Whole blood mRNA in prostate cancer reveals a four-gene androgen regulated panel

Benjamin C Thomas1,2, Jonathan D Kay1,2,3, Suraj Menon4,5, Sarah L Vowler4,5, Sarah N Dawson4, Laura J Bucklow2, Hayley J Luxton2,3, Thomas Johnston1,2, Charlie E Massie1,6, Michelle Pugh7, Anne Y Warren8, Peter Barker9, Keith Burling9, Andy G Lynch10, Anne George1, Johanna Burge1, Marie Corcoran1, Sara Stearn1, Alastair D Lamb1, Naomi L Sharma1, Greg L Shaw1,7, David E Neal1,2 and Hayley C Whitaker1,2,3

1Uro-Oncology Research Group, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, UK
2Biomarker Initiative, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, UK
3Molecular Diagnostics and Therapeutics Group, University College London, London, UK
4Bioinformatics and Statistics Core Facility, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, UK
5Astra Zeneca, 2 Riverside, Granta Park, Cambridge, UK
6University of Cambridge, Department of Clinical Biochemistry, Cambridge, UK
7Molecular and Computational Diagnostics Group, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, UK
8Genomics Core Facility, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, UK
9Department of Histopathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
10National Institute for Health Research Cambridge Biomedical Research Centre Core Biochemistry Assay Laboratory, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
11Computational Biology Group, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, UK
12University College Hospital, Department of Urology, London, UK
13Nuffield Department of Surgical Sciences, John Radcliffe Hospital, Headington, Oxford, UK

Abstract

Due to increased sensitivity, the expression of circulating nucleotides is rapidly gaining popularity in cancer diagnosis. Whole blood mRNA has been used in studies on a number of cancers, most notably two separate studies that used whole blood mRNA to define non-overlapping signatures of prostate cancer that has become castration independent. Prostate cancer is known to rely on androgens for initial growth, and there is increasing evidence on the importance of the androgen axis in advanced disease. Using whole blood mRNA samples from patients with prostate cancer, we have identified the four-gene panel of FAM129A, MME, KRT7 and SOD2 in circulating mRNA that are differentially expressed in a discovery cohort of metastatic samples. Validation of these genes at the mRNA and protein level was undertaken in additional cohorts defined by risk of relapse following surgery and hormone status. All the four genes were downregulated at the mRNA level in the circulation and in primary tissue, but this was not always reflected in tissue protein expression. MME demonstrated significant differences in the hormone cohorts, whereas FAM129A is downregulated at the mRNA level but is raised at the protein level in tumours. Using published ChIP-seq data, we have demonstrated that this may be due to AR binding at the FAM129A and MME loci in multiple cell lines. These data suggest that whole blood mRNA of androgen-regulated genes has the potential to be used for diagnosis and monitoring of prostate cancer.
Introduction

Prostate cancer is the leading cause of cancer death in men (Office for National Statistics 2013). Prostate cancer growth relies upon androgens, such as testosterone, binding to the androgen receptor (AR) and promoting AR binding to DNA via androgen response elements where it regulates the transcription of androgen-regulated genes (Claessens et al. 2014). Due to the importance of the AR signalling axis, treatment for prostate cancer relies upon blocking the production of androgens or their binding to the AR. Although this treatment is effective, patients will eventually relapse as prostate cancer cells continue to grow, despite continued hormone therapy. This stage of disease was originally termed hormone independent or relapsed, but recent evidence, including the use of drugs to block adrenal androgen synthesis (abiraterone) or binding to the AR (enzalutamide), suggests that the AR axis remains active in advanced prostate cancer and is a valid therapeutic target (Mateo et al. 2014). This is supported by the critical involvement of androgen-regulated genes to predict aggressive prostate cancer in a recently published gene signature found in tissue (Klein et al. 2014).

Diagnostic prostate cancer biomarkers primarily detect differences between benign prostates and those containing primary tumour. Despite the large number of published manuscripts describing biomarkers for detecting prostate cancer, the detection of androgen-regulated prostate-specific antigen (PSA) in serum remains the primary biochemical test used for diagnosis (Partin et al. 1993, Kattan et al. 1998, Egger et al. 2013). PSA is not an ideal prostate cancer biomarker as it can be elevated by a number of benign conditions including benign prostatic hyperplasia and prostatitis. PSA has a sensitivity of approximately 80% but specificity as low as 20%, resulting in false-positive and -negative results (Catalona et al. 1994, Thompson et al. 2005). PCA3, a small non-coding RNA that can be detected in urine, is increasingly being used for prostate cancer diagnosis, particularly in patients with raised PSA and at least one negative biopsy (de Kok et al. 2002, Haese et al. 2008). This is despite the requirement for prostatic massage and reports of relatively poor sensitivity (65%) and specificity (60%) (Hessels & Schalken 2009). Unlike PCA3, PSA has shown some promise in predicting aggressive disease and relapse following treatment, but this remains an area of prostate research where novel markers could significantly have an impact on clinical decisions (Botchorishvili et al. 2009, Augustin et al. 2013).

Over 30 years ago, Leon et al. showed that circulating DNA levels were increased in cancer patients compared with healthy controls (Leon et al. 1977). Circulating nucleic acids as biomarkers have advantages over proteins such as the ability to be amplified and detected with high sensitivity and specificity. Expression arrays and real-time quantitative PCR (qPCR) allow quantification of many genes in a single experiment (Schwarzenbach et al. 2011). More recently, whole blood mRNA has been investigated for the discovery and development of biomarkers, as RNA is more labile and may more accurately reflect any early changes in cells leading to tumour development (Miura et al. 2005, Papadopoulou et al. 2006, Williams 2010). The PAXgene system is used for the storage and purification of RNA from blood and provides storage of blood samples for 50 months at −20°C (Rainen et al. 2002). Its use has enabled the investigation of differences between whole blood mRNA expression levels in patient samples with haematological and rheumatological diseases and breast, thyroid and prostate cancers (Li et al. 2004, Batliwalla et al. 2005, Lewis et al. 2011, Yang et al. 2011).

There have been a number of recent publications on prostate cancer, which utilise PAXgene samples for prognosis to identify those primary prostate cancers that are likely to be life limiting (indolent primary vs aggressive disease) (Liong et al. 2012, Olmos et al. 2012, Ross et al. 2012, Danila et al. 2013, Marin-Aguilera et al. 2015). In back-to-back publications, Olmos et al. and Ross et al. showed different gene signatures, both capable of predicting shortened survival in patients with castration-resistant prostate cancer (Olmos et al. 2012, Ross et al. 2012). Using genes pre-selected by literature review, Ross et al. used unamplified whole blood mRNA to define a six-gene model that separated castrate-resistant patients into a low-risk group with a median survival of 34.9 months compared with a high-risk group with a median survival of 7.8 months (Ross et al. 2012). Using amplified whole blood mRNA, Olmos et al. used Affymetrix expression arrays and Bayesian latent process decomposition to identify a signature that divided their cohort into four groups with survival ranging from 9.2 to 25.6 months (Olmos et al. 2012). That these two manuscripts showed no overlap between the genes in their signature for castration-resistant disease may reflect the use of candidate genes and RNA amplification, which is known to introduce bias in expression arrays in the Olmos, but not the Ross, manuscript (Kitchen et al. 2011). Other groups have used RNA from urine or the peripheral circulation...
to study previously identified and well-characterised markers for prostate cancer using PCR-based methods and found they were robust, particularly when used in conjunction with circulating tumour cells (Quek et al. 2012, Danila et al. 2013).

Here, we describe the identification of a whole blood mRNA gene signature from unamplified PAXgene samples using Illumina HT-12 expression arrays. Genes were identified using the extremes of prostate pathology; a metastatic and benign discovery cohort. Differentially expressed genes were further validated in additional primary cohorts with defined hormonal status or risk stratification. In addition, the gene panel was investigated at both the gene and protein levels in the circulation and primary prostate tissue and AR binding to the gene locus determined.

Materials and methods

Unless otherwise stated, all reagents were purchased from Sigma Aldrich.

Study design

An overview of the study design is given in Fig. 1A. All PAXgene samples were taken from patients enrolled in the ProMPT trial at Addenbrookes Hospital, Cambridge (National Institute of Health and Research ID 5837) between 2010 and 2012. Informed consent was obtained from each patient after the purpose and nature of all procedures used were fully explained to them.

The discovery cohort used PAXgene samples from 22 patients; 12 control patients with an elevated PSA and...
negative TRUS biopsy (benign) and ten patients with metastatic prostate cancer. These patients also had core biopsy specimens taken from the primary prostate available for immunohistochemistry. An additional ten samples from men with raised PSA who had undergone template biopsy were analysed for comparison with the benign control group. The clinical characteristics are given in Table 1.

The gene signature was validated in two cohorts: a defined hormone status cohort and a defined risk cohort. Hormone status was categorised as patients not receiving any hormone treatment (hormone naïve), patients receiving hormone treatment and responding (stable disease) or patients no longer responding to hormone therapy (hormone relapsed). For the risk cohort, patients were defined as at low, intermediate or high risk based on the following clinical criteria: low risk – PSA ≤10ng/mL, Gleason 6, ≥T2; intermediate risk – PSA ≥10 ng/mL but ≤20 ng/mL, Gleason 7, ≥T2 or high risk – PSA ≥20 ng/mL, Gleason ≥8, ≥T3. The clinical characteristics of the risk cohort are given in Table 2.

For this study, hormone relapsed patients are defined as those with two consecutive PSA rises to ≥0.2 ng/mL. The clinical characteristics are given in Table 3.

### RNA extraction and cDNA formation
Whole blood (2.5 mL) was taken from patients using PAXgene tubes (GenProbe, San Diego, CA, USA) and stored according to the manufacturer’s instructions. RNA was extracted using the PAXgene RNA Blood kit (Qiagen) and eluted in 80 µL elution buffer. RNA quantification was performed by absorbance (OD A260 nm) on the Nanodrop ND1000 instrument (Thermo Scientific). RNA (500 ng) per sample was reverse transcribed to cDNA using High-Capacity RNA-to-cDNA Master Mix (Applied Biosystems).

### RNA expression array
Using 22 samples in the discovery cohort (benign (n = 12) and metastatic (n = 10)), gene expression analysis

---

**Table 1** Clinical data for the discovery cohort and template biopsy-confirmed benign cohort.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Metastatic (n = 10)</th>
<th>Benign TURP (n = 12)</th>
<th>Benign template (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age in years* (range)</td>
<td>65.0 (54.0–83.0)</td>
<td>66.5 (52.0–80.0)</td>
<td>61.0 (50.0–67.0)</td>
</tr>
<tr>
<td>Median PSA* (range)</td>
<td>43.0 (3.9–609.0)</td>
<td>7.7 (4.2–18.6)</td>
<td>10.8 (3.8–15.5)</td>
</tr>
<tr>
<td>Gleason sum score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 (%)</td>
<td>0/10 (0%)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>7 (%)</td>
<td>1/10 (20%)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>≥8 (%)</td>
<td>9/10 (90%)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Clinical stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>7/10 (70%)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>T4</td>
<td>3/10 (30%)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Hormone status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hormone naïve</td>
<td>6/10 (60%)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>On hormone therapy</td>
<td>4/10 (40%)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Hormone relapsed</td>
<td>0/10 (0%)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

n=number of patients with percentage given in in brackets afterwards. *Age and PSA data collected at the time of diagnosis.

**Table 2** Clinical data for the defined hormone status cohort.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Low risk (n = 12)</th>
<th>Moderate risk (n = 21)</th>
<th>High risk (n = 14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age in years* (range)</td>
<td>61.0 (51.0–71.0)</td>
<td>66.0 (53.0–87.0)</td>
<td>69.0 (55.0–79.0)</td>
</tr>
<tr>
<td>Median PSA* (range)</td>
<td>5.7 (2.7–10.0)</td>
<td>9.0 (2.1–18.5)</td>
<td>14.0 (4.0–149.0)</td>
</tr>
<tr>
<td>Gleason score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 (%)</td>
<td>12/12 (100%)</td>
<td>0/21 (0%)</td>
<td>1/14 (7%)</td>
</tr>
<tr>
<td>7 (%)</td>
<td>0/12 (0%)</td>
<td>21/21 (100%)</td>
<td>1/14 (7%)</td>
</tr>
<tr>
<td>≥8 (%)</td>
<td>0/12 (0%)</td>
<td>0/21 (0%)</td>
<td>12/14 (86%)</td>
</tr>
<tr>
<td>Clinical stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>12/12 (100%)</td>
<td>19/21 (91%)</td>
<td>7/14 (50%)</td>
</tr>
<tr>
<td>T3</td>
<td>0/12 (0%)</td>
<td>2/21 (9%)</td>
<td>7/14 (50%)</td>
</tr>
<tr>
<td>T4</td>
<td>0/12 (0%)</td>
<td>0/21 (0%)</td>
<td>0/14 (0%)</td>
</tr>
<tr>
<td>+ve biopsy cores (%)</td>
<td>21% (6–67%)</td>
<td>31% (8–73%)</td>
<td>78% (17–100%)</td>
</tr>
</tbody>
</table>

n=number of patients with percentage given in in brackets afterwards. *Age and PSA data collected at the time of diagnosis.
was carried out on Illumina Human HT12 version 4 arrays (Illumina, San Diego, CA, USA). Data analyses were carried out on R using Bioconductor packages (Gentleman 2004). Raw intensity data from the array scanner was processed using the BASH and HULK algorithms as implemented in the beadarray package (Dunning 2007, Cairns 2008). Log2 transformation and quantile normalisation of the data were performed across all sample groups. Differential expression analysis was carried out using the limma package (Smyth 2005). Genes were selected for further analysis by ranking according to an unadjusted $P$-value and a cut-off of 0.01 applied to stratify genes more likely to be significantly altered. Following interrogation of the top-ranked genes, data were ranked by $P$-value and then fold change. Further stratification of genes was completed using published datasets to identify differentially expressed genes associated with poor prognosis (Glinsky et al. 2004, Taylor et al. 2010). Genes with published antibodies for immunohistochemistry or good differential expression between tumour and benign glands in Human Protein Atlas were preferentially selected. How genes were selected and what criteria were taken into account are shown in Supplementary Table 2. A false discovery rate (FDR) of 0.39 was calculated for the top 100 probes indicating that only 61/100 were likely to be true positives and that additional validation was required to confirm any candidate gene.

### Quantitative real-time polymerase chain reaction

Quantitative real-time polymerase chain reaction (qPCR) was performed in triplicate using 2 pmol Sigma primers, 5 µL SYBR Green and 10 ng cDNA template using the Applied Biosystems 7900HT Real-Time PCR system. Primer sequences are given in Supplementary Table 3. Ct values were calculated and the expression of target genes was normalised against the expression of RPLP2 housekeeping gene using the $\delta\delta$Ct method. RPLP2 was chosen as a housekeeping gene by analysing all genes in the expression array with high expression and minimal variation (Supplementary Fig. 1, see section on supplementary data given at the end of this article).

### Chromatin immunoprecipitation

Chromatin was immunoprecipitated from LNCaP cells treated with either 1 nM R1881 or vehicle (ethanol) for 1 h as described in Massie and Mills (2009) using 10 µg AR n-20 antibody (sc-816; Santa Cruz Biotechnology) and an equal mixture of Dynabeads Protein A and G (10001D and 10003D; Life Technologies). Primer sequences are shown in Supplementary Table 4. Results are shown relative to the vehicle control. $P$-values were calculated using Wilcoxon rank-sum test.

### Immunohistochemistry

For immunohistochemistry, tissue cores taken from the primary prostate of patients within the discovery cohort were used. In addition, a TMA was constructed from multiple benign and tumour samples from 104 patients who underwent radical prostatectomy as described previously (Whitaker et al. 2010, 2013). Immunohistochemistry was also performed on a hormone relapsed TMA comprising 74 patients. Nine of these patients were on hormone treatment and

#### Table 3 Clinical data for the risk stratified.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Hormone naïve ($n=9$)</th>
<th>Stable disease ($n=12$)</th>
<th>Hormone relapsed ($n=12$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age in years* (range)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median PSA* (range)</td>
<td>75.0 (64.0–93.0)</td>
<td>66.5 (61.0–88.0)</td>
<td>71.0 (51.0–82.0)</td>
</tr>
<tr>
<td>Gleason sum score</td>
<td>123.0 (12.5–511.6)</td>
<td>85.0 (7.4–6470.0)</td>
<td>41.1 (2.0–405.0)</td>
</tr>
<tr>
<td>No biopsy taken, diagnosis based on clinical criteria</td>
<td>3/9 (33%)</td>
<td>6/12 (50%)</td>
<td>4/12 (34%)</td>
</tr>
<tr>
<td>6</td>
<td>0/10 (0%)</td>
<td>1/12 (8%)</td>
<td>1/12 (8%)</td>
</tr>
<tr>
<td>7</td>
<td>1/10 (11%)</td>
<td>3/12 (25%)</td>
<td>1/12 (8%)</td>
</tr>
<tr>
<td>≥8</td>
<td>5/10 (56%)</td>
<td>2/12 (17%)</td>
<td>6/12 (50%)</td>
</tr>
<tr>
<td>Clinical stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>0/9 (0%)</td>
<td>1/12 (8%)</td>
<td>1/12 (8%)</td>
</tr>
<tr>
<td>T3</td>
<td>7/9 (77%)</td>
<td>9/12 (75%)</td>
<td>8/12 (67%)</td>
</tr>
<tr>
<td>T4</td>
<td>3/10 (33%)</td>
<td>2/12 (17%)</td>
<td>3/12 (25%)</td>
</tr>
</tbody>
</table>

$n=$ number of patients with percentage given in brackets afterwards. *Age and PSA data collected at the time of diagnosis. The median proportion of positive cores is also given for each group with the range shown in brackets.
Table 4  Clinical data for the hormone-relapsed (HR) TMA generated in Cambridge.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Number of patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHRH, no progression (hormonal therapy cohort)</td>
<td>9 (15.2)</td>
</tr>
<tr>
<td>Relapse on LHRH</td>
<td>19 (32.2)</td>
</tr>
<tr>
<td>Relapse on complete androgen blockade</td>
<td>18 (30.5)</td>
</tr>
<tr>
<td>Relapse following external beam radiation</td>
<td>9 (15.2)</td>
</tr>
<tr>
<td>Salvage TURP after relapse</td>
<td>4 (6.8)</td>
</tr>
</tbody>
</table>

n= number of patients with percentage given in brackets afterwards.

response, while 50 initially responded to treatment but subsequently underwent biochemical or clinical relapse as described previously (Ramos-Montoya et al. 2014). The clinical characteristics are given in Table 4.

All immunohistochemistry was performed using a Bondmax Autostainer (Leica) using the following antibodies: FAM129A (1:750; Atlas Antibodies, Stockholm, Sweden), KRT7 (2.8 g/L; Novocastra, Solms, Germany), SOD2 (1:1500; Atlas Antibodies) and MME (1:50; Novocastra). Citrate buffer pH 6.0 (ER1) was used for all antigen retrieval. KRT7 has been validated and is sold for routine use in the clinical setting using the same conditions and Bondmax Autostainer (Addenbrookes Hospital). The specificity of antibody staining by FAM129A, SOD2 and MME was validated using qPCR and immunohistochemistry of formalin-fixed paraffin-embedded LNCaP cell pellets transiently transfected with siRNA against each target or a non-targeting control (Supplementary Fig. 2). Loss of staining in the specific cellular compartment in up to 50% of cells in the siRNA group was considered specific. Confirmation of pathology was conducted by a uropathologist (AW) for all samples. All scoring was done independently by two observers (one an independent specialist and the other a uro-oncology pathologist) who were blinded to the TMA plan. Staining was classified into the following categories: none (−), weak (+), moderate (++), and high (+++), based on intensity. For the TMA, where multiple pathologies were present in a single core, each was scored independently and multiple scores given for distinct grades of cancer. For core biopsies where cancer was present, it was scored independently of the surrounding benign tissue and a single score given for each patient.

**ELISA**

All assays were performed on the Meso Scale Discovery (MSD) platform (Maryland, USA). Normal bind plates were coated at 4°C overnight with 30 μL goat anti-MME antibody (1.44 μg/mL; R&D Systems) before washing and blocking with MSD blocker A. After diluting 1:5 in Delfia diluent II (PerkinElmer), 50 μL serum sample or standard were added to each well and incubated for 2 h at room temperature. After washing, 25 μL biotinylated goat anti-MME antibody (1:100; R&D Systems) were added per well and incubated at room temperature for 1 h. Signal was detected using streptavidin sulphoTAG (1:1000, MSD) incubated for 30 min at room temperature. The wells were washed and 150 μL MSD read buffer added before detection on a Sector 6000 plate reader. Standards ranging from 39 to 40,000 pg/mL were made using recombinant MME (R&D Systems) and Delfia diluent II. A standard curve was run alongside all samples to ensure reproducibility of the assay.

**Androgen regulation of LNCaP cells and AR ChIP-seq**

Gene expression data were generated as part of a previously published study (Massie et al. 2011). Briefly, LNCaP cells grown for 72 h in steroid depleted the medium before treatment with the synthetic androgen R1881. Samples included three time zero samples; ten vehicle (ethanol) control samples taken at 2, 4, 8, 12 and 24 h in duplicate; and 36 androgen (R1881)-treated samples taken every 30 min for 4 h and then every hour until 24 h following treatment (with replicates at 1, 2, 4, 8, 12, 16, 20 and 24 h). Gene expression data were analysed as mentioned earlier (Massie et al. 2011). The AR ChIP-seq experiments have previously been described (Wang et al. 2009, Yu et al. 2010, Massie et al. 2011).

**Data analysis and modelling**

Grouped P-values were calculated using Kruskal–Wallis test. Pairwise analysis was performed using Mann–Whitney two-tailed t-test. For both analysis, a P-value of <0.05 was considered to be statistically significant. Previously published expression array data from prostate tissue were analysed using recursive partitioning to predict time to biochemical recurrence from 179 patients (29 benign, 131 localised prostate cancer and 19 metastatic disease) (Taylor et al. 2010). Recursive partitioning was also used to determine qPCR cut-offs between groups in the defined hormone status cohort and the defined risk cohort using the ctree function in R (Hothorn 2006). Models were fitted for all four genes and serum PSA individually.
Results

Identification of genes capable of discriminating between benign and metastatic patients

To discover if any gene demonstrated significant differential expression between benign and tumour cohorts, a discovery cohort was selected, which represented the extremes of prostate disease: ten metastatic and 12 benign samples. To ensure that there were no fundamental differences between the blood samples from each group, baseline haematological parameters such as haemoglobin and white blood cells were assessed and were not significantly different (Supplementary Fig. 3).

The mRNA from the discovery cohort samples were assessed using Illumina Human HT12 v4 expression arrays with and without haemoglobin reduction, which has been reported to influence measurement of accurate expression using gene arrays (Fan & Hegde 2005, Vartanian et al. 2009). Although the signal-to-noise ratio was much lower in the haemoglobin-reduced samples, when candidate genes were validated by qPCR using the original PAXgene samples without haemoglobin reduction, they were not significant, suggesting that haemoglobin mRNA reduction had altered the samples significantly (Supplementary Fig. 1A and data not shown). The 100 top differentially expressed genes in the non-haemoglobin-reduced samples are shown in Supplementary Fig. 4. Six genes were taken forward for further analysis: family with sequence similarity 129, member A (FAM129A) also known as NIBAN, pyrophosphatase (inorganic) 1 (PPA1), dolichyl-phosphate mannosyltransferase poly peptide 3 (DPM3), superoxide dismutase-2 (SOD2), keratin-7 (KRT7) and membrane metalloendopeptidase (MME) also known as CD10. Although unrelated, all of these genes have been previously associated with cancer, and DPM3, SOD2, KRT7 and MME have previously been linked to prostate cancer (Supplementary Table 1). The genes identified in previous PAXgene studies failed to cluster our discovery cohort into benign and metastatic samples, probably as a result of different sample processing methods (data not shown) (Olmos et al. 2012, Ross et al. 2012).

Figure 1B shows a heat map of microarray data of probes for the six genes that were amongst the top 1000 most differentially expressed probes. Boxplots of these probes are shown in Fig. 1C for the benign and metastatic groups, along with the P-values from limma tests for differential expression between the groups. Further validation of these results using qPCR of fresh cDNA from the same patients showed that three genes, KRT7, DPM3 and PPA1, were not significantly different between benign and metastatic groups despite DPM3 being previously linked to prostate cancer cell invasion (Fig. 1D) (Manos et al. 2001). This result was not unexpected given the high FDR seen in the array probably due to the wide clinical parameters in the discovery cohort, non-specific binding to probes on the expression array or the increased sensitivity of qPCR. As KRT7 is routinely used as a basal cell marker in prostate cancer, we took the decision to allow it to remain in the selected gene panel (Supplementary Table 1) (Manos et al. 2001, Ashida et al. 2006). The remaining three genes FAM129A, SOD2 and MME achieved statistical significance (P<0.05). The benign group within the discovery cohort had raised PSA and at least one negative TRUS biopsy and therefore it was possible that up to 30% may go on to have a later diagnosis of prostate cancer (Shariat & Roehrborn 2008). To establish that this benign cohort was a valid control group, qPCR for FAM129A, SOD2, MME and KRT7 was compared between the TRUS-confirmed benign group and a cohort of benign prostates confirmed by template biopsy with a much lower probability of having cancer (<10%). This showed no significant differences and confirmed the benign group as a valid control (Supplementary Fig. 5).

Circulating mRNA may reflect gene expression and protein levels in tumour tissue

It is hypothesised that the mRNA collected in PAXgene samples could be influenced by expression of specific genes in primary tumour tissue or be altered due to the immune response (Olmos et al. 2012, Ross et al. 2012, Marin-Aguilera et al. 2015). As no tissue RNA was available from our discovery cohort, we examined gene expression of the four genes identified in a published independent cohort (Taylor et al. 2010). Expression of all four genes, FAM129A, SOD2, KRT7 and MME, was significantly reduced in metastatic samples when compared with benign, consistent with the results we found in circulating mRNA (Fig. 2A). To establish whether changes in circulating mRNA gene expression and tissue mRNA expression accurately reflected protein expression in the tumour, the diagnostic biopsies from the benign prostates or primary tumours of the discovery cohort were examined using immunohistochemistry (IHC) for FAM129A, SOD2, MME and KRT7 (Fig. 2B). FAM129A showed cytoplasmic staining and expression was reduced, albeit not significantly (P=0.17), in metastatic biopsy tissue when compared with benign. SOD2 showed ubiquitous punctate staining, consistent with its known mitochondrial localisation (Wispe et al. 1989). In contrast
to the mRNA expression and published data, protein levels increased in metastatic samples compared with benign, although not significantly ($P=0.070$) (Bostwick et al. 2000). Both KRT7 and MME appeared to be associated with the plasma membrane, consistent with previous publications (Glass & Fuchs 1988, Zheng et al. 2010). Protein expression of both decreased significantly with the metastatic group consistent with the whole blood mRNA (Fig. 1). However, detection of MME protein in serum was not consistently reduced when measured by ELISA using samples from the same patients (Fig. 2C).

### Localised disease stratified by risk

Patients with localised prostate cancer that may be destined for surgery can be stratified into risk groups for relapse using a number of criteria including PSA, stage and grade (D’Amico et al. 1999). Using recursive partitioning of the risk cohort, serum PSA could distinguish low and high groups with a $\geq 11.3$ ng/mL cut-off. We stratified 47 patients into low-, intermediate- and high-risk groups as defined in the “Materials and methods” section and tested their whole blood mRNA levels of FAM129A, KRT7, SOD2 and MME (Fig. 3A). Only circulating KRT7 and MME mRNA demonstrated any significant ability to discriminate between the risk groups. KRT7 was able to discriminate between low- and intermediate-risk groups ($P=0.0061$), and this was confirmed using recursive partitioning where nodes represent the point at which the most significant differences between the two arms exist (cut-off $\geq 0.00042$) (Fig. 3B). Overall PSA $>11.3$ ng/mL performs better than KRT7, but with a PSA $\leq 11.3$ ng/mL, they perform similarly. Both KRT7 and MME could differentiate between intermediate- and high-risk groups ($P=0.0019$ and $P=0.011$, respectively). This result was consistent with significant decreases in
tissue expression of KRT7 and MME in localised prostate cancer from previously published data (Fig. 3C, P<0.0001 and P=0.013, respectively) (Taylor et al. 2010). In the same dataset, FAM129A exhibited no change between benign and localised tissue groups, whereas SOD2 expression was raised in controls (P=0.018) consistent with our findings at the protein level in the primary tissue biopsies (Fig. 2B).

Low expression of FAM129A, SOD2 and MME mRNA was able to predict poor prognosis with multiple probes in data published by Taylor et al. (2010), whereas low tissue
mRNA expression of FAM129A, KRT7 and SOD2 was also linked to poor outcome in data published by Glinsky et al. (2004) (Supplementary Fig. 6).

To establish whether changes in tissue mRNA and circulating mRNA in our PAXgene samples reflected protein expression in primary prostate tissue, IHC was performed using an in-house Cambridge TMA for FAM129A, KRT7, SOD2 and MME proteins (Fig. 3D). All four proteins showed a significant difference between the benign and Gleason groups ($P<0.0001$). In contrast to circulating mRNA expression and primary tissue biopsies from metastatic patients, protein levels of FAM129A showed a significant increase in localised tumours compared with benign tissue. KRT7, SOD2 and MME exhibited a decrease in protein expression when compared with benign tissue. This decrease was particularly striking for KRT7 and MME.

### Androgen regulation of the four-gene panel

Within the discovery cohort, the metastatic patients were receiving a range of treatments that may have influenced the circulating mRNA levels of different genes. We used a well-defined hormone sensitivity cohort to determine the effect of hormone treatments on the circulating mRNA levels of the four-gene panel of FAM129A, SOD2, KRT7 and MME (clinical data are given in Table 3). These patients had never received hormone treatment (hormone naïve), were receiving hormone treatment and responding (hormonal therapy), or were no longer responding to hormone therapy (hormone relapsed). Using a Kruskal–Wallis test, detection of the mRNA coding for all four genes was significantly different between all groups for each gene ($P<0.05$) (Fig. 4A). Circulating expression of mRNA coding for FAM129A, SOD2 and MME showed a significant reduction in those patients responding to hormones compared with those receiving no treatment that may indicate androgen regulation of those genes. To determine cut-off points associated with hormone status, recursive partitioning of the qPCR data was performed (Fig. 4B). This was consistent with recursive partitioning of FAM129A, SOD2 and MME, demonstrating that the hormone-naïve (HN) group could be distinguished from the hormone-treated groups (cut-offs >0.337 (FAM129A), 1.162 (SOD2) and 0.054 (MME)). None of the four genes

![Figure 4](image_url)

**Figure 4**

Hormone regulation of FAM129A, KRT7, SOD2 and MME in patient cohorts. (A) The circulating mRNA for the four genes was determined in three hormone cohorts by qPCR: patients yet to receive treatment (hormone naive), patients receiving hormone treatment and continuing to respond (hormonal therapy), and patients who have become hormone refractory (hormone relapsed). Ct values were calculated for all conditions and the expression of target genes was normalised against the expression of RPLP2 housekeeping gene using the $\delta\delta$Ct method. Kruskal–Wallis tests for each gene tested the probability of statistically significant differences between the groups: FAM129A – $P=0.0085$, KRT7 – $P=0.038$, SOD2 – $P=0.014$ and MME – $P=0.0079$. All $P$-values shown were calculated using Mann–Whitney two-tailed $t$-test between two groups. (B) Recursive partitioning was performed using Ct values from the qPCR validation to predict cut-offs for each group within the data. Only genes showing significant results are shown. Ct cut-off values and 95% confidence intervals are indicated. HN – never received hormone treatment (hormone naïve), HT – receiving hormone treatment and responding (hormonal therapy), and HR – receiving hormone treatment and showing biochemical or clinical relapse (hormone relapsed). (C) Alterations in the protein levels of FAM129A, KRT7, SOD2 and MME with hormone status was determined by IHC using a hormone relapsed TMA described in Supplementary Table 7. All IHC was performed on the Bondmax Autostainer using conditions described previously. Staining is shown in brown with nuclei shown in blue and was classified into the following categories: none, weak, moderate and high, based on intensity. $P$-values across all groups were calculated using Kruskal–Wallis test and pairwise comparisons using a Mann–Whitney two-tailed $t$-test. A full colour version of this figure is available at http://dx.doi.org/10.1530/ERC-16-0287.
were able to distinguish hormone-relapsed patients from those responding to treatment using Mann–Whitney test. However, even in the small numbers of patients in these cohorts, recursive partitioning of the data for MME showed some ability to distinguish patients on hormonal therapy (HT, node 2) from hormone-relapsed patients (HR, nodes 4 and 5) (Fig. 4A and B).

To establish if these changes in circulating mRNA reflected protein expression in a defined hormone cohort, we performed IHC for FAM129A, KRT7, SOD2 and MME on a hormone-relapsed TMA comprising patients who were responding to hormone therapy (HT) or relapsed following hormone treatment (hormone relapsed (HR)) (Table 4). These results were compared with those of the untreated hormone-naïve patients used in the in-house Cambridge TMA (Fig. 3D). KRT7, SOD2 and MME, but not FAM129A, showed highly significant differences across the three groups. There were no statistically significant pairwise analyses for either FAM129A or KRT7 at the protein level. However, both SOD2 and MME showed significant differences at the protein level when patients were treated with hormone (hormone naive vs on hormonal therapy) ($P<0.0001$ (SOD2) and $P=0.0028$ (MME)).

Figure 5
Androgen regulation of the four-gene panel. (A) Androgen regulation of the four genes was determined in the LNCaP cell line by treating with androgens and taking samples for expression analysis over the following 24 h (Massie et al. 2011). Filled circles represent androgen treatment, open circles represent vehicle controls. Data for all good probes are shown for FAM129A. Data for other genes are shown in Supplementary Fig. 6. ACF = autocorrelation (function). A measure greater than zero means that consecutive time points are nearer each other than time points chosen at random. (B) UCSC genome browser view of the FAM129A and MME loci showing AR binding profiles in prostate cancer cells from three independent studies (Wang et al. 2009, Yu et al. 2010, Massie et al. 2011). Coloured blocks indicate AR peaks identified in each of the three chIP studies, red peaks show the AR ChIP-seq profile for two cell lines (Massie et al. 2011). Below gene annotations are ENCODE tracks indicating promoter, enhancer, DNase hypersensitivity and transcription factor binding profiles. Arrows indicate promoters and direction of transcription. Data for KRT7 and SOD2 are shown in Supplementary Fig. 7. (C) ChIP for the AR in the LNCaP cell line following starvation for 48h in charcoal-stripped media and treatment with $10^{-8}$M R1881 for 1 h (black bars). Data are shown as relative to vehicle-treated cells (grey bars). $P$-values are calculated using Wilcoxon rank-sum test. A full colour version of this figure is available at http://dx.doi.org/10.1530/ERC-16-0287.
To determine whether these observed changes in circulating mRNA and tissue protein were a result of androgen regulation of the four-gene panel, we examined data from an androgen time course of the LNCaP cell line which showed that only FAM129A expression was decreased following androgen treatment, consistent with our data (Fig. 5A and Supplementary Fig. 7). To determine if the AR binds to potential binding sites in the promoters of the four-gene panel, we examined published ChIP-seq data from a variety of cell lines (Wang et al. 2009, Yu et al. 2010, Massie et al. 2011). These results showed convincing AR binding in multiple prostate cancer cell lines for FAM129A and MME (Fig. 5B). There was evidence for AR binding at the KRT7 locus, but only in the two ERG-positive VCaP cell line studies suggesting possible ERG involvement in KRT7 regulation (Yu et al. 2010, Massie et al. 2011) (Supplementary Fig. 7). AR binding to the SOD2 locus was seen in two cell lines (LNCaP and VCaP), but only in data produced in our laboratory (Massie et al. 2011). To confirm AR binding at the promoters of these genes, chromatin immunoprecipitation (chIP) was performed in the LNCaP cell line using an anti-AR antibody (Fig. 5C). Following 1 h of treatment with the synthetic androgen, R1881, AR binding was significantly enriched at the FAM129A, KRT7 and MME promoters, but not at the SOD2 promoter.

Discussion

The majority of diagnostic biomarkers are identified through studies comparing benign and primary prostate cancer. All previous studies using PAXgene samples to determine circulating mRNA signatures in prostate cancer have examined primary indolent vs aggressive prostate cancer either using pre-selected genes of interest or by amplifying the mRNA that is known to introduce bias (Kitchen et al. 2011, Olmos et al. 2012, Ross et al. 2012, Danila et al. 2013, Marin-Aguilera et al. 2015). In this study, we have described the identification of genes initially using a benign vs metastatic comparison using Illumina Human HT12 v4 expression arrays without amplification or use of haemoglobin reduction, which has been reported to improve data quality in PAXgene studies (Fan & Hegde 2005, Vartanian et al. 2009). Although we found that haemoglobin reduction did improve the data quality as defined by Illumina (Supplementary Fig. 1), all subsequent qPCR validations in non-haemoglobin-reduced samples demonstrated that the mRNA in haemoglobin-reduced samples did not accurately reflect the mRNA in the biological sample.

Using expression arrays and native samples, we compared the two extremes of prostate pathology in the discovery cohort comprising metastatic vs benign samples (Table 1) and identified six genes (FAM129A, PPA1, DPM3, SOD2, KRT7 and MME) for further study (Supplementary Table 1). Of these genes, DPM3, SOD2, KRT7 and MME have previously been linked to prostate cancer with DPM3 and MME also linked to aggressive disease, although no common pathway links these genes (Manos et al. 2001, Voutsadakis et al. 2012, Ho et al. 2013). KRT7 is widely used in cancer diagnosis as a marker for basal cells, which are known to be lost in prostate cancer during neoplastic transformation (Brimeo & Epstein 2012). Previous studies have suggested that loss of SOD2 expression is an early event in tumour development, providing an environment for increased free radicals and the development of mutations (Hempel et al. 2011). However, high levels of SOD2 have been reported in metastatic tissue and circulating prostate tumour cells where they are thought to enhance metastases (Hempel et al. 2011, Giesing et al. 2012). MME protein is reduced in extra-capsular vs organ-confined prostate cancer, and MME\textsubscript{high}/AGR2\textsubscript{low} expression is more closely associated with primary disease, whereas MME\textsubscript{low}/AGR2\textsubscript{high} is more commonly seen in metastatic tissue (Kim et al. 2012, Ho et al. 2013). There have also been previous reports of FAM129A overexpression in renal and head and neck cancers where it has been shown to regulate cell stress and p53 degradation to promote cancer cell survival (Adachi et al. 2004, Cerutti et al. 2006, Matsumoto et al. 2006, Patel et al. 2011, Ji et al. 2012). We also reported a high FDR for our discovery cohort (0.39), indicating that although we were enriching for differentially expressed genes, only 61/100 were likely to be positive. We believe that this is most likely due to the wide range of clinical characteristics in the discovery cohort that had a range of hormone treatments and risk stages. Therefore, further examination of any identified genes in well-defined cohorts was essential.

Four of these genes (FAM129A, KRT7, SOD2 and MME) were characterised in circulating mRNA, tissue mRNA and tissue protein levels in a defined hormone cohort and a risk-stratified cohort based upon stage, Gleason grade and PSA data (D’Amico et al. 1999) (Fig. 1D and Tables 2 and 3). The data generated in all of these cohorts are summarised in Table 5. All of the genes that we identified were downregulated in prostate cancer, which presents...
challenges in using them as practical biomarkers, as technical assay issues can easily give rise to false-positive results. The sensitivity of the methodology used for detection could limit the detection of aggressive forms of cancer, and loss of signal due to epigenetic changes or mutations cannot always be ruled out.

There was concordance in the expression of FAM129A, KRT7, SOD2 and MME in circulating mRNA and published metastatic tissue mRNA, suggesting that circulating mRNA may accurately reflect expression that occurs in metastatic tumour tissue either directly or via an immune response (Figs 1C, D and 2A) (Taylor et al. 2010). This is consistent with the overexpression of a number of genes identified in circulating tumour cells derived from the tumour tissue (Marin-Aguilera et al. 2015). In our study, downregulation of these four genes could also be confirmed at the protein level in primary tissue biopsies from metastatic men for KRT7 and MME but not SOD2 or FAM129A (Fig. 2B). This may reflect the small sample numbers tested or the effects of variable fixation on thin needle core biopsies. These results were supported by the downregulation of KRT7, SOD2 and MME protein levels in localised tumours, which suggests that these three genes, and subsequently proteins, are all downregulated in all tumours compared with benign (Fig. 3D and Table 5). Although FAM129A protein was downregulated in biopsies from metastatic patients, this result was not significant (P=0.17), and subsequent analysis in localised prostate cancer demonstrated a clear upregulation in primary tumours (Figs 2B and 3D). For SOD2, the conflicting protein level data between primary biopsy of metastatic disease and localised disease are consistent with other conflicts in metastatic tissue and circulating tumour cells vs primary disease (Hempel et al. 2011, Giesing et al. 2012). There were no significant changes in MME protein levels in both the discovery and the defined risk cohorts, as well as in serum protein levels (Figs 1, 2C and 3). This may reflect the greater sensitivity frequently seen in genomic assays (Miura et al. 2005, Papadopoulou et al. 2006, Williams 2010). Despite reports of changes in MME protein in extra-capsular disease, we saw no evidence of significant differences between localised and metastatic cohorts. It is possible that without the addition of the known prostate biomarker AGR2, MME does not have sufficient power alone (Bu et al. 2011, Kani et al. 2013).

We hypothesised that some of the differences seen between the metastatic and localised cohorts were as a result of androgen regulation of the identified genes. This was investigated using published ChIP-seq data, which showed strong AR binding to the FAM129A and MME loci in multiple cell lines consistent with previous reports of altered regulation of FAM129A in cells treated with the anti-androgen, bicalutamide (Rothermund et al. 2005, Wang et al. 2009, Yu et al. 2010, Massie et al. 2011). This also suggests that MME and FAM129A are novel androgen-regulated genes, and this was confirmed by demonstrating enrichment of AR binding to the MME, FAM129A and KRT7 promoters following treatment with androgen (Fig. 5C). These data suggest that FAM129A expression at the mRNA and protein level may initially increase in hormone-sensitive tissue before decreasing in patients with hormone relapsed disease. This is consistent with an alternatively regulated AR signalling axis and alterations in AR binding sites previously reported in advanced disease (Tsao et al. 2012, Sharma et al. 2013). This supports the inclusion of androgen-regulated genes in a recently published 17-gene panel to predict aggressive prostate cancer (Klein et al. 2014).

This study has identified a four-gene panel of circulating mRNA that can distinguish metastatic from benign disease in a small patient cohort. Although there was a frequent concordance between circulating mRNA and tissue mRNA, there were frequently differences between expression of FAM129A, KRT7, SOD2 and MME
at the mRNA and protein level in many of the cohorts. Androgen regulation of FAM129A, MME and KRT7 confirms the continuing role of the AR in advanced prostate cancer. This study demonstrates altered AR-regulated gene expression during different stages of the prostate cancer disease pathway that can be detected using a minimally invasive blood test. Further work is required to determine if an AR-regulated gene panel, either alone or in combination with genes identified in other PAXgene studies, could have clinical utility in predicting aggressive disease or monitoring response to hormone-manipulating treatments such as bicalutamide, abiraterone and enzalutamide.

Supplementary data
This is linked to the online version of the paper at http://dx.doi.org/10.1530/ERC-16-0287.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
The authors acknowledge the support of the University of Cambridge, Cancer Research UK, and Hutchison Whampoa Limited in completing this study. They would also like to acknowledge the support of the National Institute for Health Research (NIHR) that funds the Cambridge Bio-medical Research Centre, Cambridge UK, and the NCRI (ProMPT), which has funded tissue collections in Cambridge, and PCUK that has funded H W and J K for a proportion of the time spent working on this manuscript.

Author contribution statement
B T performed expression arrays, qPCR and clinical data analysis and wrote sections of the manuscript. J K performed haemoglobin reduction experiments, assisted IHC scoring, contributed to data analysis and writing. S M performed bioinformatics for haemoglobin reduction and illumina expression arrays. S V and S D performed all statistical analysis and contributed to manuscript. L B performed IHC and assisted in scoring, assisted in the development of the MME ELISA and data analysis. H L performed the chIP experiments, analysed the data generated and wrote the relevant section of the manuscript. T J performed template biopsy vs benign controls, collated and analysed associated clinical data. C M and A L analysed data from LNCaP expression and ChIP-seq studies and provided contributions to the manuscript; M P performed expression arrays and provided advice on haemoglobin reduction experiments. A W involved in IHC scoring, producing TMAs, staging and grading all patients and contributed to manuscript preparation. K B developed and performed MME ELISA on MSD platform and contributed to manuscript writing. A G, J B, M C and S S obtained informed consent from all patients, collected patient samples, and collected and helped analyse all clinical data; A L, N S and G S collected samples and analysed clinical data for the HR TMA. D N helped analyse clinical data and contributed to manuscript. H W supervised all work, IHC scoring, analysed data, wrote manuscript and contributed some funding.

Acknowledgements
The authors would like to thank the Genomics Core Facility and Histopathology and in Situ Hybridisation Core Facility, Cancer Research UK Cambridge Institute, for their assistance with the work described in this manuscript.

References
Hempen N, Carrico PM & Melendez JA 2011 Manganese superoxide dismutase (Sod2) and redox-control of signaling events that drive metastasis. Anti-Cancer Agents in Medicinal Chemistry 11 191–201. (doi:10.2174/1568026097895259911)


Received in final form 4 August 2016
Accepted 10 August 2016