use as a serum marker for PCa is far from ideal, due to the lack of specificity; it has only 20% specificity[31]. This review synthesized the current knowledge about early diagnosis of PCa with %p2PSA and Phi determination. The present meta-analysis indicated that %p2PSA and Phi had a high cancer detection rate. Most importantly, %p2PSA and Phi may be useful as tumor markers for high risk PCa detection.

PSA cannot accurately distinguish between benign and malignant pathology in the tPSA range of 2–10 ng/ml. In addition, more indolent PCas were detected, which will not develop into clinical significant PCas[32]. New markers that can accurately detect as well as differentiate patients with aggressive and non-aggressive PCa are needed. According to preliminary investigations and observational studies, %p2PSA and Phi may improve discrimination between men with and without PCa, and they are associated with aggressiveness of PCa, as indicated by Gleason score[6,10–12]. But some studies did not demonstrate a relationship between Gleason score and p2PSA level[7]. So we carried out a meta-analysis to further confirm the ability of %p2PSA and Phi in PCa patient's diagnosis, and derive a more precise estimation of their ability to predicate aggressive PCa.

To date, only one systematic review has been conducted to assess the ability of %p2PSA and Phi to diagnose PCa. Both %p2PSA and Phi improved the accuracy of PCa detection compared to PSA and %fPSA, particularly with the PSA range of 2–10 ng/ml[33]. However, only the papers published before December 2011 were included, and subgroup analysis or meta-regression were not performed to explore the sources of heterogeneity. In addition, the diagnostic accuracy of %p2PSA and Phi for differentiating aggressive PCa from non-aggressive PCa has not yet been summarized. In our meta-analysis,

Table 2 | Subgroup analyses of %p2PSA diagnostic value

Variable	Data sets	Sensitivity (95% CI)	Specificity (95% CI)	Diagnostic OR(95% CI)	I^2, *p*-value	AUC (95% CI)
Total	14	0.86(0.84–0.87)	0.40(0.39–0.42)	4.11(3.46–4.87)	29.1%,0.1448	0.72(0.67–0.77)
Design						
Retrospective	6	0.90(0.88–0.92)	0.30(0.28–0.33)	4.29(2.87–6.42)	61.6%,0.0231	0.96(0.91–1.01)
Prospective	7	0.76(0.73–0.79)	0.54(0.51–0.56)	3.85(3.15–4.71)	0.0%,0.6260	0.70(0.66–0.74)
Mixed	1	0.90	0.34	N/A	N/A	0.72(0.70–0.75)
Ethnicity						
Caucasian	12	0.86(0.85–0.87)	0.38(0.37–0.40)	4.05(3.36–4.87)	36.1%,0.1016	0.71(0.65–0.77)
Asian	2	0.76(0.64–0.85)	0.55(0.50–0.60)	5.37(2.93–9.84)	0.0%,0.7095	0.50(0.50–0.50)
QUADAS						
≥12	8	0.83(0.81–0.85)	0.44(0.42–0.46)	4.00(3.38–4.70)	0.0%, 0.7132	0.69(0.65–0.73)
<12	6	0.89(0.87–0.91)	0.34(0.32–0.36)	4.40(2.98–6.52)	63.7%,0.0171	0.78(0.78–0.89)

Abbreviation: QUADAS, Quality Assessment of Diagnostic Accuracy Studies; N/A, not available.

Table 3 | Subgroup analysis of Phi diagnostic value

Variable	Data sets	Sensitivity(95% CI)	Specificity(95% CI)	Diagnostic OR(95% CI)	I^2, *p*-value	AUC (95% CI)
Total	14	0.85(0.83–0.86)	0.45(0.44–0.47)	4.82(3.90–5.97)	46.4%,0.0289	0.70(0.65–0.74)
Design						
Retrospective	3	0.92(0.89–0.95)	0.39(0.35–0.43)	6.05(3.98–9.18)	0.0%,0.4700	0.66(0.23–1.08)
Prospective	9	0.80(0.77–0.82)	0.52(0.49–0.54)	4.98(3.61–6.87)	51.3%,0.0368	0.71(0.66–0.76)
Mixed	2	0.86(0.84–0.88)	0.39(0.37–0.42)	4.04(2.73–5.97)	71.1%,0.0629	0.50(0.50–0.50)
Ethnicity						
Caucasian	12	0.85(0.83–0.86)	0.44(0.42–0.46)	4.85(3.85–6.11)	52.5%,0.0168	0.69(0.64–0.74)
Asian	2	0.76(0.64–0.85)	0.56(0.51–0.61)	4.76(2.53–8.98)	4.1%, 0.3071	0.50(0.50–0.50)
QUADAS						
≥14	5	0.89(0.87–0.91)	0.43(0.40–0.45)	5.44(4.29–6.89)	0.0%,0.8272	0.78(0.71–0.85)
<14	9	0.82(0.80–0.84)	0.47(0.45–0.49)	4.46(3.34–5.96)	55.9%,0.0201	0.67(0.62–0.71)

QUADAS: Quality Assessment of Diagnostic Accuracy Studies.

articles published in 2012, 2013 and 2014 were also included. To the best of our knowledge, this was the first attempt to synthesize existing literature to evaluate the diagnostic accuracy of %p2PSA and Phi for aggressive PCa.

Continued emergence of promising data indicate that %p2PSA and Phi are better parameters than tPSA and fPSA for the detection of PCa and aggressive PCa[27]. Based on synthesis of the results in our meta-analysis, the diagnostic test may be particularly useful in patients with diagnostically challenging PSA scores ranging 2–10 ng/ml, as sensitivity reaches 85%–86%, and specificity reaches 40%–45% for %p2PSA and Phi respectively. As shown previously, the PSA test had varying sensitivity and poor specificity, only about 20% for specificity[31]. Our results showed that %p2PSA and Phi test have potential to complement PSA screening due to their high specificity. We also observed that %p2PSA and Phi have nearly the same effect in diagnosing PCa. Both %p2PSA and Phi showed a high accuracy for detecting PCa (AUC of 0.72 and 0.70 respectively). SROC of our systematic review showed a curvilinear relationship between sensitivity and 1-specificity, and the Spearman's correlation coefficient was 0.705 ($P = 0.005$) and 0.744 ($P = 0.002$) for %p2PSA and Phi respectively, which may be due to a threshold effect.

The detection of aggressive PCa is the main concern of current discussions on the usefulness of PCa detection markers. Recent studies have shown that higher %p2PSA and Phi values were associated with increased probability of detecting Gleason score ≥ 7 PCa and were higher in aggressive cancers, so %p2PSA and Phi may be able to predict more aggressive PCa on prostatic biopsies[5]. Phi provided the best discrimination between aggressive and non-aggressive PCa than other markers[5]. According to our pooled results, the sensitivity for discriminating between aggressive and non-aggressive PCa was very high, 96% and 90% for %p2PSA and Phi respectively, which would miss a small number of aggressive PCa, but the specificity was relatively low, only 9% and 17% for %p2PSA and Phi respectively. The AUC for Phi (0.67) exceeded that of the %p2PSA (0.54) in discriminating aggressive versus non-aggressive PCa. Increased Phi levels and %p2PSA might be helpful in identifying patients harboring more aggressive forms of the disease. Some cutoffs were used for %p2PSA and Phi, at a threshold of 1.2 and 25.5 for %p2PSA and Phi respectively, about 4% of aggressive cancers would have been missed[8]. And these cutoffs did not allow significant improvement in differentiation between Gleason score 4 + 3 = 7 and 3 + 4 = 7 cancers[5]. Phi values did not differ with age and race[11,24], this suggests that Phi may be more prosperous to be used in a wider scope of all men, irrespective of race, age and background in predicting aggressive PCa. A limited number of studies were included, therefore it was difficult to draw a definitive conclusion for their ability to discriminate. Hence, further research should aim at determining the potential aggressiveness of PCa in patients with high %p2PSA and Phi values.

When interpreting the results, the heterogeneity caused by the different optimum cutoff points determining whether a test is positive or negative must be taken into consideration, since there is actually no consensus regarding the most appropriate level, in part, due to differences in study design and methodology. The cut-off points were only reported in 7 and 9 studies for %p2PSA and Phi

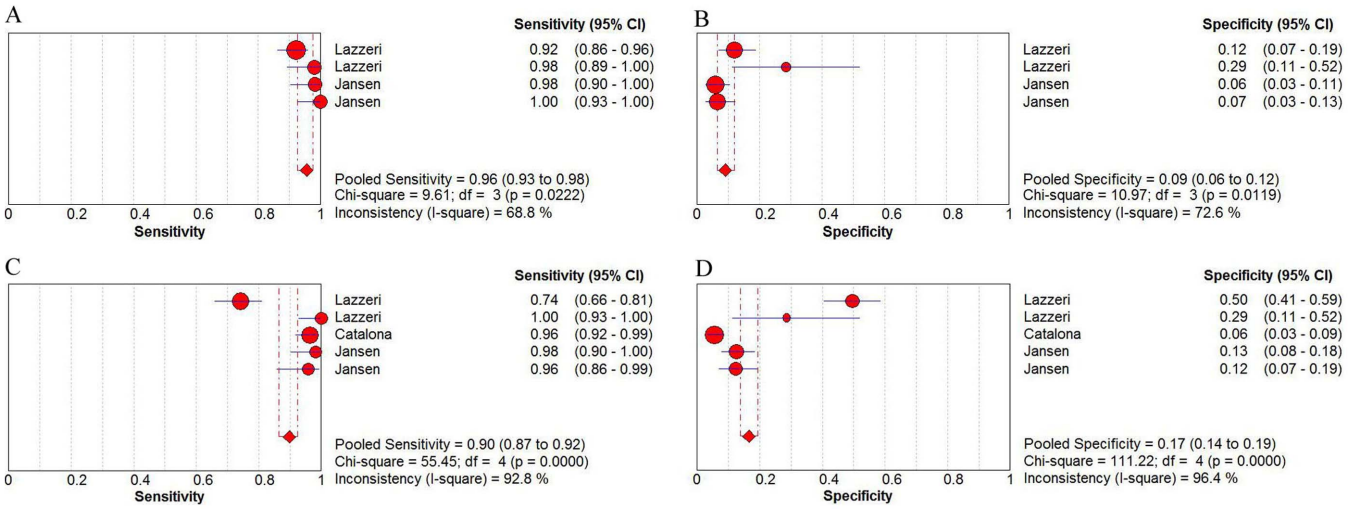

Figure 3 | Forest plot of sensitivities and specificities of %p2PSA and Phi for the diagnosis of aggressive PCa. (A) sensitivity for %p2PSA; (B) specificity for %p2PSA; (C) sensitivity for Phi; (D) specificity for Phi.

Table 4 | Diagnostic accuracy of %p2PSA and Phi for Gleason score ≥7 VS Gleason score<7

Variable	Data sets	Sensitivity (95% CI)	Specificity (95% CI)	Diagnostic OR(95% CI)	I^2, *p*-value	AUC (95% CI)
%p2PSA	4	0.96(0.93–0.98)	0.09(0.06–0.12)	3.77(1.13–12.57)	43.5%,0.1507	0.54(−0.52–1.61)
Phi	5	0.90(0.87–0.92)	0.17(0.14–0.19)	3.06(1.61–5.84)	29.5%,0.2248	0.67(0.57–0.77)

respectively, and the cut-off values were quite different for all the studies. Increased Phi values were associated with increasing PCa detection rate, when the phi values were <25, 25–34.9, 35–54.9 and ≥55, the PCa detection rates were 11%,18.1%, 32.7% and 52.1% respectively[11]. The full clinical potential and the threshold of %p2PSA and Phi for diagnosing PCa and identification of aggressive disease need to be determined. If the maximum number of aggressive PCas were to be detected, it might be acceptable to not detect some of the less aggressive, clinically insignificant PCas.

The present meta-analysis has a number of limitations that must be taken into account. First, the preanalytic and analytic management of blood samples were different in the studies such as the case when archived serum was used in some studies, and different calibrations were used in the literatures, including Hybritech and WHO standardization. PSA and fPSA were measured using Hybritech Tandem assays except for one study. Second, the main bias of all the studies was that the inclusion criteria was based on the risk of having PCa, the final decision on whether to biopsy or not was based on elevated PSA values and not based on the %p2PSA or Phi, which would cause selection bias. Third, only one study was a randomized study, differences in study designs including prospective, retrospective and mixed, population characteristics, cutoff points and the number of cores obtained in prostate biopsies (rang, 6–24) may contribute to the heterogeneity. Fourth, another limitation was potential publication bias, since yet unpublished studies and non-English studies were excluded, however, we did not confirm the presence of publication bias. We should also note that the number of studies included for analysis for the purpose of discriminating aggressive and non-aggressive PCa was relatively low. Due to limited literatures, subgroup-analyses and meta-regression were not performed. Finally, the best strategy to correlate %p2PSA and phi with aggressiveness is to perform a correlation between preoperative %p2PSA, phi and whole gland sample after radical prostatectomy; unfortunately only the paper by Guazzoni et al[12] focused on this topic. Due to the pathological difference between prostate biopsy and sample after radical prostatectomy, we did not include this study into our meta-analysis.

Due to the limitations mentioned above, the results of this meta-analysis should be interpreted with care. Nevertheless, our study was conducted at an appropriate time, because enough data was available to use meta-analytical methods. And we provided the most up-to-date information on this topic. However, it is necessary to conduct the large multi-center randomized-controlled studies or prospective trials with similar methodology to evaluate the ability of %p2PSA and Phi in predicting aggressive PCa in the future. In addition, similar comparison outcomes should be based on the same preanalytic and analytic management for blood samples, same calibration technology (Hybritech or WHO), same TRUS prostate biopsy techniques and biopsy cores or based on the pathology of prostatectomy specimen.

In conclusion, our meta-analysis suggested that both %p2PSA and Phi have high diagnostic accuracy rates and can be used in early PCa diagnosis. In addition, %p2PSA and Phi might be potential biomarkers in predicating aggressive PCa. In our opinion, if the data will be confirmed by further larger sample size studies, the two biomarkers might be helpful in guiding biopsy decisions.

Methods

Literature search and selection criteria. Our meta-analysis was performed following the guidelines concerning Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) statement[34] (Supplementary Table 5). A systematic literature search on PubMed was conducted for articles published from January 2000 to March 2014. The strategy was to use and combine the following terms included in their titles, abstracts, or keywords lists: "prostatic neoplasm", "prostate cancer", "p2PSA", "[−2]proPSA", "[−2]proenzyme prostate specific antigen", "Prostate Health Index", "Phi", "diagnosis", "sensitivity", and "specificity". Additionally, we reviewed the reference list of each relevant paper by hand. Two researchers (W Wang and M Wang) independently reviewed all identified titles, abstracts and manuscripts to determine if a study was suitable for the meta-analysis.

Studies that met the following criteria were included in the meta-analysis: (1) case-control or cohort design; (2) diagnostic test about %p2PSA or Phi for PCa; (3) histological results were based on prostate biopsy; (4) data on sensitivity and specificity could be extracted. Non-English language papers or studies with only abstracts available for information extraction were excluded. When more than one article was published based on the same patient population or same project, only the most recent or complete report was used, to avoid overlapping between cohorts.

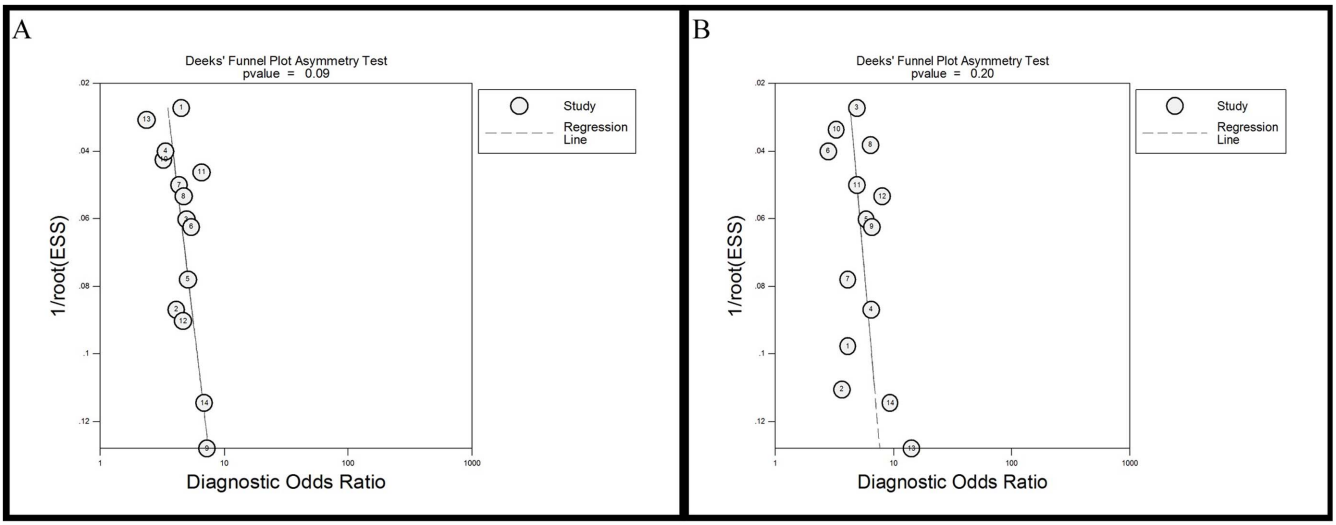

Figure 4 | Linear regression test of funnel plot asymmetry for %p2PSA (A) and Phi (B). The statistically non-significant p value of 0.09 (A) and 0.20 (B) for the slope coefficient suggest symmetry in the data.

Data extraction and quality assessment. We collected information from each eligible study and entered the data into a structured database including: last name of the first author, year of publication, country of origin, study design, ethnicity, age range of participants, the tPSA and fPSA range, sample size, indication for biopsy, type of biopsy used, the number of biopsy-confirmed PCa and the values of true positive (TP), false-positive (FP), false negative (FN), true negative (TN), summary receiver operating curve (SROC) and cut-off points if available. Of these studies, Gleason score ≥ 7 was defined as high Gleason score and aggressive, and Gleason score <7 was defined as low Gleason score and non-aggressive. The study quality was assessed with Quality Assessment of Diagnostic Accuracy Studies (QUADAS)[35,36]. Two authors extracted the data and undertook the quality assessment independently. If disagreements appeared, a third author helped to resolve the problem.

Data analysis and presentation. Based on the collected data, 2×2 contingency tables were created to calculate the indices of diagnostic validity, the main outcome data included pooled sensitivity (Se) and specificity (Sp), each value was determined with 95% confidence intervals. A Forest plot was used to display the results of the meta-analysis. Methodological heterogeneity was assessed during selection.

We determined the existence of a threshold effect, which is the primary cause of heterogeneity in test accuracy studies, defined as the use of different cut-offs or thresholds in different studies to determine a positive test result. The analysis of the diagnostic threshold was assessed using the graphic representation of "Se" vs "1-Sp" on a receiver operating characteristic (ROC) plane and also by calculating the Spearman correlation coefficient. The ROC plane provides a graphic representation of the pairs of Se and Sp; characteristically its points show a typical pattern of "shoulder arm" if the threshold effect exists. The results were synthesized and represented graphically in a forest plot. If there was evidence of a threshold effect, the studies were combined to create a SROC and we calculated an additional measurement of accuracy of the technique (Q*), and obtained the area under the curve (AUC). Possible sources of heterogeneity were assessed through chi-square, the null hypothesis that the studies are homogeneous was rejected with a p value of <0.05, and heterogeneity was quantified using the I^2 statistic which was interpreted as low (25%–50%), moderate (51%–75%) and high (>75%) levels of heterogeneity. The Mantel-Haenszel fixed effect model was used if heterogeneity was not found; otherwise, the Dersimonian-Laird random effect model was used. In addition, subgroup analyses were also conducted by collection of sera (prospective, retrospective), Ethnicities (Caucasian and Asian) and QUADAS score. Potential publication bias was examined using Deeks' funnel plots[37], with $P < 0.05$ for the slope coefficient indicating significant asymmetry. All meta-analyses were performed using MetaDisc (Version 1.4)[38] and STATA (Version 10.0).

1. Sokoll, L. J. *et al.* Proenzyme psa for the early detection of prostate cancer in the 2.5–4.0 ng/ml total psa range: preliminary analysis. *Urology* **61**, 274–6 (2003).
2. Schroder, F. H. & Roobol, M. J. Defining the optimal prostate-specific antigen threshold for the diagnosis of prostate cancer. *Curr Opin Urol* **19**, 227–31 (2009).
3. Hori, S., Blanchet, J. S. & McLoughlin, J. From prostate-specific antigen (PSA) to precursor PSA (proPSA) isoforms: a review of the emerging role of proPSAs in the detection and management of early prostate cancer. *BJU Int* **112**, 717–28 (2013).
4. Khan, M. A. *et al.* Evaluation of proprostate specific antigen for early detection of prostate cancer in men with a total prostate specific antigen range of 4.0 to 10.0 ng/ml. *J Urol* **170**, 723–6 (2003).
5. Stephan, C. *et al.* Multicenter evaluation of [−2]proprostate-specific antigen and the prostate health index for detecting prostate cancer. *Clin Chem* **59**, 306–14 (2013).
6. Lazzeri, M. *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* **188**, 1137–43 (2012).
7. Jansen, F. H. *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* **57**, 921–7 (2010).
8. Lazzeri, M. *et al.* Clinical performance of serum prostate-specific antigen isoform [−2]proPSA (p2PSA) and its derivatives, %p2PSA and the prostate health index (PHI), in men with a family history of prostate cancer: results from a multicentre European study, the PROMEtheuS project. *BJU Int* **112**, 313–21 (2013).
9. Lazzeri, M. *et al.* Relationship of chronic histologic prostatic inflammation in biopsy specimens with serum isoform [−2]proPSA (p2PSA), %p2PSA, and prostate health index in men with a total prostate-specific antigen of 4–10 ng/ml and normal digital rectal examination. *Urology* **83**, 606–12 (2014).
10. Guazzoni, G. *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* **60**, 214–22 (2011).
11. Catalona, W. J. *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* **185**, 1650–5 (2011).
12. Guazzoni, G. *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* **61**, 455–66 (2012).
13. Stephan, C. *et al.* 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. *Clin Chem* **59**, 280–8 (2013).
14. Ferro, M. *et al.* Predicting prostate biopsy outcome: prostate health index (phi) and prostate cancer antigen 3 (PCA3) are useful biomarkers. *Clin Chim Acta* **413**, 1274–8 (2012).
15. Mikolajczyk, S. D. *et al.* Proenzyme forms of prostate-specific antigen in serum improve the detection of prostate cancer. *Clin Chem* **50**, 1017–25 (2004).
16. Loeb, S. *et al.* Prospective multicenter evaluation of the Beckman Coulter Prostate Health Index using WHO calibration. *J Urol* **189**, 1702–6 (2013).
17. Lughezzani, G. *et al.* Multicenter European External Validation of a Prostate Health Index-based Nomogram for Predicting Prostate Cancer at Extended Biopsy. *Eur Urol* (2013).
18. Fossati, N. *et al.* Clinical performance of serum isoform [−2]proPSA (p2PSA) and its derivatives, namely %p2PSA and PHI (Prostate Health Index) in men younger than 60 years of age: results from a multicentric European study. *BJU Int* (2014).
19. Lughezzani, G. *et al.* Development and internal validation of a Prostate Health Index based nomogram for predicting prostate cancer at extended biopsy. *J Urol* **188**, 1144–50 (2012).
20. Scattoni, V. *et al.* Head-to-head comparison of prostate health index and urinary PCA3 for predicting cancer at initial or repeat biopsy. *J Urol* **190**, 496–501 (2013).
21. Perdona, S. *et al.* Prostate health index (phi) and prostate cancer antigen 3 (PCA3) significantly improve diagnostic accuracy in patients undergoing prostate biopsy. *Prostate* **73**, 227–35 (2013).
22. Ferro, M. *et al.* Prostate Health Index (Phi) and Prostate Cancer Antigen 3 (PCA3) significantly improve prostate cancer detection at initial biopsy in a total PSA range of 2–10 ng/ml. *PLoS One* **8**, e67687 (2013).
23. Lazzeri, M. *et al.* Serum isoform [−2]proPSA derivatives significantly improve prediction of prostate cancer at initial biopsy in a total PSA range of 2–10 ng/ml: a multicentric European study. *Eur Urol* **63**, 986–94 (2013).
24. Ito, K. *et al.* Diagnostic significance of [−2]pro-PSA and prostate dimension-adjusted PSA-related indices in men with total PSA in the 2.0-10.0 ng/mL range. *World J Urol* **31**, 305–11 (2013).
25. Le, B. V. *et al.* [−2]Proenzyme prostate specific antigen is more accurate than total and free prostate specific antigen in differentiating prostate cancer from benign disease in a prospective prostate cancer screening study. *J Urol* **183**, 1355–9 (2010).
26. Sokoll, L. J. *et al.* A prospective, multicenter, National Cancer Institute Early Detection Research Network study of [−2]proPSA: improving prostate cancer detection and correlating with cancer aggressiveness. *Cancer Epidemiol Biomarkers Prev* **19**, 1193–200 (2010).
27. Stephan, C. *et al.* A [−2]proPSA-based artificial neural network significantly improves differentiation between prostate cancer and benign prostatic diseases. *Prostate* **69**, 198–207 (2009).
28. Sokoll, L. J. *et al.* [−2]proenzyme prostate specific antigen for prostate cancer detection: a national cancer institute early detection research network validation study. *J Urol* **180**, 539–43; discussion 543 (2008).
29. Catalona, W. J. *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* **170**, 2181–5 (2003).
30. Ng, C. F. *et al.* The Prostate Health Index in predicting initial prostate biopsy outcomes in Asian men with prostate-specific antigen levels of 4–10 ng/mL. *Int Urol Nephrol* (2013).
31. van Leeuwen, P. J. *et al.* Prostate cancer mortality in screen and clinically detected prostate cancer: estimating the screening benefit. *Eur J Cancer* **46**, 377–83 (2010).
32. Mikolajczyk, S. D., Marks, L. S., Partin, A. W. & Rittenhouse, H. G. Free prostate-specific antigen in serum is becoming more complex. *Urology* **59**, 797–802 (2002).
33. Filella, X. & Gimenez, N. Evaluation of [−2] proPSA and Prostate Health Index (phi) for the detection of prostate cancer: a systematic review and meta-analysis. *Clin Chem Lab Med* **51**, 729–39 (2013).
34. Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* **339**, b2700 (2009).
35. Whiting, P., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. M. & Kleijnen, J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* **3**, 25 (2003).
36. Whiting, P. *et al.* Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* **8**,1–234 (2004).
37. Deeks, J. J., Macaskill, P. & Irwig, L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* **58**, 882–93 (2005).
38. Zamora, J., Abraira, V., Muriel, A., Khan, K. & Coomarasamy, A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* **6**, 31 (2006).

Acknowledgments

No funding was used for this project.

Author contributions

Conceived and designed the experiments: Y.T., J.F.X. Performed the experiments: W.Y.W.,

M.L.W. Analyzed the data: L.W., M.L.W. Contributed reagents/material/analysis tools: T.S.A., L.W., Wrote the manuscript: W.Y.W., M.L.W., Reference collection and data management: Y.T., T.S.A.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wang, W.Y. *et al*. Diagnostic ability of %p2PSA and prostate health index for aggressive prostate cancer: a meta-analysis. *Sci. Rep.* **4**, 5012; DOI:10.1038/srep05012 (2014).