

Keywords: prostate cancer; biomarker; 450K Infinium Methylation BeadChip; DNA methylation; gene expression

DNA methylation status is more reliable than gene expression at detecting cancer in prostate biopsy

A Paziewska¹, M Dabrowska², K Goryca², A Antoniewicz³, J Dobruch⁴, M Mikula², D Jarosz^{1,2}, L Zapala³, A Borowka⁴ and J Ostrowski^{*,1,2}

¹Department of Gastroenterology and Hepatology, Medical Center for Postgraduate Education, 01-813 Warsaw, Poland;

²Department of Genetics, Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, 02-781 Warsaw, Poland;

³Department of Urology, Multidisciplinary Hospital Warsaw-Miedzylesie, 04-749 Warsaw, Poland and ⁴Department of Urology, Medical Center for Postgraduate Education, 01-813 Warsaw, Poland

Background: We analysed critically the potential usefulness of RNA- and DNA-based biomarkers in supporting conventional histological diagnostic tests for prostate carcinoma (PCa) detection.

Methods: Microarray profiling of gene expression and DNA methylation was performed on 16 benign prostatic hyperplasia (BPH) and 32 cancerous and non-cancerous prostate samples extracted by radical prostatectomy. The predictive value of the selected biomarkers was validated by qPCR-based methods using tissue samples extracted from the 58 prostates and, separately, using 227 prostate core biopsies.

Results: HOXC6, AMACR and PCA3 expression showed the best discrimination between PCa and BPH. All three genes were previously reported as the most promising mRNA-based markers for distinguishing cancerous lesions from benign prostate lesions; however, none were sufficiently sensitive and specific to meet the criteria for a PCa diagnostic biomarker. By contrast, DNA methylation levels of the APC, TACC2, RARB, DGKZ and HES5 promoter regions achieved high discriminating sensitivity and specificity, with area under the curve (AUCs) reaching 0.95 – 1.0. Only a small overlap was detected between the DNA methylation levels of PCa-positive and PCa-negative needle biopsies, with AUCs ranging between 0.854 and 0.899.

Conclusions: DNA methylation-based biomarkers reflect the prostate malignancy and might be useful in supporting clinical decisions for suspected PCa following an initial negative prostate biopsy.

Although histological evaluation remains the gold standard of cancer diagnosis, molecular diagnostics may complement conventional histopathological analyses. Diagnostic biomarkers should be sensitive and specific enough to show whether disease is present or absent; however, the levels of most reported biomarkers overlap between tumour and normal tissues. Consequently, it is difficult to identify an ideal biomarker with enough discriminatory power for an accurate diagnosis.

Prostate cancer (PCa) is the leading cause of cancer-related morbidity and mortality in the Polish population (Didkowska *et al*,

2012). The suspicion of PCa results from elevated levels of serum prostate-specific antigen (PSA) and palpable alterations within the prostate upon digital rectal examination (DRE). The examination can be further supported by more specific imaging methods, including transrectal ultrasound (TRUS) and magnetic resonance imaging scans. All of these diagnostic tests are used in PCa prevention programs. Although the final diagnosis of PCa is based on histological evaluation of the prostatic tissue, the most accurate diagnosis is achieved by histopathological examination of the surgically removed gland. This approach, however, cannot be

*Correspondence: Professor J Ostrowski; E-mail: jostrow@warman.com.pl

Received 8 March 2014; revised 24 April 2014; accepted 20 May 2014; published online 17 June 2014

© 2014 Cancer Research UK. All rights reserved 0007 – 0920/14

applied to each patient with a suspicion of cancer. Instead, TRUS-guided prostate core biopsy is recognised as a standard procedure for PCa diagnosis (Serefoglu *et al*, 2013).

The systemic sextant biopsy method introduced by Hodge *et al* (1989) revolutionised the early diagnosis of PCa, but the inherent sampling error of classical needle-directed biopsies results in substantial numbers of missed cancers, especially in younger patients and patients with a larger prostate (Vashi *et al*, 1998; Babaian *et al*, 2000; Eskicorapci *et al*, 2005). Although 12-, 18-, 20- and 24-core prostate biopsies (Pepe and Aragona, 2007; Ravery *et al*, 2008; Scattoni *et al*, 2008, 2010) gradually improve the PCa detection rate, the risk of complications increases with the number of biopsies, including bleeding, urinary obstruction, vasovagal reaction and infection (Anastasiadis *et al*, 2013; Serefoglu *et al*, 2013). Consequently, some authors advise saturation biopsy of the prostate as the next diagnostic step after an initial negative biopsy but when prostate cancer is strongly suspected (Rabets *et al*, 2004; Jones *et al*, 2006). If so, it is important to objectify and standardise the procedure by which patients with suspected PCa and an initial negative biopsy are selected for further biopsy.

TRUS-guided 10- to 12-core biopsies may miss up to 30% of PCa (Jones *et al*, 2006). Furthermore, the false-negative rate of the 6- and 12-core biopsy techniques performed *ex vivo* on prostates removed due to biopsy-proven cancer was >25% (Svetec *et al*, 1998; Fink *et al*, 2001; Serefoglu *et al*, 2013). Thus, under-diagnosis of PCa is not uncommon and, therefore, standard hematoxylin and eosin (H&E) staining of core biopsies could be assisted by monitoring differentially expressed genes (Hessels *et al*, 2004; Bhavsar *et al*, 2013) and patterns of DNA methylation (Kim *et al*, 2011a; Day and Bianco-Miotto, 2013) as an initial biopsy strategy.

Here, we analysed critically the potential usefulness of selected biomarkers for supporting conventional histological tests for PCa diagnosis. The strategy included microarray-based expression studies and genome-wide DNA methylation profiling followed by multiplex PCR-based techniques to validate biomarker association with tumour occurrence. The final aim was to establish the sensitivity and specificity of the new molecular markers for detecting tumour even in the absence of dysplastic cells in a needle biopsy. Such biomarkers might facilitate the decision between watchful waiting or an additional biopsy in cases of suspected PCa with an initial negative biopsy.

MATERIALS AND METHODS

Patients were prospectively selected for the study at the Department of Urology, Medical Center for Postgraduate Education, Warsaw, Poland. None received any previous hormonal therapy or radiotherapy to the prostate. The study protocol was approved by the Medical Center Bioethics Committee, and all patients signed informed consent before inclusion.

Prostate tissue samples were collected from 58 patients who underwent radical prostatectomy. The prostates were sectioned immediately after surgical removal and two suitable tissue pieces were cut from tumoral and non-tumoral areas in PCa patients who underwent radical prostatectomy and from normal areas in benign prostatic hyperplasia (BPH) patients who underwent simple prostatectomy. The amount of specimen removed did not interfere with the histological assessment of the remaining prostate. In another group of 227 patients with suspected PCa (based on DRE and/or elevated serum PSA (PSA >4 ng ml⁻¹) levels) who were subjected to the initial biopsy approach, sets of 13 TRUS-guided core biopsies were extracted using standard techniques and 18G needles. The tissues removed from the surgically resected prostates and one random (out of 13) of the TRUS-guided core biopsies

from each biopsy set were frozen in RNALater (Qiagen, Hilden, Germany) and stored at -80 °C until further processing.

The remaining tissues from the surgically resected prostates and needle biopsies were embedded into paraffin according to routine procedures. Sections were stained with H&E and examined for PCa diagnosis and Gleason scoring by two experienced pathologists in a blinded fashion. Histological examination performed after prostatectomy resulted in a diagnosis of PCa in 39 patients (median age, 63 years; range: 49–82) and BPH in 19 patients (median age, 68 years; range: 58–86). Histological examination of core biopsies resulted in a diagnosis of PCa in 131 patients (median age, 71 years; range: 41–89): 87 tumours were graded as well- or moderately-well differentiated (Gleason 4–7) and 44 as poorly differentiated (Gleason 8–10). The presence of high-grade prostatic intra-epithelial neoplasia (PIN) was recorded in core biopsies from 37 patients (median age, 70 years; range: 55–81). The remaining 59 patients were classified as having benign disease (median age, 66 years; range: 51–85).

RNA and DNA extraction. To obtain histologically well-characterised samples for molecular analyses, macrodissection of specimens removed after prostatectomy was performed as described previously (Skrzypczak *et al*, 2010). Several series of cryostat sections were prepared from different parts of all malignant, non-malignant and normal prostate areas using a Microm HM 505E (Zeiss, Oberkochen, Germany). The upper and lower sections from each cryosection were evaluated by an experienced pathologist, and the nucleic acids were isolated from cryostat sections with known epithelial cell content.

Total RNA and DNA were isolated from surgically resected prostate specimens using the MirVana miRNA Isolation Kit (Ambion, Carlsbad, CA, USA), and the QIAamp DNA Mini Kit (Qiagen), respectively, following manufacturer instructions. RNA quality and quantity were analysed in a NanoDrop spectrophotometer. Samples with A260/A280 ratios of 1.8–2.1 were assessed using an Agilent 2100 Bioanalyzer. Each sample used for microarray analysis had a RNA integrity number of 7.6–9.6. Trizol reagent (Invitrogen, Carlsbad, CA, USA) was used to isolate RNA and DNA from each core biopsy. Briefly, the biopsy was grinded with disposable pestle (Kontes; 749520-0590) in Trizol reagent, then chloroform was added and phases were separated by centrifugation at 12 000 RPM for 15 min at 4 °C. Aqueous phase containing RNA was further processed with RNeasy Mini Kit (Qiagen) with on-column DNA digestion. DNA was isolated by mixing interphase/organic phase with ethanol followed by phenol/ethanol supernatant removal and further DNA recovery with a QIAamp DNA Mini Kit (Qiagen).

Gene expression microarray analysis. Whole-transcriptome profiling was performed by AROS Applied Biotechnology services using a HT-12 v4 Expression BeadChip (Illumina, San Diego, CA, USA). The average bead signals from the chip were quantile-normalised with no background correction. All computations were done using R 2.15.0 software with the Bioconductor extension (Gentleman *et al*, 2004). Measurements were filtered according to the ratio between the interquartile range (IQR) and the median. Only probes with IQR/median > 1/8 were selected for analysis, and 22 820/47 231 probes passed filtering. Genes showing differential expression were selected according to the *P*-value from the *t*-test (Welch variant) after correction for multiple hypothesis testing with the Benjamini–Hochberg algorithm. Adjusted *P*-values < 0.05 were considered significant.

DNA methylation assays. Whole-genome DNA methylation profiling was performed by AROS Applied Biotechnology services using a 450K Infinium Methylation BeadChip (Illumina). Genomic DNAs were bisulfite-converted using the EZ-96 DNA Methylation kit (Zymo Research, Orange, CA, USA). Probes with

missing intensity signals were discarded (21 098 probes out of 485 577). Probes were divided into design classes and annotated to genome location according to the IlluminaHumanMethylation450k.db library. Intensities were normalised with the BMIQ package (version 1.1; Teschendorff *et al*, 2013) using default parameters. Probes mapping to differentially methylated sites were selected according to the *P*-value from the *t*-test (Welch variant) after correction for multiple hypothesis testing with the Benjamini–Hochberg algorithm.

Quantitative(q) RT-PCR. qRT-PCR was performed as described previously (Mikula *et al*, 2011) using SYBR Green chemistry. The geometric mean expression of RPLP0 and UBC mRNA was used as a normalisation factor after evaluation of housekeeping genes with GeNorm software (Vandesompele *et al*, 2002). Gene expression results were calculated using the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). Primers were from PrimerBank (Wang *et al*, 2011). The results were analysed with the Mann–Whitney U test using GraphPad Prism 5 software (GraphPad Software Inc., La Jolla, CA, USA). *P*-values <0.05 were considered significant.

Methylation-Specific (MS)-qPCR. DNA (1 μ g) was bisulfite-converted with the EpiTect Bisulfite Kit (Qiagen) according to the manufacturer's protocol. Primers spanning methylated cytosines were designed with Methyl Primer Express (Life Technologies, Carlsbad, CA, USA) and are available upon the request. qPCR reactions were performed on an ABI 7900HT thermocycler equipped with a 384-well block in 5- μ l-reaction mixtures containing 2.5 μ l 2X SensiMix SYBR Hi-ROX (Bioline, London, UK), 2- μ l DNA (10 ng) and 50-nM primers. qPCR was carried out for 40 cycles consisting of 15 s of denaturation at 95 °C and hybridisation for 1 min at 60 °C in a 384-well reaction plate. Results were expressed as percentages of the methylation level of fully methylated DNA from the EpiTect PCR Control DNA Set (Qiagen) normalised to input DNA.

Predictors based on multiple gene expression/methylation.

Linear models combining expression/methylation of multiple genes were fitted for the first validation sets (qRT-PCR/MS-qPCR) to create appropriate classifiers. Calculations were performed independently for cancerous and adjacent non-cancerous sample sets, resulting in two sets of predictions in each case. Models fitted to the first data sets were applied for the second, validation data

sets. Values predicted by models for each sample were used to create ROC curves. Four ROC curves were created for each expression and methylation: models fitted to cancerous and adjacent non-cancerous sample sets applied to the first and the second validation data sets.

RESULTS

To select biomarkers that can distinguish between BPH and cancerous prostatic tissues, we conducted a two-step selection strategy. The discovery step selected genes differentially expressed and differentially methylated between non-malignant, BPH and malignant specimens removed from tumoral prostates. Standard Illumina array technologies were used, followed by qPCR. The validation step compared the expression and methylation of the selected genes in prostate core biopsies.

Profiling of gene expression. Altogether, 48 RNA samples represented 16 surgically removed glands with benign disease, BPH and 16 pairs of matched cancerous and non-cancerous tissue samples extracted from the prostates with malignant pathology. The presence of at least 40% dysplastic cells (except two samples with 10% dysplastic cells) in the cancerous tissues and the absence of dysplastic cells in the non-cancerous tissues were ascertained by histological examination. RNA samples were hybridised on the Human HT-12 v4 Expression BeadChip microarray.

The pair-wise comparison between BPH and the combined cancerous and non-cancerous samples selected from cancer-bearing prostates identified 12 genes showing at least a threefold change (FC) in expression and an adjusted *P*-value <0.05. None of these genes differentiated between cancerous and non-cancerous tissue samples (Table 1).

The expression status of the 12 genes selected by microarray (Table 1) and that of PCA3, a known biomarker of PCa (Rittenhouse *et al*, 2013), was validated by qRT-PCR using RNA extracted from 19 BPH samples, 20 cancerous prostate samples with a dysplastic cell content <10%, 22 cancerous prostate samples with a dysplastic cell content ranging between 30% and 90%, and 27 adjacent non-cancerous prostate samples. Seven were positive biomarkers (showing elevated expression in cancerous prostate) and six were negative (showing reduced expression in

Table 1. Adjusted *P*-values (pVals Adj) and fold changes (FCs) of genes selected by the microarray study

| Gene name | Gene ID | PCa vs BPH | | PCa vs PCa 0% | |
|---|----------|------------|------|---------------|-------|
| | | pVals Adj | FCs | pVals Adj | FCs |
| Claudin 3 | CLDN3 | 8.08E-05 | 3.08 | 0.459 | 1.23 |
| T-cell receptor gamma alternate reading frame protein | TARP | 0.00094 | 3.24 | 0.0914 | 1.69 |
| Chromosome 15 open reading frame 48 | C15orf48 | 0.00363 | 3.48 | 0.355 | 1.73 |
| Alpha-methylacyl-CoA racemase | AMACR | 0.0371 | 5.42 | 0.0973 | 10.34 |
| Homeobox C6 | HOXC6 | 0.0204 | 3.11 | 0.183 | 3.22 |
| Interleukin 6 | IL6 | 0.0165 | 4.57 | 0.446 | 0.53 |
| LIM domain only 3 (rhombotin-like 2) | LMO3 | 0.00424 | 0.28 | 0.364 | 0.78 |
| Chemokine (C-X-C motif) ligand 13 | CXCL13 | 0.00753 | 0.20 | 0.945 | 1.05 |
| Bone morphogenetic protein 5 | BMP5 | 0.00810 | 0.22 | 0.818 | 0.92 |
| Prostaglandin-endoperoxide synthase 1 | PTGS1 | 0.00934 | 0.21 | 0.307 | 0.75 |
| Ring finger protein 112 | RNF112 | 0.0222 | 0.33 | 0.0670 | 0.73 |
| NEL-like 2 (chicken) | NELL2 | 0.0255 | 0.15 | 0.480 | 0.68 |

Abbreviations: BPH = benign prostatic hyperplasia; PCa = prostate cancer; PCa 0% = non-cancerous samples from cancer-bearing prostates.

cancerous prostate). The gene expression levels and the FCs measured by qRT-PCR were consistent with the microarray results. The expression of all 13 genes was highly distinguishable between BPH and cancer-bearing prostate ($P < 0.0001$), independently of the dysplastic cell content (Supplementary Figure S1).

Receiver operating characteristic (ROC) curves and area under the curve (AUC) were calculated using qRT-PCR expression values. The AUCs represent an appropriate assessment of the relationship between the sensitivity and the specificity of a biomarker over all cutoffs (Hsu *et al.*, 2014). As shown in Figure 1, the expression of nine genes (AMACR, HOXC6, PCA3, TARP, C15orf48, LMO3, NELL2, PTGS1 and RNF112) was highly discriminating between BPH specimens and cancerous prostate specimens cut from surgically resected glands, with AUCs ranging between 0.909 and 0.98. The AUC values of the other four genes (CLDN3, BMP5, IL6 and CXCL13) ranged between 0.835 and 0.894. Similar analysis of the differentially expressed genes between BPH specimens and non-cancerous parts of the glands-exhibiting malignant pathology also showed moderate to high discriminative properties, with AUCs ranging between 0.671 and 0.943.

The second validation step was performed by qRT-PCR with RNA samples isolated from one of the 13 TRUS-guided core biopsies obtained from 131, 59 and 37 patients with histologically diagnosed PCa, BPH and high-grade PIN, respectively. As shown in Supplementary Figure S2, the expression of 11 genes differed between benign and malignant prostate pathology ($P < 0.0001$), whereas that of only 2 genes (PCA3 and HOXC6) differed between simple BPH and high-grade PIN. The AUC-ROC values confirmed

the rather moderate discriminative ability of gene expression, with an AUC > 0.8 for HOXC6, AMACR and PCA3 only (Figure 1). Combination of expression of HOXC6, AMACR and PCA3 resulted in AUC equal 0.887 (Figure 2, yellow line).

Profiling of DNA methylation. Genome-wide DNA methylation profiling was conducted using the Illumina Infinium Human Methylation 450 BeadChip and the same tissue specimens used for gene expression profiling: 16 BPH samples and 32 cancerous and non-cancerous prostate samples. Candidate gene transcription start regions were selected according to adjusted P -values and differences in the methylated fraction between BPH specimens and malignant/non-malignant prostate specimens. A total of 18 256/12 339 CpG sites showed statistical significance with adjusted P -values $< 0.05/0.01$ and ≥ 0.12 mean methylation difference between BPH and PCa. Of these, 1517/1019 targeted methylation sites were located in close proximity to transcription start sites (TSS) (between TSS and 200 base pairs upstream of TSS (TSS200 region)). For the validation study, we selected eight TSS200 regions for which significant ($P < 0.05$) differences in methylation levels were observed at all CpG sites mapped, and in which differences in methylated fractions were > 0.12 (Table 2).

Of note, the expression of genes selected for methylation-based classifiers assessed on transcript level using microarrays was not significantly altered (Supplementary Table S1). Only single probe for RARB gene reached $P < 0.05$ level and none of the FCs was > 1.5 .

The methylation status of the selected gene promoter regions (Table 2) was validated by MS-qPCR using genomic DNA extracted from 14 BPH samples, 15 cancerous prostate

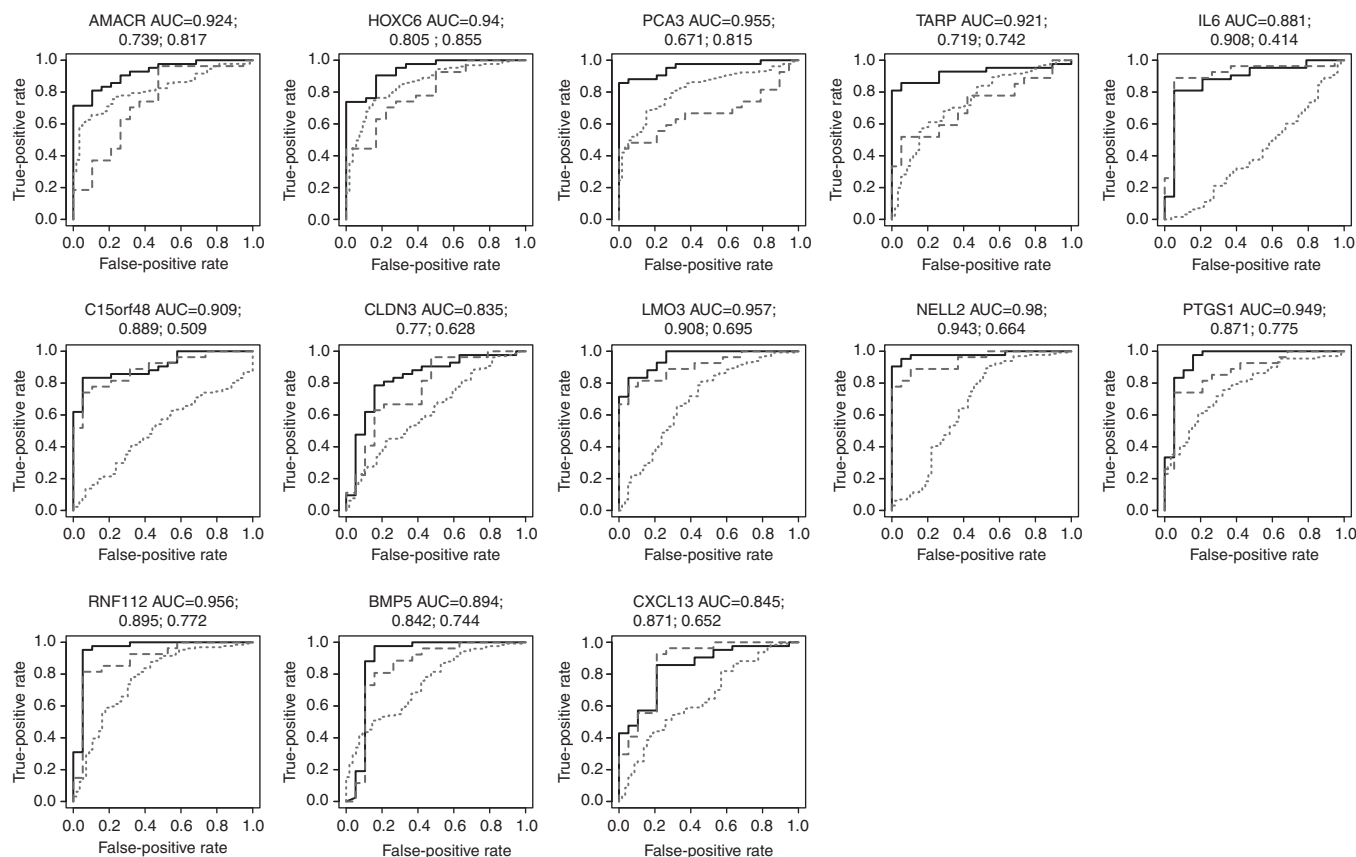


Figure 1. ROC (receiver operating characteristic) curves for classifiers based on single gene expression. A solid line was plotted for discovery phase sets comprising 19 BPH samples and 42 cancerous prostate samples (with low and high dysplastic cell content). A dashed blue line was plotted for BPH samples compared with 27 adjacent non-cancerous prostate samples. A dotted red line was plotted for a validation set of samples comprising 131 and 59 patients histologically diagnosed with PCa and BPH, respectively. AUC values are given in the same order. The full colour version of this figure is available at *British Journal of Cancer* online.

samples with low dysplastic cell content, 15 cancerous prostate samples with high dysplastic cell content and 26 adjacent non-cancerous prostate samples. The methylation levels of seven regions were significantly higher ($P < 0.0001$) in cancerous tissues than in BPH samples, and six were also higher than in adjacent non-cancerous prostate tissues ($P < 0.001 - P < 0.0001$) (Supplementary Figure S3). As shown by AUC-ROC, differences in the methylation of six regions (APC, TACC2, RARB, DGKZ, C5orf4 and HES5) appeared to be highly discriminating between BPH specimens and cancerous prostate specimens, with AUCs ranging between 0.95 and 1.0 (Figure 3). The AUC values calculated for differences in the methylation of the APC, TACC2,

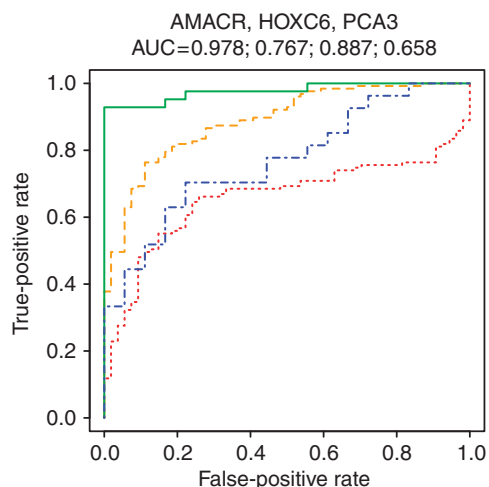


Figure 2. ROC (receiver operating characteristic) curves for classifiers based on linear combination of AMACR, HOXC6 and PCA3 expression. A solid green line was plotted for discovery phase sets comprising 19 BPH samples and 42 cancerous prostate samples (with low and high dysplastic cell content)—version 1. A dashed blue line was plotted for BPH samples compared with 27 adjacent non-cancerous prostate samples—version 2. A dashed yellow line was plotted for a validation set of samples comprising 131 and 59 patients histologically diagnosed with PCa and BPH, respectively, fitted with parameters from version 1 model. A dotted red line was plotted for validation set of samples fitted with parameters from version 2 model. AUC values are given in the same order.

RARB, mir10B, DGKZ, C5orf4 and HES5 regions between BPH and non-cancerous parts of glands—exhibiting malignant pathology ranged between 0.555 and 0.89.

The second validation step performed by MS-qPCR on core biopsies from 131, 59 and 36 patients with histologically diagnosed PCa, BPH and high-grade PIN, respectively, showed highly significant differences in the methylation status of the APC, TACC2, RARB, DGKZ, C5orf4 and HES5 regions between BPH and PCa ($P < 0.0001$). The methylation status of four regions (APC, TACC2, DGKZ and C5orf4) also distinguished BPH from high-grade PIN ($P < 0.05 - P < 0.001$) (Supplementary Figure S4). The methylation levels demonstrated an unexpectedly small overlap between core biopsies extracted from prostates classified as PCa negative and PCa positive based on microscopic examination of H&E-stained biopsy tissues. Consequently, AUC-ROC values confirmed the high discriminatory ability of methylation status, with AUCs ranging between 0.854 and 0.899 for the APC, TACC2, RARB, DGKZ and HES5 regions (Figure 3). Combination of methylation levels of RARB, HES5 and C5orf4 resulted in AUC equal 0.909 (Figure 4, yellow line). Other three gene combinations performed worse (AUC below 0.9, data not shown).

DISCUSSION

Prostate Carcinoma is the most common cancer and leading cause of cancer death in men in different populations. It is a heterogeneous disease with regard to clinical course and molecular alterations (Al-Maghrebi *et al*, 2012). The tests used in clinical practice to diagnose PCa, including the measurement of serum PSA, DRE and TRUS-guided biopsies, have significant limitations (Javed and Langley, 2013). More than 60% of biopsies conducted based on PSA test results ultimately turn up negative (Altintas *et al*, 2013). Although the practical use of molecular markers is still questionable for some clinicians, others believe that newly established biomarkers might complement clinical parameters.

Tumour-associated RNA- and DNA-based markers are typically identified by high-throughput methods and subsequently validated by standard molecular methods such as qPCR. A similar approach was used in this study. The selection of potential biomarkers was conducted using microarray profiling of gene expression and DNA methylation in prostate specimens extracted from glands after

Table 2. Gene promoters selected for the validation phase

| Symbol | Gene ID | Gene name | Bmin TSS200 | Bmax TSS200 | pmin | TSS200 probes | TSS200 significant probes |
|--------|--------------|--|-------------|-------------|----------|---------------|---------------------------|
| APC | NM_000038 | Adenomatous polyposis coli | 0.144 | 0.24 | 0.00017 | 9 | 9 |
| mir10B | NR_029609 | MicroRNA 10b (MIR10B) | 0.190 | 0.26 | 6.06E-05 | 6 | 6 |
| TACC2 | NM_006997 | Transforming, acidic coiled-coil containing protein 2 | 0.139 | 0.30 | 0.00014 | 5 | 5 |
| HES5 | NM_001010926 | Hairy and enhancer of split 5 (Drosophila) | 0.21 | 0.28 | 7.69E-05 | 4 | 4 |
| DGKZ | NM_001105540 | Diacylglycerol kinase, zeta | 0.164 | 0.191 | 0.00010 | 3 | 3 |
| RARB | NM_000965 | Retinoic acid receptor, beta | 0.128 | 0.198 | 0.00012 | 3 | 3 |
| ITGB2 | NM_000211 | Integrin, beta 2 (complement component 3 receptor 3 and 4 subunit) | 0.148 | 0.22 | 0.0040 | 3 | 3 |
| FAXDC2 | NM_032385 | Fatty acid hydroxylase domain containing 2 | 0.125 | 0.139 | 0.0069 | 2 | 2 |

Bmin and Bmax denote the lowest and the highest difference, respectively, between observed methylation rates in prostates resected from PCa patients and BPH patients for all probes mapping to the TSS200 region of a given gene. pmin denotes the lowest *P*-value (corrected) among all probes mapping for the given TSS200 region.

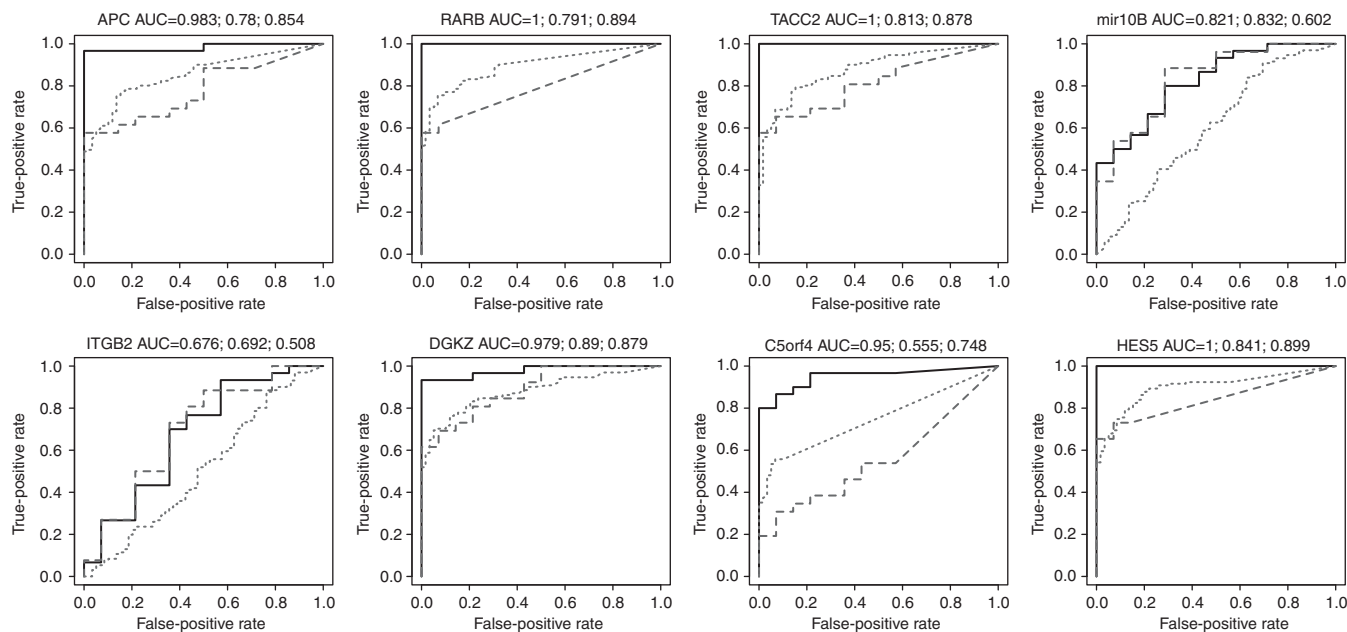


Figure 3. ROC (receiver operating characteristic) curves for classifiers based on single gene methylation. A solid line was plotted for discovery phase sets comprising 14 BPH samples and 30 cancerous prostate samples (with low and high dysplastic cell content, respectively). A dashed blue line was plotted for BPH samples compared with 26 adjacent non-cancerous prostate samples. A dotted red line was plotted for a validation set of samples consisting of 131 and 59 patients histologically diagnosed with PCa and BPH, respectively. AUC values are given in the same order. The full colour version of this figure is available at *British Journal of Cancer* online.

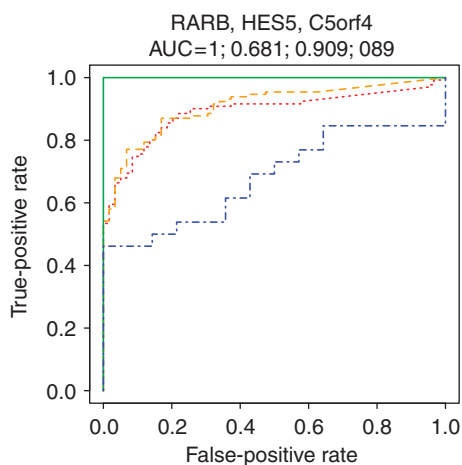


Figure 4. ROC (receiver operating characteristic) curves for classifiers based on linear combination of RARB, HES5 and C5orf4 methylation. A solid green line was plotted for discovery phase sets comprising 14 BPH samples and 30 cancerous prostate samples (with low and high dysplastic cell content)—version 1. A dashed blue line was plotted for BPH samples compared with 26 adjacent non-cancerous prostate samples—version 2. A dashed yellow line was plotted for a validation set of samples comprising 131 and 59 patients histologically diagnosed with PCa and BPH, respectively, fitted with parameters from version 1 model. A dotted red line was plotted for validation set of samples fitted with parameters from version 2 model. AUC values are given in the same order.

prostatectomy. We decided to select candidate genes among those, that differentiated both cancerous prostate samples and adjacent non-cancerous prostate samples from BPH to utilise the chance of hitting tumour with an extra biopsy or detecting the effect of

adjacent tumour on surrounding normal tissue. Next, the predictive value of the selected biomarkers was validated by qPCR-based methods using tissue samples extracted from resected prostates and, separately, in one out of 13-core biopsies from prostates that were negative and positive for PCa.

Previously reported large-scale surveys identified thousands of genes or hypermethylated CpG regions that are differentially expressed between cancerous and benign prostate samples (reviewed in Bradford *et al*, 2006; Sørensen and Ørntoft, 2010; Jerónimo *et al*, 2011; Lucas and Heath, 2012; Willard and Koochekpour, 2012; Bhavsar *et al*, 2013; Chen *et al*, 2013; Day and Bianco-Miotto, 2013); however, because of high inter-individual variations, most of them are not recognised as putative markers for PCa diagnosis. This statement is consistent with our results presented in this study.

Only 12 mRNAs fulfilled the restricted criteria (≥ 3 FC in expression and an adjusted P -value < 0.05) in the pair-wise comparison between BPH and carcinoma-bearing prostates. Notably, the expression of PCA3, an approved diagnostic PCa biomarker, did not meet these criteria (FDR = 0.088, FC = 2.1). The validation step, performed on prostate specimens extracted from surgically resected glands, confirmed the high to moderate potential of all 13 genes (including PCA3) to distinguish benign samples from cancerous samples (Supplementary Figure S1); however, although five (LMO3, NELL2, PTGS1, RNF112 and CXCL13) of six negative markers also discriminated between BPH and adjacent non-cancerous prostate samples (with AUCs ranging between 0.871 and 0.943), the same was true for two (IL6, C15orf48; AUC = 0.908, 0.889) of seven positive biomarkers (Figure 1).

In the second step of validation, we measured the expression of the 13 selected genes in single-needle biopsies selected from sets of TRUS-guided biopsies that were negative or positive for PCa. Only three markers (HOXC6, AMACR and PCA3) were able to distinguish benign from malignant prostates with acceptable levels of sensitivity and specificity (AUCs were > 0.8) (Figure 1). Even the

combination of those three genes were far from perfect discrimination (AUC = 0.887, Figure 2). These three genes were previously reported by other authors as probably the most promising mRNA-based markers for distinguishing cancerous from benign prostate tissue (Hessels *et al*, 2004; McCabe *et al*, 2008; Neves *et al*, 2008; Vinarskaja *et al*, 2011; Al-Maghrebi *et al*, 2012). However, although the expression of HOXC6, AMACR and PCA3 was significantly different between benign prostates and adjacent non-cancerous tissues in the first validation analyses, transcript levels showed a significant overlap (Supplementary Figure S1). Although basic classification in medicine is binary (Lasko *et al*, 2005), gene expression measurements generate a numeric value on a continuous scale. In a consequence, choosing a threshold for converting mRNA levels of the studied genes into the binary results appeared to be a fundamental challenge when higher specificity resulted in lower sensitivity. Therefore, from a practical point of view, even these three biomarkers could hardly 'rescue' most false-negative diagnosis from pathological assessment.

DNA methylation, occurring predominantly at CpG islands, is the most studied epigenetic modification and is frequently found within gene promoter regions. Genome-wide DNA methylation studies compared non-malignant prostate tissue and tumour tissue (review in Day and Bianco-Miotto, 2013) and found that the promoter hypermethylation of several genes, including FAM84A (NSE1) and SPOCK2 (Yang *et al*, 2013), EFEMP1 (Kim *et al*, 2011b), HOXD3 and BMP7 (Kron *et al*, 2009), GSTP1 (Goering *et al*, 2012), AOX1 and SPON2 (Kim *et al*, 2012), might serve as optimal biomarkers for differentiating PCa from non-malignant glands. Furthermore, there is substantial overlap in candidate methylated genes between different studies of genome-wide DNA methylation in prostate cancer (Day and Bianco-Miotto, 2013).

Our study confirmed that the analysis of promoter region methylation may discriminate between PCa and BPH. The DNA methylation of six (APC, TACC2, RARB, DGKZ, C5orf4 and HES5) out of eight selected promoter regions progressively increased with the neoplastic cell content of the cancerous specimens (Supplementary Figures S3 and S4), achieving high sensitivity and specificity, with AUCs reaching 0.95–1.0 (Figure 3). An unexpectedly small overlap in methylation levels was detected between PCa-positive and PCa-negative needle biopsies. Five methylation-based markers (APC, TACC2, RARB, DGKZ and HES5) AUCs were between 0.854 and 0.899. The combination of RARB, HES5 and C5orf4 has the AUC equal 0.909 (Figure 4). AUCs > 0.90 are considered excellent (Altintas *et al*, 2013). APC and RARB promoter methylation is affected in the vast majority of PCa cases (Goering *et al*, 2012). Furthermore, the combination of the methylation status of APC and that of RARB, together with that of two other genes, GSTP1 and RASSF1a (Florl *et al*, 2004), could be used to stratify patients as low or high risk for PCa when DNA methylation was measured in urine collected after prostate massage (Rouprêt *et al*, 2007).

Methylation-based classifiers have proved to be superior to expression-based ones. Although the best performing expression marker AMACR can detect < 30% of cancer cases with 100% specificity (Figure 1, red dotted line) the methylation of any of three genes: RARB, HES5 and DGKZ can detect > 50% cancer cases with 100% specificity (Figure 3, red dotted line). The same was true for classifiers based on combination of the three genes expression/methylation. AMACR, HOXC6 and PCA3 expression combination detected < 40% cancer cases (Figure 2, yellow dashed line), whereas RARB, HES5 and C5orf4 methylation detected over 50% cancer samples with a 100% specificity (Figure 4, yellow dashed line).

An ideal biomarker should consistently and completely discriminate between diseased and normal tissues. Currently we are witnessing an era in which molecular diagnostics utilising nucleic acids are reliable enough for use in patient care and for

broad clinical application; however, despite the tremendous number of potential biomarkers found using high-throughput genomic methods, only a few have been vigorously validated and are used for clinical diagnosis. A successful example of such a transition is a blood-based screening test for colorectal cancer that detects hypermethylation of the SEPT9 gene (Warren *et al*, 2011). From the perspective of clinical utility, any biomarker that meets the rigorous threshold of specificity and sensitivity for a certain disease state also requires the development of pre-analytical sample processing and readout techniques that are applicable in a typical clinical laboratory setting. With this in mind, biomarkers based on DNA (or its methylation) for which the readout is usually dichotomous (presence/absence of mutation or DNA methylation in a certain region) are better suited than mRNA-based biomarkers for which the expression signal is regarded as continuous. Furthermore, aberrant DNA methylation usually occurs at an early stage in cancer (where it silences tumour suppressor genes), making DNA methylation biomarkers good targets for early cancer detection. In addition, methylated DNA is far more stable than mRNA, does not require special handling requirements, can be easily detected retrospectively in archived samples, and is generally amenable to reliable analysis of patient samples.

Bearing in mind that the diagnosis of prostate cancer based on histological examination of core biopsies may be subject to significant error, we can assume that, among the patients selected for the second validation step, the detection rate from the histological examination could not have achieved 100% accuracy. Parts of core biopsies representing carcinoma-bearing prostates could be extracted from non-cancerous prostate areas, while some patients classified as negative for PCa did in fact have disease. Such events are not uncommon because the 20-core biopsy approach can miss up to 10% of cancers (Fink *et al*, 2001), and PCa can mimic benign prostate glands (Gaudin and Reuter, 1997). Thus, the most important attribute of a diagnostic biomarker for PCa would be the ability to detect cancer even when neoplastic cells are missed in needle biopsies, or in minimal residual cancer after radical prostatectomy.

In summary, two large-scale surveys backed-up by qPCR-based measurements allowed us to establish, validate and compare the diagnostic potential of PCa biomarkers; however, further validation in clinical cohorts is required to confirm the low utility of mRNA-based biomarkers and the significantly greater utility of DNA methylation-based biomarkers. If they are confirmed, methylation-based biomarkers might support clinical decision-making for patients with suspected PCa following negative initial prostate biopsy.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Science and Higher Education grant N N403 177340.

REFERENCES

- Altintas DM, Allioli N, Decaussin M, de Bernard S, Ruffion A, Samarut J, Vlaeminck-Guillem V (2013) Differentially expressed androgen-regulated genes in androgen-sensitive tissues reveal potential biomarkers of early prostate cancer. *PLoS One* **8**: e66278.
- Anastasiadis A, Zapala L, Cordeiro E, Antoniewicz A, Dimitriadis G, De Reijke T (2013) Complications of prostate biopsy. *Expert Rev Anticancer Ther* **13**: 829–837.
- Babaian RJ, Toi A, Kamoi K, Troncoso P, Sweet J, Evans R, Johnston D, Chen M (2000) A comparative analysis of sextant and an extended 11-core multisite directed biopsy strategy. *J Urol* **163**: 152–157.
- Bhavsar T, McCue P, Birbe R (2013) Molecular diagnosis of prostate cancer: are we up to age? *Semin Oncol* **40**: 259–275.

- Bradford TJ, Tomlins SA, Wang X, Chinnaiyan AM (2006) Molecular markers of prostate cancer. *Urol Oncol* **24**: 538–551.
- Chen Y, Li J, Yu X, Li S, Zhang X, Mo Z, Hu Y (2013) APC gene hypermethylation and prostate cancer: a systematic review and meta-analysis. *Eur J Hum Genet* **21**: 929–935.
- Day TK, Bianco-Miotto T (2013) Common gene pathways and families altered by DNA methylation in breast and prostate cancers. *Endocr Relat Cancer* **20**: R215–R232.
- Didkowska J, Wojciechowska U, Zatonski W (2012) *Cancer in Poland in 2011*. Warsaw: National Cancer Registry. Available at: http://onkologia.org.pl/wp-content/uploads/COI_Nowotwory2013_web.pdf.
- Esikorapci SY, Guliyev F, Akdogan B, Dogan HS, Ergen A, Ozen H (2005) Individualization of the biopsy protocol according to the prostate gland volume for prostate cancer detection. *J Urol* **173**: 1536–1540.
- Fink KG, Hutarew G, Lumper W, Jungwirth A, Dietze O, Schmeller NT (2001) Prostate cancer detection with two sets of ten-core compared with two sets of sextant biopsies. *Urology* **58**: 735–739.
- Florl AR, Steinhoff C, Müller M, Seifert H-H, Hader C, Engers R, Ackermann R, Schulz WA (2004) Coordinate hypermethylation at specific genes in prostate carcinoma precedes LINE-1 hypomethylation. *Br J Cancer* **91**: 985–994.
- Gaudin PB, Reuter VE (1997) Benign mimics of prostatic adenocarcinoma on needle biopsy. *Anat Pathol* **2**: 111–134.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smyth G, Tierney L, Yang JYH, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: R80.
- Goering W, Kloth M, Schulz WA (2012) DNA methylation changes in prostate cancer. *Methods Mol Biol* **863**: 47–66.
- Hessels D, Verhaegh GW, Schalken JA, Witjes JA (2004) Applicability of biomarkers in the early diagnosis of prostate cancer. *Expert Rev Mol Diagn* **4**: 513–526.
- Hodge KK, McNeal JE, Terris MK, Stamey TA (1989) Random systematic versus directed ultrasound guided transrectal core biopsies of the prostate. *J Urol* **142**: 71–74 (discussion 74–75).
- Hsu M-J, Chang Y-CI, Hsueh H-M (2014) Biomarker selection for medical diagnosis using the partial area under the ROC curve. *BMC Res Notes* **7**: 25.
- Javed S, Langley SEM (2013) Importance of HOX genes in normal prostate gland formation, prostate cancer development and its early detection. *BJU Int* **113**: 535–540.
- Jerónimo C, Bastian PJ, Bjartell A, Carbone GM, Catto JWF, Clark SJ, Henrique R, Nelson WG, Shariat SF (2011) Epigenetics in prostate cancer: biologic and clinical relevance. *Eur Urol* **60**: 753–766.
- Jones JS, Patel A, Schoenfeld L, Rabets JC, Zippe CD, Magi-Galluzzi C (2006) Saturation technique does not improve cancer detection as an initial prostate biopsy strategy. *J Urol* **175**: 485–488.
- Kim JH, Dhanasekaran SM, Prensner JR, Cao X, Robinson D, Kalyana-Sundaram S, Huang C, Shankar S, Jing X, Iyer M, Hu M, Sam L, Grasso C, Maher CA, Palanisamy N, Mehra R, Kominsky HD, Siddiqui J, Yu J, Qin ZS, Chinnaiyan AM (2011a) Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. *Genome Res* **21**: 1028–1041.
- Kim JW, Kim S-T, Turner AR, Young T, Smith S, Liu W, Lindberg J, Egevad L, Gronberg H, Isaacs WB, Xu J (2012) Identification of new differentially methylated genes that have potential functional consequences in prostate cancer. *PLoS One* **7**: e48455.
- Kim Y-J, Yoon H-Y, Kim S-K, Kim Y-W, Kim E-J, Kim IY, Kim W-J (2011b) EFEMP1 as a novel DNA methylation marker for prostate cancer: array-based DNA methylation and expression profiling. *Clin Cancer Res* **17**: 4523–4530.
- Kron K, Pethe V, Briollais L, Sadikovic B, Ozelic H, Sunderji A, Venkateswaran V, Pinthus J, Fleshner N, van der Kwast T, Bapat B (2009) Discovery of novel hypermethylated genes in prostate cancer using genomic CpG island microarrays. *PLoS One* **4**: e4830.
- Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L (2005) The use of receiver operating characteristic curves in biomedical informatics. *J Biomed Inform* **38**: 404–415.
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402–408.
- Lucas SM, Heath EI (2012) Current challenges in development of differentially expressed and prognostic prostate cancer biomarkers. *Prostate Cancer* **2012**: 640968.
- Al-Maghrebi M, Kehinde EO, Anim JT, Sheikh M (2012) The role of combined measurement of tissue mRNA levels of AMACR and survivin in the diagnosis and risk stratification of patients with suspected prostate cancer. *Int Urol Nephrol* **44**: 1681–1689.
- McCabe CD, Spyropoulos DD, Martin D, Moreno CS (2008) Genome-wide analysis of the homeobox C6 transcriptional network in prostate cancer. *Cancer Res* **68**: 1988–1996.
- Mikula M, Rubel T, Karczmarski J, Goryca K, Dadlez M, Ostrowski J (2011) Integrating proteomic and transcriptomic high-throughput surveys for search of new biomarkers of colon tumors. *Funct Integr Genomics* **11**: 215–224.
- Neves AF, Araújo TG, WKFS Biase, Meola J, Alcântara TM, Freitas DG, Goulart LR (2008) Combined analysis of multiple mRNA markers by RT-PCR assay for prostate cancer diagnosis. *Clin Biochem* **41**: 1191–1198.
- Pepe P, Aragone F (2007) Saturation prostate needle biopsy and prostate cancer detection at initial and repeat evaluation. *Urology* **70**: 1131–1135.
- Rabets JC, Jones JS, Patel A, Zippe CD (2004) Prostate cancer detection with office based saturation biopsy in a repeat biopsy population. *J Urol* **172**: 94–97.
- Ravery V, Dominique S, Panhard X, Toublanc M, Boccon-Gibod L, Boccon-Gibod L (2008) The 20-core prostate biopsy protocol—a new gold standard? *J Urol* **179**: 504–507.
- Rittenhouse H, Blase A, Shamel B, Schalken J, Groskopf J (2013) The long and winding road to FDA approval of a novel prostate cancer test: our story. *Clin Chem* **59**: 32–34.
- Rouprêt M, Hupertan V, Yates DR, Catto JWF, Rehman I, Meuth M, Ricci S, Lacave R, Cancel-Tassin G, de la Taille A, Rozet F, Cathelineau X, Vallancien G, Hamdy FC, Cussenot O (2007) Molecular detection of localized prostate cancer using quantitative methylation-specific PCR on urinary cells obtained following prostate massage. *Clin Cancer Res* **13**: 1720–1725.
- Scattoni V, Raber M, Abdollah F, Roscigno M, Dehò F, Angiolilli D, Maccagnano C, Gallina A, Capitanio U, Freschi M, Doglioni C, Rigatti P, Montorsi F (2010) Biopsy schemes with the fewest cores for detecting 95% of the prostate cancers detected by a 24-core biopsy. *Eur Urol* **57**: 1–8.
- Scattoni V, Roscigno M, Raber M, Dehò F, Maga T, Zanoni M, Riva M, Sangalli M, Nava L, Mazzoccoli B, Freschi M, Guazzoni G, Rigatti P, Montorsi F (2008) Initial extended transrectal prostate biopsy—are more prostate cancers detected with 18 cores than with 12 cores? *J Urol* **179**: 1327–1331 (discussion 1331).
- Serefoglu EC, Altinova S, Ugras NS, Akincioglu E, Asil E, Balbay MD (2013) How reliable is 12-core prostate biopsy procedure in the detection of prostate cancer? *Can Urol Assoc J* **7**: E293–E298.
- Skrzypczak M, Goryca K, Rubel T, Paziewska A, Mikula M, Jarosz D, Pachlewski J, Oledzki J, Ostrowski J (2010) Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. *PLoS One* **5**: e13091.
- Sørensen KD, Ørntoft TF (2010) Discovery of prostate cancer biomarkers by microarray gene expression profiling. *Expert Rev Mol Diagn* **10**: 49–64.
- Svetec D, McCabe K, Peretsman S, Klein E, Levin H, Optenberg S, Thompson I (1998) Prostate rebiopsy is a poor surrogate of treatment efficacy in localized prostate cancer. *J Urol* **159**: 1606–1608.
- Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S (2013) A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**: 189–196.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paep A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**: RESEARCH0034.
- Vashi AR, Wojno KJ, Gillespie B, Oesterling JE (1998) A model for the number of cores per prostate biopsy based on patient age and prostate gland volume. *J Urol* **159**: 920–924.
- Vinarskaja A, Yamanaka M, Ingenwerth M, Schulz WA (2011) DNA Methylation and the HOXC6 Paradox in Prostate Cancer. *Cancers (Basel)* **3**: 3714–3725.
- Wang X, Spandidos A, Wang H, Seed B (2011) PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res* **40**: D1144–D1149.

Warren JD, Xiong W, Bunker AM, Vaughn CP, Furtado LV, Roberts WL, Fang JC, Samowitz WS, Heichman KA (2011) Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. *BMC Medicine* **9**: 133.

Willard SS, Koochekpour S (2012) Regulators of gene expression as biomarkers for prostate cancer. *Am J Cancer Res* **2**: 620–657.

Yang B, Bhusari S, Kueck J, Weeratunga P, Wagner J, Levenson G, Huang W, Jarrard DF (2013) Methylation profiling defines an extensive field defect in

histologically normal prostate tissues associated with prostate cancer. *Neoplasia* **15**: 399–408.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)