
Article:
Drayton, R.M., Rehman, I., Clarke, R. et al. (13 more authors) (2015) Identification and Diagnostic Performance of a Small RNA within the PCA3 and BMCC1 Gene Locus That Potentially Targets mRNA. Cancer Epidemiology, Biomarkers and Prevention, 24 (1). 268 - 275. ISSN 1055-9965

https://doi.org/10.1158/1055-9965.EPI-14-0377

Identification and diagnostic performance of a small RNA within the PCA3 and BMCC1 gene locus that potentially targets mRNA

Ross M Drayton [1], Ishtiaq Rehman [1], Raymond Clarke [2], Zhongming Zhao [3], Karl Pang[1], Saiful Miah[1], Robert Stoehr [4], Arndt Hartmann [4], Sheila Blizard [1], Martin Lavin [2], Helen E. Bryant [1], Elena S. Martens-Uzunova [5], Guido Jenster [5], Freddie C. Hamdy [6], Robert A. Gardiner [2] and James W.F. Catto [1]*

1. Academic Urology Unit and Academic Unit of Molecular Oncology, University of Sheffield, UK;
2. Department of Urology, University of Queensland, Australia;
3. Departments of Biomedical Informatics and Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN 37203, USA
4. Department of Pathology, University of Erlangen, Germany;
5. Department of Urology, Josephine Nefkens Institute, Erasmus MC, The Netherlands;
6. Nuffield Department of Surgery, University of Oxford, UK

* Correspondence to;
James Catto
Academic Urology Unit and Academic Unit of Molecular Oncology,
G Floor, The Medical School,
University of Sheffield,
Beech Hill Road,
SHEFFIELD,
S10 2RX, United Kingdom

Tel: +44 (0)114 226 1229
Fax: +44 (0)114 271 2268
Email: j.catto@sheffield.ac.uk

Running title: Small RNA within PCA3

Word Count: 3310
Abstract word count: 278
Tables and Figures: 5

Disclosure of potential conflicts of interest. The authors declare no competing conflicts of interest.

Abstract

Background

PCA3 is a long non-coding RNA (lncRNA) with unknown function, upregulated in prostate cancer. LncRNAs may be processed into smaller active species. We hypothesized this role for PCA3 and searched for small transcribed RNAs within this gene locus.

Methods

We computed energetically feasible RNA hairpins within the BMCC1 gene locus (encompassing PCA3). We searched a deep-sequencing prostate transcriptome for RNAs derived from these hairpins. We measured their expression using QrtPCR in three cohorts of prostate cancer tissues (n=60), exfoliated urinary cells (n=484 with cancer and n=166 controls) and in cell lines (n=22). We used *in silico* predictions and RNA knock-up to identify potential mRNA targets of short transcribed RNAs.

Results

We predicted 13 hairpins, of which PCA3-shRNA2 was most abundant within the prostate transcriptome. PCA3-shRNA2 is located within intron 1 of PCA3 and appears regulated by androgens. Expression of PCA3-shRNA2 was upregulated in malignant prostatic tissues, exfoliated urinary cells from men with prostate cancer (13-273 fold change, T Test $p<0.003$) and closely correlated to PCA3 expression ($r=0.84$ to 0.93, $p<0.001$). Urinary PCA3-shRNA2 (C-index 0.75-0.81) and PCA3 (C-index 0.78) could predict the presence of cancer in most men. PCA3-shRNA2 knock-up altered the expression of predicted target mRNAs,

including COPS2, SOX11, WDR48, TEAD1 and Noggin. PCA3-shRNA2 expression was negatively correlated with COPS2 in patient samples ($r=-0.32$, $p<0.001$).

Conclusion

We identified a short RNA transcribed from the PCA3 gene, whose expression is closely correlated to PCA3, which may target mRNAs implicated in prostate biology.

Impact

This short RNA is stable *ex vivo*, suggesting a role as a more robust target of the PCA3 assay. We identify cytoplasmic enrichment of this RNA and potential targeting of mRNAs implicated in prostate carcinogenesis.

Introduction

Prostate cancer (PCa) is the most common cancer in men (1). The diagnosis of PCa is typically based on a combination of digital rectal examination (DRE), serum PSA and prostate biopsy (2). This approach leads to over diagnosis of indolent cancer and can delay the detection of significant disease. Non-protein biomarkers have been identified to help this diagnostic approach. To date, Prostate Cancer Associated 3 (PCA3) appears the most promising and received FDA approval in 2012 (3). The diagnostic assay (PROGENSA™) uses quantitative rtPCR to measure the expression of the PCA3 and PSA mRNA in post-prostatic massage urine samples (4). The PCA3 score (PCA3/PSA mRNA ratio x 1000) is higher in men with PCa than in controls and may be used to guide the need for further prostate biopsy (5, 6).

PCA3 is a long non-coding RNA (lncRNA) that is over-expressed in most PCa specimens. The PCA3 gene is located on chromosome 9q21-22 in an antisense orientation within intron 6 of the BMCC1/PRUNE2 gene (7)(8). PCA3 expression is predominantly restricted to the prostate, under androgen regulation and the gene may produce multiple transcripts through splicing and alternate polyadenylation (8). The adoption of PCA3 assay into clinical practice has been hampered by several factors. Firstly, the assay detects a lncRNA (gene length 25kb, PCR target sequence is 380 bases) that is unstable *ex vivo*. Target transcripts require protection from RNAses prior to analysis, making the test expensive and vulnerable to transportation errors. Secondly, the function of PCA3 is unknown, producing a biological gap in knowledge. Finally, the test has a

poor sensitivity for PCa and is mainly used in conjunction with serum PSA in men with a previous negative biopsy (5).

The importance of lncRNAs in human health and disease is becoming clearer (9) (10). The GENCODE consortium recently annotated 9,277 lncRNA genes, corresponding to 14,880 transcripts (11). In contrast to protein coding genes, lncRNAs typically have few gene exons, can be processed into active short species (12, 13)(14) and are not conserved from primitive species (15). As no functional role for PCA3 has been assigned to date, we wondered whether this ncRNA might encode an active shorter species. To test this hypothesis, we searched for possible short ncRNAs derived from sequence within the boundaries of the BMCC1 primary transcript, which spans PCA3, and evaluated their translational role. Here we report the outcomes of this work.

Materials and Methods

Identification of hairpin RNA structures

To identify short potentially transcribed ncRNAs, we searched the BMCC1 gene locus for predicted and energetically feasible RNA hairpins using ProMir II (16) and MiPred (17). These programs compare random sequences within the target hairpin using structure-sequence composition and minimum of free energy of the secondary structure. We then searched a small RNA transcriptome generated from malignant prostatic epithelium using deep sequencing (18) for RNA sequences derived from these predicted hairpins. We measured the expression of any identified short RNA sequences using custom stem loop primers (TaqMan

small RNA assays, Applied Biosystems, UK) with realtime qPCR (as described (19)) in cell lines and human tissues samples.

Cell lines and androgen regulation of RNA

We examined a panel of cell lines representing prostate cancer (DU145, LNCap, LNCap-LN3, LNCaP-pro5, PC2, PC3M, PC3M-ln4) and other common human malignancies (A549 and NCI-H460 (lung), AN3CA (endometrial), EJ/T24, RT112 and RT4 (bladder), HCT-116 (colorectal), HEK293 (human embryonic kidney), HeLa (Vulval), Jurkat (T-Cell lymphoma), MCF-7 and T47D (breast), MRC5 (lung fibroblasts), SKOV-3 (ovarian) and WM793 (melanoma)). Cell lines were grown in appropriate media according to standard methods (20). To determine androgen regulation of candidate RNAs, we examined their expression in LNCap cells (chosen for their androgen dependency) growing in androgen-depleted media (phenol red free RPMI-1640 and 10% charcoal stripped serum (Sigma)) with no (0nM), 1nM and 10nM added testosterone (7).

Expression in Prostatic RNA

We measured the expression of candidate short RNAs in prostatic tissues and urinary samples enriched for prostatic cells and secretions. Frozen malignant and histologically benign prostatic tissues were obtained using laser-capture microdissection of fresh radical prostatectomy specimens from the University of Erlangen, Germany. Each sample was marked by a dedicated uropathologist (AH) and tissue extracted from 10 x 10uM sections (>80% tumor purity). RNA from exfoliated prostatic urinary cells and secretions was collected following vigorous prostatic massage in men with PCa and matching controls within separate pilot

and validation cohorts from the University of Sheffield, UK. Controls were matched for age and PSA, and selected if they had undergone 2 or more prostatic biopsies without finding cancer. Following massage, the first 10-20mls of urine was collected and centrifuged. The cell pellet was then washed twice in PBS before storage. All samples were frozen at -80^{o}C until use. Ethics committee approval was in place before commencement of this study.

RNA extraction and quantification

Total RNA was extracted using the mirVana™ extraction kit (Ambion, TX) and measured using a 2100 Bioanalyzer (Agilent, Cheshire, UK) (as described elsewhere (19)). RNA expression was determined using realtime quantified rtPCR with primers for PCA3, BMCC1, PSA ((4, 7)) and two custom stem loop hairpin primers for PCA3-shRNA2 (termed "a" and "b": target sequences [ACTGCACTCCAGCCTGGGCA] and [CACTGCACTCCAGCCTGGGCA] Ambion: (Assay IDs: SCSGJ090, CSHSNF8 respectively) using qrtPCR (21). Expression of PCA3, BMCC1 and PCA3-shRNA2 was normalized to PSA and fold change calculated using ΔCt values (21). For RNA localization studies, we extracted separate nuclear and cytoplasmic RNA fractions from cells using standardized methodology (methods detailed in (22)).

Cloning primary transcript using 3'RACE

To determine the sequence (and genomic origin) of the primary RNA transcript producing our short RNA, we performed 3' rapid amplification of cDNA ends (RACE) in PC3 cells using the GeneRacer kit according to manufacturers guidelines (Life technologies, UK). Briefly, total RNA was precipitated, cleaned,

polyadenylated, adaptor ligated and amplified with hot start PCR using primers to the shRNA and the adaptor. The target sequences were cloned into *E. coli* (Top10, Life Technologies, UK), followed by extraction, purification and Sanger sequencing. Sequences were aligned, (Sequencher 5.1, Gene Codes))) and genomic matches were identified using BLAST (NCBI).

mRNA Target analysis and RNA knock-up

We identified putative target mRNAs with complementary sequences to our candidate short RNA using TargetScan (Vsn. 4.2, www.targetscan.org). We determined prostate cancer-specific expression of these mRNAs using publically available gene expression data (Arrayexpress ID: E-GEOD-8218 (23)). We analyzed cellular functions and pathway enrichment for these mRNAs with DAVID Bioinformatics Resource of identified mRNAs (24). We focused upon those with carcinogenic or prostate specific biological functions.

For exploratory analysis of targeting, we examined expression of selected mRNA targets in LNCaP cells following PCA3-shRNA2 knock-up. Briefly, we transfected cells with a custom made hairpin precursor designed to generate PCA3-shRNA2, or with an equal amount of control (scrambled) RNA (both from Ambion, UK) using Lipofectamine RNAiMAX (Life Technologies, UK) (methods detailed in (21)). We determined success of transfection using qrtPCR (as described above). All assays were performed in triplicate. We measured the expression of potential targets using qrtPCR (primer sequences and reaction conditions given in supplementary table 3) in these cell lines and in the exfoliated prostatic urinary cells from cohort 2.

Statistical analysis

RNA expression was compared between cells and tissues using Student's T test or ANOVA, and correlated with other RNAs using Pearson's coefficient within SPSS Vsn. 14.0 (SPSS Inc, Illinois)). Graphs were plotted using PRISM 6.0 (GraphPad Software inc.). The ability of each RNA to detect prostate cancer was determined using concordance indices and plotted using ROC curves (25). All tests were two sided and a p value of <0.05 taken as the threshold of significance.

Results

Identification of expressed short RNAs sequences within PCA3 and BMCC1

In Silico analysis of the BMCC1 locus identified 13 potential RNA hairpins (supplementary table 1). Each was derived from sequence within an intron of BMCC1 and most located around the PCA3 locus. MiPred classified ten of these as likely to be real. A search of the prostate transcriptome identified 5 of these RNAs, including RNA2 (which we termed PCA3-shRNA2 (short RNA number 2), supplementary figure 1a) that accounted for 72/79 (91%) of hits. PCA3-shRNA2 is located within intron 1 of the PCA3 gene adjacent to a region of high species conservation (Figure 1a).

Alignment of the transcriptomic sequences to the genome revealed two potential 5' start sequences for PCA3-shRNA2; namely ACUG and a minority member starting with CACUG (figure 1b). We designed Taqman assays to each (given that the 5' end of short RNA is vital for mRNA-targeting) and named these assays

PCA3-shRNA2a and PCA3-shRNA2b, respectively. We measured their expression in the 22 cell lines. We detected expression of PCA3-shRNA2 in all 7 prostate and 15 other cancer cell lines (figure 1c, supplementary figure 2). PCA3 and PCA3-shRNA2 expression did not vary significantly with organ of origin for these cells. We normalized RNA expression to PSA mRNA, as we were keen to compare with the commercial PCA3 assay (which uses this reference gene). Expression of the target RNA was similar whether using normalized PCA3-shRNA2a or PCA3-shRNA2b primers (data not shown: $r=0.98$, $p<0.001$). We identified a correlation between the expression of PCA3 mRNA and PCA3-shRNA2 (PCA3-shRNA2a assay: $r=0.92$, $p<0.001$ and PCA3-shRNA2b assay $r=0.93$, $p<0.001$), which was closer than for BMCC1 ($r=0.67$ for PCA3-shRNA2a and $r=0.72$ for PCA3-shRNA2b, $p<0.001$).

These data support our transcriptomic analysis, but do not prove our short RNA is derived from sequence within the PCA3 intron. To analyze this, we used 3' RACE to clone the primary transcript from the PCA3-shRNA2a primer. Sequenced RACE products from selected colonies aligned to the PCA3 intronic locus and supported our *in silico* prediction of a 98bp hairpin (red box, figure 1d). A BLAST search of this 98bp sequence revealed strong (97%) homology for only one locus in the genome, i.e. that within PCA3 intron 1 (supplementary figure 1b).

It is known that many RNAs important in prostate carcinogenesis are regulated by the androgen receptor. In LNCap cells (chosen for their androgen dependency), both PCA3 (2.1 ± 0.31 fold change (mean±st. dev.)) and PCA3-

shRNA2 (2.75 ± 0.23 fold change (mean±st. dev.)) were upregulated in a dose dependent manner (supplementary figure 3) by testosterone. The changes were less than seen for PSA (219.0±25.2 fold upregulation (mean±st. dev.)).

Expression of PCA3-shRNA2 in prostate tissue

Having identified the existence and origin of this short RNA, we investigated its expression in malignant and benign prostatic tissues from 60 radical prostatectomy specimens (Table 1, Figure 1e). We identified that expression of PCA3 and PCA3-shRNA2 were correlated ($r=0.88$, $p<0.001$) and upregulation of each RNA in malignant tissues when compared with benign samples. The extent of this difference was largest for PSA (8.6±1.2 fold change (mean±st. dev.)) and least for PCA3-shRNA2 (1.4±1 data not shown). When normalized to PSA, we found significant differences in expression for PCA3-shRNA2, PCA3 and BMCC1 between malignant and benign tissues (T test $p<0.01$, and supplementary figure 4). Once again there was close correlation in the detection of PCA3-shRNA2 using PCA3-shRNA2a or PCA3-shRNA2b primers ($r=0.99$, $p<0.001$). There was less correlation between PCA3-shRNA2 and BMCC1 expression ($r=0.51$, $p<0.001$).

Analysis of PCA3-shRNA2 in urinary samples

The clinical utility for PCA3 is a test for prostate cancer using exfoliated prostatic urinary cells. To explore this role for PCA3-shRNA2, we examined expression in 179 post-DRE urinary samples (table 1) from men with (n=129) and without prostate cancer (n=50). Once again we identified close correlation between PCA3 and PCA3-shRNA2 expression (figure 2a, $r=0.84$, $p<0.001$), between the two

PCA3-shRNA2 assays (i.e. PCA3-shRNA2a Vs. PCA3-shRNA2b, $r=0.95$ $p<0.001$), and less close correlation with BMCC1 expression ($r=0.30$, $p<0.001$). Overall there was upregulation of PCA3 (86.2±53.1 fold change (mean±st. dev.)), BMCC1 (2.7±0.1) and PCA3-shRNA2 (273±0.1) in specimens from men with cancer, when compared to controls (all T Test $p<0.003$, figure 2b). This allowed the identification of malignancy in most men, (figure 2c: concordance indices suggest that PCA3 (C-index 0.78) and PCA3-shRNA2 (C-index 0.75) had similar accuracy for cancer, and were superior to BMCC1 (C-index 0.66)). Scatterplots comparing PCA3 with PCA3-shRNA2 expression (figures 1e and 2a) suggested less variation in malignant samples than benign samples. Thus we plotted RNA expression in the frozen tissues and urinary samples according to cancer presence (supplementary figure 5) and saw less variation for malignant ($r=0.90$, $p<0.001$) than for benign ($r=0.58$, $p<0.001$) samples.

To explore the robustness of these findings, we examined a separate larger validation cohort of 471 urinary samples (figure 3). Samples were collected and processed in a similar manner to the pilot cohort. Quantitative analysis revealed PCA3-shRNA2 expression was higher in samples from men with prostate cancer than controls (13.0±2.8 fold upregulation (mean±st. dev.) in malignant samples, T Test $p<0.001$, supplementary figure 6). Expression of PCA3-shRNA2 did not vary with tumor stage (figure 3b), as reported for PCA3 (4), but could correctly identify PCa (C-index 0.81, figure 3c), supporting our pilot exploration outcomes.

The functional role of PCA3-shRNA2

To date, little is known about the function of PCA3. To explore a functional role for PCA3-shRNA2, we investigated its cellular localization. QrtPCR of total and nuclear fractions revealed a cytoplasmic enrichment (Nuclear:Cytoplasmic ratio =0.6) for PCA3-shRNA2, close to that seen for established microRNAs (supplementary figure 7), and very different from PCA3 (with its mostly nuclear localization). As this suggests a potential mRNA targeting capacity, we searched the genome for complementary sequences. Using TargetScan we identified 178 mRNAs with complementary seed sequences (supplementary table 2). Gene enrichment analysis revealed significant associations (Bonferroni adjusted $p<0.05$) with pathways important for cell regulation (such as cell adhesion and growth, cell signaling) and prostate biology (such as response to steroids, TGF-β signaling and urogenital development). We annotated these mRNAs with their expression in human prostate cancer samples (23), and preferentially selected those known to be down-regulated in cancer (reflecting our hypothesized targeting by upregulated PCA3-shRNA2: defined as fold change <1.0 and T Test $p<0.05$) or implicated in prostate cancer biology, and having high predicted binding affinity (e.g. 8-mer seed). The resultant panel (table 2) included interesting potential targets; such as ETS variant genes 1 and 5 (ETV1 and ETV5), mitogen-activated protein kinase kinase kinase 1 (MAPK31), noggin, N-cadherin and TEA domain family member 1 (SV40 transcriptional enhancer factor).

We transfected DU145 cells (chosen as they have low endogenous PCA3-shRNA2 expression) with the PCA3-shRNA2 plasmid and a scrambled RNA sequence, and measured RNA expression of these 12 predicted targets (supplementary figure

8a). We identified reciprocal knock down of COPS2 (COP9 signalosome subunit 2), SOX11 (sex determining region Y - box 11), WDR48, TEAD1 and Noggin, suggestive of targeting (table 2). We measured the mRNA expression of the two strongest candidates (COPS2 and SOX11) in the largest urinary sample cohort to look for biological associations *in vivo*. We identified reduced expression of COPS2 in the prostate cancer samples (fold change 0.29±0.5, T test $p<0.001$) when compared with controls, and a significant inverse correlation between the expression of COPS2 and PCA3-shRNA2 ($r=-0.32$, $p<0.001$, supplementary figure 8b,c). Non-significant lower expression for SOX11 was also seen in malignant samples (fold change 0.74±1.5, $p=0.08$) when compared to controls and this mRNA was not significantly correlated with PCA3-shRNA2 expression ($r=-0.1$, $p=0.48$).

Discussion

It is known that many transcribed RNAs do not encode proteins. These are termed ncRNAs and are currently best classified according to size and cellular location. Whilst a fraction of short RNAs, known as microRNAs (around 20-22bps in size), have been extensively studied (14), little is known about the function of most long ncRNAs (reviewed in (9, 10)). Identified roles for longer ncRNAs include direct involvement in chromatin remodeling and androgen receptor regulation (26), and processing into shorter more-active ncRNAs. For example, many microRNAs are clustered together and derived from single primary transcripts (such as miRs-24-2/27a/23a) (27). Recently, Rogler et al. reported RNase MRP (a 268bp non-coding RNA component of mitochondrial

RNA processing endoribonuclease) was the source for two short (around 20bp) RNAs important in the biology of cartilage–hair hypoplasia (13). As such, we hypothesised that one role for PCA3 could be as a source for short biologically active species.

Here we present a combination of *in silico* and *in vitro* data suggesting a short RNA hairpin is produced during processing of the PCA3 primary transcript, and that this may have a biological activity. This short RNA is located within intron 1 of PCA3, close to a sequence of species conservation, suggesting biological protection. Our findings have direct and indirect clinical implications. Firstly, expression of the short ncRNA appeared closely correlated with that for PCA3. This was expected, given our data suggesting the short RNA is derived from the PCA3 transcript. In contrast to long mRNAs, short ncRNAs are stable molecules and do not decay with repeated freeze thawing or prolonged storage at room temperature without RNAse inhibition. For example, we recently reported that short RNAs do not dramatically degrade with prolonged storage at room temperature (in plain clean universal containers without RNAase inhibitors) and with freeze thawing (28). As such, PCA3-shRNA2 may be a more stable biomarker for prostate cancer than PCA3. Assays to detect PCA3-shRNA2 would not be so vulnerable to delays in handling or variations in stringency in collection, and so should be more reproducible. In post-DRE urinary cell pellets from two large patient cohorts, we found that PCA3-shRNA2 detected cancer with a similar accuracy to PCA3. Whilst the PCA3 test is currently normalized to PSA mRNA expression, it is likely that short RNAs (such as prostate specific microRNAs) could replace the need for this mRNA.

Secondly, our data suggest a potential functional role for PCA3 and derivative short RNAs. An unbiased genome-wide computational search identified genes and pathways implicated in the biology of prostate cancer. Whilst many annotated genes in these pathway enrichment datasets are implicated in cell homeostasis and regulation, relatively few are annotated for steroidal regulated pathways, TGF-β signaling and urogenital development. As such, the identification of genes involved in these pathways is extremely pertinent and adds support for a role of PCA3-shRNA2 in prostate biology. Our preliminary targeting analysis identified changes in COPS2, SOX11, WDR48, TEAD1 and Noggin with PCA3-shRNA2 upregulation. These mRNAs play roles in the regulation of gene transcription, urogenital tract development and in cell growth and signaling. As such, they appear ideal carcinogenic gene candidates. In a further analysis we explored the expression of COPS2 and SOX11 in exfoliated urinary cell pellets. We found that COPS2 expression was correlated to PCA3-shRNA2, suggesting biological validation, and that a trend for SOX11 was also seen. COPS2 is a transcription co-repressor that underwent a four-fold loss of expression in cells with PCA3-shRNA2 knock-up. COPS2 is a component of the COP9 signalosome complex that acts to regulate the ubiquitin conjugation pathway during various cellular and developmental processes, including phosphorylation of p53 and c-jun. COPS2 is abundantly expressed in most human tissues, suggesting an important role in cellular homeostasis, and has not been studied in depth with respect to human malignancies. SOX11 is a transcription factor belonging to the SRY-related HMG-box (SOX) family. These regulate multiple biological processes, such as hematopoiesis, vasculogenesis

and cardiogenesis during embryonic development (29), and some members are negative regulators of the WNT-beta-catenin-TCF pathway (30) which is implicated in prostate biology. To date, whilst Katoh reported reduced expression of SOX7 in PCA cells, SOX11 function and expression has not been reported in PCa. Our data now suggest the need for further analysis of our candidate target mRNAs. Of the other predicted targets, noggin appears particularly interesting. Noggin is an antagonist of bone morphogenetic proteins (BMP) (31), which has been reported to be down regulated in prostate cancer cells (32) (33). Noggin loss leads to the development of osteoblastic bone metastatses. Reversal of this loss may be used to palliate or diminish the activity of osteolytic malignant disease.

In conclusion, we have found evidence of a small active RNA that is derived from the PCA3 gene locus and probably co-expressed with PCA3 ncRNA. This may be a more suitable target of the PCA3 biomarker assay and could start to identify roles for this ncRNA in prostate biology.

Acknowledgements.

J.W.F. Catto was supported by a GSK Clinician Scientist fellowship and project grants from Yorkshire Cancer Research (Grant number S305PA), Astellas Educational Foundation and the European Union (European Community's Seventh Framework Programme. Grant Numbers: FP7/2007-2013, HEALTH-F2-2007-201438). H.E. Bryant was supported by an RCUK research fellowship

References

1. Center MM, Jemal A, Lortet-Tieulent J, Ward E, Ferlay J, Brawley O, et al. International variation in prostate cancer incidence and mortality rates. Eur Urol. 2012;61:1079-92.

2. Heidenreich A, Bellmunt J, Bolla M, Joniau S, Mason M, Matveev V, et al. EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and treatment of clinically localised disease. Eur Urol. 2011;59:61-71.

3. de Kok JB, Verhaegh GW, Roelofs RW, Hessels D, Kiemeney LA, Aalders TW, et al. DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. Cancer Res. 2002;62:2695-8.

4. Hessels D, Klein Gunnewiek JM, van Oort I, Karthaus HF, van Leenders GJ, van Balken B, et al. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. Eur Urol. 2003;44:8-15; discussion -6.

5. Hessels D, Schalken JA. The use of PCA3 in the diagnosis of prostate cancer. Nat Rev Urol. 2009;6:255-61.

6. van Gils MP, Hessels D, van Hooij O, Jannink SA, Peelen WP, Hanssen SL, et al. The time-resolved fluorescence-based PCA3 test on urinary sediments after digital rectal examination; a Dutch multicenter validation of the diagnostic performance. Clin Cancer Res. 2007;13:939-43.

7. Clarke RA, Zhao Z, Guo AY, Roper K, Teng L, Fang ZM, et al. New genomic structure for prostate cancer specific gene PCA3 within BMCC1: implications for prostate cancer detection and progression. PLoS One. 2009;4:e4995.

8. Salagierski M, Verhaegh GW, Jannink SA, Smit FP, Hessels D, Schalken JA. Differential expression of PCA3 and its overlapping PRUNE2 transcript in prostate cancer. Prostate. 2010;70:70-8.

9. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009;136:629-41.

10. Martens-Uzunova ES, Bottcher R, Croce CM, Jenster G, Visakorpi T, Calin GA. Long Noncoding RNA in Prostate, Bladder, and Kidney Cancer. Eur Urol. 2014 Jun;65(6):1140-51

11. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22:1775-89.

12. Rother S, Meister G. Small RNAs derived from longer non-coding RNAs. Biochimie. 2011;93:1905-15.

13. Rogler LE, Kosmyna B, Moskowotz D, Bebawee R, Rahimzadeh J, Kutchko K, et al. Small RNAs derived from lncRNA RNase MRP have gene-silencing

activity relevant to human cartilage–hair hypoplasia. Human Molecular Genetics. 2014;23:368-82.

14. Catto JW, Alcaraz A, Bjartell AS, De Vere White R, Evans CP, Fussel S, et al. MicroRNA in Prostate, Bladder, and Kidney Cancer: A Systematic Review. Eur Urol. 2011;59:671-81.

15. Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, et al. Rapid turnover of long noncoding RNAs and the evolution of gene expression. PLoS Genet. 2012;8:e1002841.

16. Nam JW, Kim J, Kim SK, Zhang BT. ProMiR II: a web server for the probabilistic prediction of clustered, nonclustered, conserved and nonconserved microRNAs. Nucleic Acids Res. 2006;34:W455-8.

17. Jiang P, Wu H, Wang W, Ma W, Sun X, Lu Z. MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. Nucleic Acids Res. 2007;35:W339-44.

18. Martens-Uzunova ES, Jalava SE, Dits NF, van Leenders GJ, Moller S, Trapman J, et al. Diagnostic and prognostic signatures from the small non-coding RNA transcriptome in prostate cancer. Oncogene. 2012;31:978-91.

19. Dudziec E, Miah S, Choudhury HMZ, Owen HC, Blizard C, Glover M, et al. Hypermethylation of CpG Islands and Shores around specific MicroRNAs and Mirtrons is associated with the phenotype and presence of Bladder Cancer. Clin Cancer Res. 2011;17:1287-96.

20. Leiblich A, Cross SS, Catto JW, Phillips JT, Leung HY, Hamdy FC, et al. Lactate dehydrogenase-B is silenced by promoter hypermethylation in human prostate cancer. Oncogene. 2006;25:2953-60.

21. Catto JW, Miah S, Owen HC, Bryant H, Dudziec E, Larre S, et al. Distinct microRNA alterations characterize high and low grade bladder cancer. Cancer Res. 2009;69:8472-81.

22. Herold A, Teixeira L, Izaurralde E. Genome-wide analysis of nuclear mRNA export pathways in Drosophila. Embo J. 2003;22:2472-83.

23. Stuart RO, Wachsman W, Berry CC, Wang-Rodriguez J, Wasserman L, Klacansky I, et al. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. Proc Natl Acad Sci U S A. 2004;101:615-20.

24. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4:44-57.

25. Catto JW, Abbod MF, Linkens DA, Larre S, Rosario DJ, Hamdy FC. Neurofuzzy modeling to determine recurrence risk following radical cystectomy for non-metastatic urothelial carcinoma of the bladder Clin Cancer Res. 2009;15:3150-5.

26. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature. 2013;500:598-602.

27. Drayton RM, Dudziec E, Peters S, Bertz S, Hartmann A, Bryant H, et al. Reduced expression of microRNA-27a modulates cisplatin resistance in bladder cancer by targeting the cystine/glutamate exchanger SLC7A11. Clin Cancer Res. 2014 Apr 1;20(7):1990-2000.

28. Miah S, Dudziec E, Drayton RM, Zlotta AR, Morgan SL, Rosario DJ, et al. An evaluation of urinary microRNA reveals a high sensitivity for bladder cancer. British journal of cancer. 2012;107:123-8.

29. Stovall DB, Cao P, Sui G. SOX7: From a developmental regulator to an emerging tumor suppressor. Histol Histopathol. 2014 Apr;29(4):439-45

30. Katoh M. Expression of human SOX7 in normal tissues and tumors. International journal of molecular medicine. 2002;9:363-8.

31. Haudenschild DR, Palmer SM, Moseley TA, You Z, Reddi AH. Bone morphogenetic protein (BMP)-6 signaling and BMP antagonist noggin in prostate cancer. Cancer Res. 2004;64:8276-84.

32. Secondini C, Wetterwald A, Schwaninger R, Thalmann GN, Cecchini MG. The role of the BMP signaling antagonist noggin in the development of prostate cancer osteolytic bone metastasis. PLoS One. 2011;6:e16078.

33. Schwaninger R, Rentsch CA, Wetterwald A, van der Horst G, van Bezooijen RL, van der Pluijm G, et al. Lack of noggin expression by cancer cells is a determinant of the osteoblast response in bone metastases. Am J Pathol. 2007;170:160-75.

Tables

		1. Radical prostatectomy specimens	2. Urinary Pilot Cohort	3. Urinary Validation Cohort
Material		Tissue	Disaggregated urinary cells	
Total		60	179	471
Benign tissue/controls				
Total		29	50	116
Age				
	Median	67.0	68.3	66.2
	St. dev.	5.5	8.9	7.2
PSA				
	Median	7.1	7.8	6.2
	St. dev.	4.9	5.9	6.2
Prostate cancer				
Total		31	129	355
Stage				
	pT1-2	7	34	279
	pT3	24	30	61
	Metastatic	0	49	15
	Missing	0	7	0
Gleason sum				
	5	0	3	0
	6	11	54	191
	7	6	29	125
	8-10	14	21	31
	Missing	0	14	8
Age				
	Median	67		65.1
	St dev	5.5		7.0
PSA				
	Median	7.1	7.5	7.4
	St Dev	4.9	175.1	418.9
	0-10	17	79	248
	11-20	12	27	64
	>20	2	16	37
	Missing	0	0	9

Table 1. Patient samples analyzed in this report

Gene Id	Gene Name	8Mer	7Mer-M8	7Mer-1A	Microarray Expression: Fold Change ±1.0 & p<0.05	Fold change (DU145 & PCA3-shRNA2 (mean ± st. dev.))	Cell Adhesion	Cell Growth	Cell Surface Signalling	Growth Regulation	Negative Regulation of Proliferation	Protein Signalling	Regulation of Transcription	TGFB Signalling	Transmembrane Signalling	Urogenital Development/Sex Development
COPS2	COP9 constitutive photomorphogenic homolog subunit 2	1	1	0	0	0.24 ±0.15							1			
SOX11	SRY (sex determining region Y)-box 11	1	0	0	0	0.36 ±0.31							1			1
WDR48	WD repeat domain 48	1	0	0	1	0.51 ±0.2										
TEAD1	TEA domain family member 1	1	0	1	1	0.57 ±0.32							1			
NOG	Noggin	1	0	0	0	0.72 ±0.27			1	1	1	1			1	2
WDR1	WD repeat domain 1	1	0	0	1	0.86 ±0.32										
INVS	Inversin	1	0	0	0	1.06 ±0.53			1							1
CDH2	N-cadherin	1	0	0	1	1.37 ±1.87	1									
MAP3K1	Mitogen-activated protein kinase kinase kinase 1	1	0	0	0	1.54 ±0.93		1	1			1	1	1	1	
ETV5	Ets variant gene 5	1	0	0	1	1.55 ±0.39							1			
KIAA0515	KIAA0515	2	0	0	0	1.59 ±1.35										
ETV1	Ets variant gene 1	1	0	0	0	3.35 ±3.2							1			

Table 2. Selected Potential targets of PCA3-shRNA2

End (bp)	BMCC1 Region	Note	Free energy	GC ratio	Entropy	Promir value	MiPred Result	MiPred Confidence	Freq in Transcriptomic data	% of hits
31,365	Intron 1		-25.5	0.38	1.9508	0.0489	Pseudo	68.70%	**1**	1.3%
93,278	Intron 6 (downstream PCA3)	Repeat, first 75 bp is AluSx (SINE/Alu)	-30.8	0.45	1.97326	0.0451	Real	54.00%	**72**	91.1%
100,952	Intron 6 (downstream PCA3)	Repeat, MER5A (DNA/MER1_type)	-36.44	0.45	1.97809	0.0617	Real	66.90%	**0**	0.0%
105,132	Intron 6 (downstream PCA3)	Conserved in 43 mammals	-27	0.33	1.8797	0.0635	Real	74.20%	**0**	0.0%
115,526	Intron 6 (downstream PCA3)		-49.7	0.3	1.85017	5.0776	Real	74.20%	**0**	0.0%
163,666	Intron 6 (upstream PCA3)		-46.3	0.46	1.98194	1.5115	Real	68.60%	**0**	0.0%
176,673	Intron 6 (upstream PCA3)	Repeat, MER5A (DNA/MER1_type)	-37	0.46	1.99233	0.0559	Real	62.90%	**0**	0.0%
181,993	Intron 6 (upstream PCA3)	Conserved in human dog (see the alignment in the end of the file)	-28.1	0.33	1.89539	0.0742	Real	61.10%	**3**	3.8%
215,235	Intron 9		-30.9	0.51	1.99248	0.1165	Pseudo	50.80%	**1**	1.3%
236,277	Intron 9	Repeat, L1HS (LINE/L1)	-34.8	0.47	1.91696	0.0531	Real	62.30%	**2**	2.5%
247,401	Intron 9	Repeat, Tigger4a (DNA/MER2_type)	-39.2	0.38	1.95071	85.884	Real	75.80%	**0**	0.0%
254,347	Intron 10	Repeat, Tigger4a (DNA/MER2_type)	-29.03	0.36	1.94008	0.2239	Real	80.00%	**0**	0.0%
268,339	Intron 12		-27.85	0.39	1.9288	0.6003	Pseudo	51.20%	**0**	0.0%

Supplementary table 1. Predicted hairpin RNAs within PCA3 and BMCC1

Supplementary table 2. Predicted mRNA targets for PCA3-shRNA2

Gene ID	Gene name
KIAA0515	KIAA0515
COPS2	COP9 constitutive photomorphogenic homolog subunit 2 (Arabidopsis)
AFF3	AF4/FMR2 family, member 3
SFRS2	splicing factor, arginine/serine-rich 2
TEAD1	TEA domain family member 1 (SV40 transcriptional enhancer factor)
ANKHD1-EIF4EBP3	ANKHD1-EIF4EBP3
ANKRD57	ankyrin repeat domain 57
BTBD3	BTB (POZ) domain containing 3
C13orf36	chromosome 13 open reading frame 36
C8orf33	chromosome 8 open reading frame 33
CDH2	cadherin 2, type 1, N-cadherin (neuronal)
CLDN22	claudin 22
DCAKD	dephospho-CoA kinase domain containing
EIF4EBP3	eukaryotic translation initiation factor 4E binding protein 3
ETV1	ets variant gene 1
ETV5	ets variant gene 5 (ets-related molecule)
FAM123A	family with sequence similarity 123A
FAM123B	family with sequence similarity 123B
FAM40B	family with sequence similarity 40, member B
FLJ20309	hypothetical protein FLJ20309
HDHD2	haloacid dehalogenase-like hydrolase domain containing 2
INVS	inversin
IRF2BP2	interferon regulatory factor 2 binding protein 2
LIN9	lin-9 homolog (C. elegans)
LMNB1	lamin B1
LSM11	LSM11, U7 small nuclear RNA associated
MAP3K1	mitogen-activated protein kinase kinase kinase 1
METTL8	methyltransferase like 8
MLLT6	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 6
NOG	noggin
ODZ4	odz, odd Oz/ten-m homolog 4 (Drosophila)
ORMDL1	ORM1-like 1 (S. cerevisiae)
P15RS	cyclin-dependent kinase 2B-inhibitor-related protein
PDS5A	PDS5, regulator of cohesion maintenance, homolog A (S. cerevisiae)
PHF21A	PHD finger protein 21A
PHKA2	phosphorylase kinase, alpha 2 (liver)
PIGA	phosphatidylinositol glycan anchor biosynthesis, class A (paroxysmal nocturnal hemoglobinuria)
PNMA1	paraneoplastic antigen MA1

PSCDBP	pleckstrin homology, Sec7 and coiled-coil domains, binding protein
PXMP4	peroxisomal membrane protein 4, 24kDa
RAPGEF2	Rap guanine nucleotide exchange factor (GEF) 2
RNF169	ring finger protein 169
SESN2	sestrin 2
SFRS12IP1	SFRS12-interacting protein 1
SFRS2B	splicing factor, arginine/serine-rich 2B
SFRS3	splicing factor, arginine/serine-rich 3
SOX11	SRY (sex determining region Y)-box 11
SSR1	signal sequence receptor, alpha (translocon-associated protein alpha)
TLR4	toll-like receptor 4
TYW3	tRNA-yW synthesizing protein 3 homolog (S. cerevisiae)
WDR1	WD repeat domain 1
WDR48	WD repeat domain 48
XPO7	exportin 7
ZC3H10	zinc finger CCCH-type containing 10
ZFAND6	zinc finger, AN1-type domain 6
HTR2C	5-hydroxytryptamine (serotonin) receptor 2C
LIN54	lin-54 homolog (C. elegans)
SPTBN1	spectrin, beta, non-erythrocytic 1
ABI1	abl-interactor 1
ADAM12	ADAM metallopeptidase domain 12 (meltrin alpha)
ANKH	ankylosis, progressive homolog (mouse)
APH1A	anterior pharynx defective 1 homolog A (C. elegans)
ARL5B	ADP-ribosylation factor-like 5B
BRPF1	bromodomain and PHD finger containing, 1
CCNL2	cyclin L2
CPEB2	cytoplasmic polyadenylation element binding protein 2
CROP	cisplatin resistance-associated overexpressed protein
CTBP2	C-terminal binding protein 2
DDX5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5
EHMT1	euchromatic histone-lysine N-methyltransferase 1
EIF3J	eukaryotic translation initiation factor 3, subunit J
EIF4G3	eukaryotic translation initiation factor 4 gamma, 3
EML1	echinoderm microtubule associated protein like 1
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog
GATAD2A	GATA zinc finger domain containing 2A
H3F3B	H3 histone, family 3B (H3.3B)
HMGB2	high-mobility group box 2
HNRNPA3	heterogeneous nuclear ribonucleoprotein A3
IKZF2	IKAROS family zinc finger 2 (Helios)
ITCH	itchy E3 ubiquitin protein ligase homolog (mouse)
KCMF1	potassium channel modulatory factor 1
KCNA4	potassium voltage-gated channel, shaker-related subfamily, member 4

LHFPL4	lipoma HMGIC fusion partner-like 4
LMO7	LIM domain 7
LOC399947	similar to expressed sequence AI593442
MAGI2	membrane associated guanylate kinase, WW and PDZ domain containing 2
MAP4K4	mitogen-activated protein kinase kinase kinase kinase 4
MBOAT1	membrane bound O-acyltransferase domain containing 1
MEIS1	Meis homeobox 1
MN1	meningioma (disrupted in balanced translocation) 1
MON2	MON2 homolog (S. cerevisiae)
NDN	necdin homolog (mouse)
NLK	nemo-like kinase
NMNAT2	nicotinamide nucleotide adenylyltransferase 2
OPCML	opioid binding protein/cell adhesion molecule-like
P2RY4	pyrimidinergic receptor P2Y, G-protein coupled, 4
PAX3	paired box 3
PBRM1	polybromo 1
PKD2	polycystic kidney disease 2 (autosomal dominant)
PPP1R8	protein phosphatase 1, regulatory (inhibitor) subunit 8
PRPF40A	PRP40 pre-mRNA processing factor 40 homolog A (S. cerevisiae)
PRPF40B	PRP40 pre-mRNA processing factor 40 homolog B (S. cerevisiae)
PURB	purine-rich element binding protein B
RAP2C	RAP2C, member of RAS oncogene family
RAPH1	Ras association (RalGDS/AF-6) and pleckstrin homology domains 1
RELT	RELT tumor necrosis factor receptor
SENP6	SUMO1/sentrin specific peptidase 6
SLC25A1	solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1
SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1
SLC38A4	solute carrier family 38, member 4
SNX22	sorting nexin 22
SOCS5	suppressor of cytokine signaling 5
ST3GAL5	ST3 beta-galactoside alpha-2,3-sialyltransferase 5
THSD7A	thrombospondin, type I, domain containing 7A
TNNI1	troponin I type 1 (skeletal, slow)
UNK	unkempt homolog (Drosophila)
USP7	ubiquitin specific peptidase 7 (herpes virus-associated)
VEGFA	vascular endothelial growth factor A
VGLL4	vestigial like 4 (Drosophila)
WSB2	WD repeat and SOCS box-containing 2
ZFAND3	zinc finger, AN1-type domain 3
ZFHX4	zinc finger homeobox 4
ZIC1	Zic family member 1 (odd-paired homolog, Drosophila)
ZNF516	zinc finger protein 516
ZNF740	zinc finger protein 740
ZNF827	zinc finger protein 827

tcag7.1228	hypothetical protein FLJ25778
ABHD13	abhydrolase domain containing 13
ANKS6	ankyrin repeat and sterile alpha motif domain containing 6
ARHGEF5	Rho guanine nucleotide exchange factor (GEF) 5
ASB8	ankyrin repeat and SOCS box-containing 8
BLID	BH3-like motif containing, cell death inducer
BRMS1L	breast cancer metastasis-suppressor 1-like
C1orf83	chromosome 1 open reading frame 83
C22orf15	chromosome 22 open reading frame 15
CD84	CD84 molecule
CEP350	centrosomal protein 350kDa
CSDE1	cold shock domain containing E1, RNA-binding
DCP1A	DCP1 decapping enzyme homolog A (S. cerevisiae)
DIP2C	DIP2 disco-interacting protein 2 homolog C (Drosophila)
DLX2	distal-less homeobox 2
FAM104A	family with sequence similarity 104, member A
GMEB2	glucocorticoid modulatory element binding protein 2
HAO1	hydroxyacid oxidase (glycolate oxidase) 1
KIAA1147	KIAA1147
KLHL20	kelch-like 20 (Drosophila)
LIMCH1	LIM and calponin homology domains 1
LPCAT3	lysophosphatidylcholine acyltransferase 3
MAF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)
MAPK1	mitogen-activated protein kinase 1
MLLT4	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 4
MTF1	metal-regulatory transcription factor 1
NAV1	neuron navigator 1
PISD	phosphatidylserine decarboxylase
RALBP1	ralA binding protein 1
RAVER1	ribonucleoprotein, PTB-binding 1
RPE	ribulose-5-phosphate-3-epimerase
SDC4	syndecan 4
SDK1	sidekick homolog 1, cell adhesion molecule (chicken)
SMAD4	SMAD family member 4
SOCS1	suppressor of cytokine signaling 1
SOCS6	suppressor of cytokine signaling 6
SRGAP3	SLIT-ROBO Rho GTPase activating protein 3
SSH2	slingshot homolog 2 (Drosophila)
TARDBP	TAR DNA binding protein
TIAM2	T-cell lymphoma invasion and metastasis 2
ZAK	sterile alpha motif and leucine zipper containing kinase AZK
ZNF263	zinc finger protein 263
ZNF618	zinc finger protein 618

Supplementary table 3. Primers and condition used to detect target mRNAs.

ID	Name	Fwd Primer	Rev Primer	Annealing Temp	Amplicon Size
KIAA0515	KIAA0515	TGGCTCACCTTCGTCATCTGA	TCATCCTCGGATACTGTTGGAA	60 °C	215
COPS2	COP9 constitutive photomorphogenic homolog subunit 2 (Arabidopsis)	TTTTACGCCAGTTACATCAGTCG	CTTCCCTCAAGTGCATTTTACCA	60 °C	234
TEAD1	TEA domain family member 1 (SV40 transcriptional enhancer factor)	GGCCGGGAATGATTCAAACAG	CAATGGAGCGACCTTGCCA	60 °C	165
CDH2	cadherin 2, type 1, N-cadherin (neuronal)	TCAGGCGTCTGTAGAGGCTT	ATGCACATCCTTCGATAAGACTG	60 °C	94
ETV1	ets variant gene 1	TGGCAGTTTTTGGTAGCTCTTC	CGGAGTGAACGGCTAAGTTTATC	60 °C	170
ETV5	ets variant gene 5 (ets-related molecule)	CAGTCAACTTCAAGAGGCTTGG	TGCTCATGGCTACAAGACGAC	60 °C	168
INVS	inversin	TGCTCTACAGAGGCTCATCGT	ACGCAATACATAAGTGGTGTTCT	60 °C	84
MAP3K1	mitogen-activated protein kinase kinase kinase 1	TCTCACCATATAGCCCTGAGGA	AGGAAAGAGTTAGGCCCTATCTG	60 °C	97
NOG	noggin	CCATGCCGAGCGAGATCAAA	TCGGAAATGATGGGGTACTGG	60 °C	337
SOX11	SRY (sex determining region Y)-box 11	AGGATTTGGATTCGTTCAGCG	AGGTCGGAGAAGTTCGCCT	60 °C	121
WDR1	WD repeat domain 1	TGGGATTTACGCAATTAGTTGGA	CCAGATAGTTGATGTACCCGGAC	60 °C	209
WDR48	WD repeat domain 48	TGGGACAATTCGCCTTTGGTC	TGTCAGGGTTTCTTAGGTCTGT	60 °C	164

Figure legends

Figure 1. Identification of PCA3 shRNA2. (a). Our potential shRNA is located in exon 6 of the BMCC1 gene. The Location of PCA3-shRNA2 is adjacent to a region of high conservation within intron 1 of PCA3. (b). Prostate cancer RNA transcriptomic data identified a relative abundance of this RNA. (c). Expression of PCA3-shRNA2 is closely correlated with PCA3 and less so to BMCC1 (DCT values normalized to PSA expression shown) in cultured cell lines. (d). 3' RACE identifies the longer hairpin structure in LNCap cells. (e). Expression of PCA3-shRNA2 is closely correlated with PCA3 in frozen benign and malignant prostatic tissues.

Figure 2. Expression of PCA3, BMCC1 and PCA3-shRNA2 in the urinary RNA from men with and without prostate cancer. (a). Expression of PCA3-shRNA was closely correlated with PCA3 expression (r=0.84). (b). Expression was higher in urinary pellets from men with prostate cancer than in benign controls for each RNA. (c). Expression of each RNA could identify the presence of the disease in most men. In comparison, PC3 and PCA3-shRNA2 expression were more reliable (c-indices 0.78, 0.75, respectively) than BMCC1 (c-index 0.66).

Figure 3. Detection of prostate cancer using PCA3-shRNA2. (a). Expression of PCA3-shRNA was correlated with PCA3 expression. (b). Expression was higher in urinary pellets from men with prostate cancer (Pca) than in benign controls, but did not vary with cancer stage (ANOVA p=0.46 between stages) (c). Expression could identify the presence of the disease in most men (C-index 0.81).

Supplementary Figure 1. Identification of the PCA3-shRNA2 hairpin. (a). Predicted hairpin of PCA3 RNA2. The bases in red are those identified within the prostate RNA transcriptome. (b). BLAST results of the 98bp fragment derived from PCA3-shRNA2 using 3'RACE, indicates that the sequence is found within PCA3 intron 1.

Supplementary figure 2. Expression of PCA3 and PCA3-shRNA2 in cell lines representing prostate cancer (blue and labeled) and other malignancies (red). The expression of the two PCA3 RNAs is shown (normalised to PSA mRNA expression) for each of 22 cell lines. The non-prostate cancer cell lines are not labeled for clarity. In order of PCA3-shRNA2 expression these are (from HCT-116 (PCA3-shRNA2 DCt=-15.85), HEK 293, A549, NCI-H460, WM793, RT112, T47D, MRC5, AN3CA, RT4, SKOV-3, EJ, MCF-7, Jurkat and HeLa (PCA3-shRNA2 DCt=5.08)).

Supplementary figure 3. Androgen regulated expression of PSA, PCA3 and PCA3 shRNA2. Expression (fold change) was determined using qrtPCR and normalized to a non-androgen regulated U1 RNA. PCA3 and PCA3 shRNA expression was directly related to testosterone concentration within growth media (0, 1nM and 10nM)

Supplementary figure 4. Expression of PCA3, BMCC1, PCA3-shRNA2a and PCA3-shRNA2b in malignant and benign prostatic tissue. Significantly higher expression is seen in malignant tissues, when compared to benign tissues, for BMCC1 (DCt: -11.5±2 vs. -14.1±3, T Test $p<0.01$ for malignant vs. benign) and PCA3-shRNA2 (e.g. PCA3-shRNA2a DCt: 2.15±5 vs. -1.17±4).

Supplementary figure 5. Correlation of PCA3 and PCA3-shRNA2 in urinary and frozen tissue according to pathology. Expression of the two RNAs was closer correlation in (a). malignant than in (b). benign samples.

Supplementary figure 6. Expression of PCA3-shRNA2 in the validation cohort of men, stratified by the presence of prostate cancer and benign prostatic hyperplasia.

Supplementary figure 7. Expression of PCA3-shRNA2 and primary/mature miRs according to nuclear and cytoplasmic localization. The ratio between nuclear and cytoplasmic RNA normalized expression is shown for PCA3-shRNA3 in red and various primary and mature microRNAs for comparison. For each mature short RNA (including PCA3-shRNA2), the majority of the transcript is expressed within the cytoplasm, in contrast to the primary miR hairpin transcript.

Supplementary figure 8. PCA3-shRNA2 in DU145 and SOX11 expression. (a). Expression of PCA3-shRNA2 RNA in cells transfected with the correct sequence, mock transfection with a scrambled RNA sequence and untransfected cells. Bars represent the mean of three independent repeats and standard deviation (error bars). (b). Expression of COPS2 is lower in the urinary cells of

patients with prostate cancer (PCa) when compared to controls (BPH) and is (c). correlated with that of PCA3-shRNA2.