BJUI
BJU International

Urological Oncology

A four-group urine risk classifier for predicting outcomes in patients with prostate cancer

Shea P. Connell*, Marcelino Yazbek-Hanna*, Frank McCarthy†, Rachel Hurst*, Martyn Webb*, Helen Curley*, Helen Walker‡, Rob Mills‡, Richard Y. Ball‡, Martin G. Sanda§, Kathryn L. Pellegrini§, Dattatraya Patil§, Antoinette S. Perry¶, Jack Schalken**, Hardev Pandha††, Hayley Whitaker‡‡, Nening Dennis†, Christine Stuttle†, Ian G. Mills§§¶¶***, Ingrid Guldvik¶¶, Movember GAP1 Urine Biomarker Consortium[1], Chris Parker†††, Daniel S. Brewer*‡‡‡, Colin S. Cooper* and Jeremy Clark*

*Norwich Medical School, University of East Anglia, Norwich, †Institute of Cancer Research, Sutton, ‡Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK, §Department of Urology, Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA, USA , ¶School of Biology and Environmental Science, Science West, University College Dublin, Dublin 4, Ireland , **Nijmegen Medical Centre, Radboud University Medical Centre, Nijmegen, The Netherlands, ††Faculty of Health and Medical Sciences, The University of Surrey, Guildford, ‡‡Molecular Diagnostics and Therapeutics Group, University College London, London, §§School of Medicine, Dentistry and Biomedical Sciences, Institute for Health Sciences, Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, UK, ¶¶Centre for Molecular Medicine, University of Oslo, Oslo, Norway, ***Nuffield Department of Surgical Sciences, University of Oxford, Oxford , †††Royal Marsden Hospital, Sutton, and ‡‡‡Earlham Institute, Norwich, UK

S.P.C., M.H. and F.M. are joint first authors.
D.S.B., J.C. and C.S.C. are joint senior authors.
[1]The Movember GAP1 Urine Biomarker Consortium members are listed in Appendix.
[Correction added on 09 July 2019, after first online publication: An author name has been updated in this version.]

Objectives

To develop a risk classifier using urine-derived extracellular vesicle (EV)-RNA capable of providing diagnostic information on disease status prior to biopsy, and prognostic information for men on active surveillance (AS).

Patients and Methods

Post-digital rectal examination urine-derived EV-RNA expression profiles ($n = 535$, multiple centres) were interrogated with a curated NanoString panel. A LASSO-based continuation ratio model was built to generate four prostate urine risk (PUR) signatures for predicting the probability of normal tissue (PUR-1), D'Amico low-risk (PUR-2), intermediate-risk (PUR-3), and high-risk (PUR-4) prostate cancer. This model was applied to a test cohort ($n = 177$) for diagnostic evaluation, and to an AS sub-cohort ($n = 87$) for prognostic evaluation.

Results

Each PUR signature was significantly associated with its corresponding clinical category ($P < 0.001$). PUR-4 status predicted the presence of clinically significant intermediate- or high-risk disease (area under the curve = 0.77, 95% confidence interval [CI] 0.70–0.84). Application of PUR provided a net benefit over current clinical practice. In an AS sub-cohort ($n = 87$), groups defined by PUR status and proportion of PUR-4 had a significant association with time to progression (interquartile range hazard ratio [HR] 2.86, 95% CI 1.83–4.47; $P < 0.001$). PUR-4, when used continuously, dichotomized patient groups with differential progression rates of 10% and 60% 5 years after urine collection (HR 8.23, 95% CI 3.26–20.81; $P < 0.001$).

Conclusion

Urine-derived EV-RNA can provide diagnostic information on aggressive prostate cancer prior to biopsy, and prognostic information for men on AS. PUR represents a new and versatile biomarker that could result in substantial alterations to current treatment of patients with prostate cancer.

Keywords

biomarker, urine, active surveillance, liquid biopsy, cell free, #ProstateCancer, #PCSM

Introduction

The progression of prostate cancer is highly heterogeneous [1], and risk assessment at the time of diagnosis is a critical step in the management of the disease. Based on the information obtained prior to treatment, key decisions are made about the likelihood of disease progression and the best course of treatment for localized disease. D'Amico

stratification [2], which classifies patients as having low, intermediate or high risk of PSA failure post-radical therapy, is based on Gleason score (GS) [3], PSA and clinical stage, and has been used as a framework for guidelines issued in the UK, Europe and the USA [4–6]. Patients with low- and some favourable intermediate-risk prostate cancer are generally offered active surveillance [4,7] (AS), while those with unfavourable intermediate-, and high-risk disease are considered for radical therapy [7]. Other classification systems, such as the CAPRA score [8], use additional clinical information, assigning simple numeric values based on age, pre-treatment PSA, GS, percentage of biopsy cores positive for cancer and clinical stage for an overall 0–10 CAPRA score. The CAPRA score has shown favourable prediction of PSA-free survival, development of metastasis and prostate cancer-specific survival [9].

Prostate cancer is often multifocal [10], with disease state often underestimated by TRUS biopsy alone [11] and overestimated by multiparametric MRI (mpMRI), most often in the case of Prostate Imaging Reporting and Data System (PI-RADS) 3 lesions [12]. Sampling issues associated with needle biopsy of the prostate have prompted the development of non-invasive urine tests for aggressive disease, which examine prostate-derived material, harvested within urine [13–15]. Recent successes in this field are illustrated by three studies carried out on whole urine for predicting the presence of GS ≥7 on initial biopsy: Tomlins et al. [13] and McKiernan et al. [14] used *PCA3* and *TMPRSS2-ERG* transcript expression levels, whilst Van Neste et al. [16] used *HOXC6* and *DLX1* in combination with traditional clinical markers. The objectives of the present study were to develop a urine classifier that can predict D'Amico and CAPRA risk group, and to additionally test its utility as a predictor of disease progression, triggering the requirement for therapeutic intervention, within an AS cohort with 5 years of clinical follow-up. As a starting point, we used 167 gene probes, many previously associated with prostate cancer progression, leading to the development of a 36-gene classifier, known as the Prostate Urine Risk (PUR) signatures.

Methods

Patient Samples and Clinical Criteria

The Movember cohort comprised first-catch post-DRE urine samples collected at diagnosis between 2009 and 2015 from urology clinics at the Norfolk and Norwich University Hospital (NNUH [Norwich, UK]), Royal Marsden Hospital (RMH [London, UK]), St James's Hospital (Dublin, Republic of Ireland) and from primary care and urology clinics of Emory Healthcare (Atlanta, GA, USA). Within the Movember cohort, 87 patients were enrolled on an AS programme at the RMH [7]. AS eligibility criteria for this programme included histologically proven prostate cancer, age 50–80 years, clinical stage T1/T2, PSA <15 ng/mL, GS ≤ 3 + 3 (GS ≤3 + 4 if age >65 years), and <50% percent positive biopsy cores. Progression was defined as the detection of disease by clinical criteria that typically trigger the requirement for therapeutic intervention. Clinical criteria of progression were either: PSA velocity >1 ng/mL per year or adverse histology on repeat biopsy, defined as primary GS ≥4 or ≥50% biopsy cores positive for cancer. mpMRI criteria for progression were either: detection of >1 cm^3 prostate tumour, an increase in volume >100% for lesions between 0.5 and 1 cm^3, or T3/4 disease [7].

D'Amico classification used GS and PSA criteria as per D'Amico et al. [2]. CAPRA classification used the criteria as described by Cooperberg et al. [8]. Sample collections and processing were ethically approved in their country of origin: NNUH samples by the East of England Research Ethics Committee, Dublin samples by St James's Hospital, RMH samples by the local ethics committee, and Emory Healthcare samples by the institutional review board of Emory University. TRUS-guided biopsy was used to provide biopsy information. Where multiple biopsies were taken the results from the closest biopsy to initial urine sample collection were used. Men were defined as having no evidence of cancer (NEC) with a PSA level normal for their age or lower [17] and, as such, were not subjected to biopsy. Metastatic disease was defined by a PSA >100 ng/mL and was excluded from analyses.

Sample Processing

For the full Movember protocol, see Data S1. Briefly, urine was centrifuged (1200*g* 10 min, 6 °C) within 30 min of collection to pellet cellular material. Supernatant extracellular vesicles (EVs) were then harvested by microfiltration as described by Miranda et al. [18] and RNA extracted (RNeasy micro kit, #74004; Qiagen, Hilden, Germany). RNA was amplified as cDNA with an Ovation PicoSL WTA system V2 (Nugen #3312-48). Then, 5–20 ng of total RNA was amplified where possible, down to 1 ng input in 10 samples. The mean (range) cDNA yields were 3.83 (1–6) μg.

Expression Analyses

NanoString expression analysis (167 probes, 164 genes; Data S2) of 100 ng cDNA was performed at the Human Dendritic Cell Laboratory, Newcastle University, UK. A total of 137 probes were selected based on previously proposed controls plus prostate cancer diagnostic and prognostic biomarkers within tissue and control probes (Data S2). Thirty additional probes were selected as overexpressed in prostate cancer samples when next-generation sequence data generated from 20 urine-derived EV-RNA samples were analysed (unpublished). Target gene sequences were provided to NanoString, who designed the probes according to their protocols [19]. Data were adjusted relative to internal positive

control probes as stated in NanoString's protocols. The ComBat algorithm was used to adjust for inter-batch and inter-cohort bias [20]. Data were adjusted by means of a correction factor for input amount by normalization to two invariant and highly expressed housekeeping gene-probes, *GAPDH* and *RPLP2*. The correction factor (CF) for a given sample *i*, was calculated as the total mean of *GAPDH* and *RPLP2* expression, divided by the sample-specific mean of *GAPDH* and *RPLP2*:

$$\mathrm{CF}_i = \frac{\sum_j \bar{x}_{\mathrm{GAPDH}_j,\mathrm{RPLP2}_j}}{n \times \bar{x}_{\mathrm{GAPDH}_i,\mathrm{RPLP2}_i}}$$

All data were expressed relative to *KLK2* as follows: samples with low *KLK2* (counts <100) were removed, and data $\log_2$ transformed. Data were further normalized by adjusting the median of each probe across all samples to 1, with the interquartile range (IQR) adjusted to that of *KLK2*. More formally, for each sample *i* and gene-probe *j*, the *KLK2* normalized value, $\hat{y}_{i,j}$ was calculated as:

$$\hat{y}_{ij} = \frac{\left(\left(\frac{y_{i,j} - \mathrm{median}_j}{\mathrm{IQR}_j}\right) \times \mathrm{IQR}_{\mathrm{KLK2}}\right) + \mathrm{Median}_{\mathrm{KLK2}}}{y_{i,\mathrm{KLK2}}}$$

No correlation was seen with respect to patient's drugs, cohort site, urine pH, colour or sample volume ($P > 0.05$, chi-squared and Spearman's rank tests; data not shown).

Model Production and Statistical Analysis

All statistical analyses and model construction were undertaken in R version 3.4.1 [21], and unless otherwise stated used base R and default parameters.

The PUR signatures were constructed from the training dataset as follows: for each probe, a univariate cumulative link model was fitted using the R package *clm*, with risk group as the outcome and NanoString expression as inputs. Each probe that had a significant association with risk group ($P < 0.05$) was used as input to the final multivariate model. A constrained continuation ratio model with an L_1 penalization was fitted to the training dataset using the *glmnetcr* library [22], an adaption of the LASSO method [23]. Default parameters were applied using the LASSO penalty and values from all probes selected by the univariate analysis used as input. The final multivariable model was selected according to the minimum Akaike information criterion and incorporated all probes not removed by the LASSO penalty. Ordinal logistic regression was undertaken using the *ordinal* library [24].

Bootstrap resampling of receiver-operating curve analyses used the *pROC* library [25] for calculation, statistical tests and production of figures, with 2000 resamples used. Random predictors were generated by randomly sampling from a uniform distribution between 0 and 1.

Survival analyses were undertaken where follow-up of patients on AS allowed, and used progression as an endpoint, as described above. Cox proportional hazards models used risk signatures as a continuous variable. Kaplan–Meier estimators were calculated based on the median optimal threshold to minimize the log-rank test *P* value from 10 000 resamples of the cohort, with replacement to ensure robustness. The costs of missing significant cancer are far higher than an unnecessary biopsy or investigation. With this considered, where multiple samples were analysed from the same patient on AS, the sample with the highest PUR-4 signature was used in survival analyses and Kaplan–Meier estimators. No multiple samples from patients on AS appeared simultaneously in either training or test datasets, minimizing the potential for overfitting and bias of the model.

Decision-curve analysis (DCA) [26] examined the potential net benefit of using PUR signatures in the clinic. Standardized net benefit was calculated with the *rmda* library [27] and presented throughout our DCAs, as it is more interpretable when compared with net benefit [28]. In order to ensure DCA was representative of a more general population, the prevalence of Gleason score within the Movember cohort was adjusted via bootstrap resampling to match that observed in a population of 219 439 men that were in the control arm of the Cluster Randomized Trial of PSA Testing for Prostate Cancer (CAP) [29]. For the biopsied men within this CAP cohort, 23.6% had GS 6, 8.7% GS 7 and 7.1% GS ≥8, with 60.6% of biopsies being prostate cancer-negative. This was used to perform stratified random sampling with replacement of the Movember cohort to produce a 'new' dataset of 300 samples. Standardized net benefit was calculated on the resampled dataset, and the process repeated for a total of 1000 resamples. The mean standardized net benefit for PUR-4 and the 'treat-all' options over all iterations were used to produce the presented figures to account for variance in sampling.

Results

Clinical Cohort

The Movember cohort comprised 535 post-DRE urine samples collected from four centres (NNUH, *n* = 312; RMH, *n* = 87; Atlanta, *n* = 85; Dublin, *n* = 17). Multiple, longitudinal samples within the Movember cohort were provided by 20 of the 87 men enrolled on an AS programme at the RMH. The median (IQR) time between collection of multiple samples was 185 (122–252) days and was treated independently from one another. Samples originated from men categorized as having either NEC (*n* = 92) or localized prostate cancer at time of urine collection, as detected by TRUS biopsy (*n* = 443), that were further subdivided into three risk categories using D'Amico criteria: low, *n* = 134;

intermediate, n = 208; and high, n = 101. Patients with metastatic cancer at collection were excluded from analyses. Further characteristics of the Movember cohort are available in Table 1.

Selection of Extracellular Vesicle Fractions and RNA Yields

Prostate markers *KLK2* and *KLK3*, were up to 28-fold higher in the EV fraction when compared to sediment (TaqMan RT-PCR, paired samples Welch t-test $P < 0.001$, data not shown). Based on these analyses and previously published results by Pellegrini et al. [30]. EVs were selected for further study.

Median urine-derived EV-RNA yields for the NNUH cohort were similar for patients in the NEC (204 ng), low- (180 ng) and intermediate-risk (221 ng) groups, and lower in the high-risk group (108 ng; Fig. S1). Yields from three patients post-radical prostatectomy were 0.8–2 ng, suggesting that most urine-derived EV-RNA originates from the prostate.

Development of the Prostate Urine Risk Signatures

Samples in the D'Amico low-, intermediate- and high-risk categories, together with NEC samples, were divided into the Movember Training dataset (two-thirds of samples; n = 358) and the Movember Test dataset (one-third of samples; n = 177) by random assignment, stratified by risk category (Table 1).

Table 1 Characteristics of the training and test cohorts

Characteristic	Training	Test
Total, *n* (%)	358 (67.0)	177 (33.0)
Collection centre		
NNUH	203	109
RMH	83	38
Dublin	9	8
Atlanta	63	22
PSA, ng/mL, mean (median; IQR)	10.6 (6.9, 6.4)	10.9 (6.9, 7)
Age, years, mean (median; IQR)	65.8 (67, 11)	67.2 (67, 11)
Family history of prostate cancer, %; no, yes, NA	3.0, 6.1, 90.8	0.6, 6.2, 93.3
First biopsy, *n* (%)	298 (82.78)	145 (81.46)
Prostate volume, mL; mean (median; IQR)	59.2 (49.8, 30.4)	61.1 (49.2, 32.8)
PSA density, ng/mL; mL, mean (median; IQR)	0.29 (0.19, 0.16)	0.29 (0.18, 0.17)
Suspicious DRE, *n*	107	52
Diagnosis, *n*	358	177
NEC, *n* (%)	62 (17.3)	30 (17.0)
D'Amico low-risk, *n* (%)	89 (24.9)	45 (25.4)
D'Amico intermediate-risk, *n* (%)	139 (38.8)	69 (39.0)
D'Amico high-risk, *n* (%)	61 (17.0)	27 (15.3)
Metastatic (bone scan), *n* (%)*	7 (2.0)	6 (3.3)
CAPRA, *n*	288	145
Low-risk (0–2), *n* (%)	97 (33.7)	49 (33.7)
Intermediate-risk (3–5), *n* (%)	108 (37.5)	53 (36.6)
High-risk (≥6), *n* (%)	83 (28.8)	43 (29.7)
GS, *n* (%)	292	144
6	119 (40.8)	64 (44.4)
7	131 (44.9)	56 (38.9)
≥ 8	42 (14.4)	24 (16.7)

GS, Gleason score; IQR, interquartile range; NEC, no evidence of cancer; NNUH, Norfolk and Norwich University Hospital; RMH, Royal Marsden Hospital.

The optimal model, as defined by the LASSO criteria in a constrained continuation ratio model (see Methods for full details) incorporated information from 36 probes (Table 2, for model coefficients see Table S1) and was applied to both training and test datasets (Fig. 1A,B). For each sample, the four-signature PUR model defined the probability of containing NEC (PUR-1), low-risk (PUR-2), intermediate-risk (PUR-3) and high-risk (PUR-4) material within samples (Fig. 1A,B). The sum of all four PUR signatures in any individual sample was 1 (PUR-1 + PUR-2 + PUR-3 + PUR-4 = 1). The strongest PUR signature for a sample was termed the primary (1°) signature, while the second highest was called the secondary (2°) signature (Fig. 1C,D).

Pre-biopsy Prediction of D'Amico Risk, CAPRA Score and Gleason Score

Primary PUR signatures (PUR-1–4) were found to be significantly associated with clinical category (NEC, low-risk, intermediate-risk, and high-risk, respectively) in both the training and test datasets ($P < 0.001$, Wald test for ordinal logistic regression in both the training and test datasets; Fig. 2A,B). A similar association was observed with CAPRA score ($P < 0.001$, Wald test for ordinal logistic regression in both the training and test datasets; Fig. S2).

Based on recommended guidelines [4–6], the distinction between D'Amico low- and intermediate-risk disease is considered critical because radical therapy is commonly recommended for patients with high- and intermediate-risk cancer. We therefore initially tested the ability of the PUR model to predict the presence of high- or intermediate-risk disease from low-risk or NEC on initial biopsy. Each of the four PUR signatures alone were able to predict the presence of significant disease (risk category ≥ intermediate, area under the curve [AUC] ≥ 0.68 for each PUR signature, test dataset; Fig. S3), and were significantly better than a random predictor ($P < 0.001$, bootstrap test, 2000 resamples); however, PUR-1 and PUR-4 were best at discerning significant disease and were equally effective; the AUCs for both PUR-4 and for PUR-1 in the training and test cohorts were 0.81 (95% CI 0.77–0.85) and 0.77 (95% CI 0.70–0.84; Fig. 2C,D), respectively.

When GS alone was considered we found that PUR-4 predicted GS ≥3 + 4 with AUCs of 0.78 (95% CI 0.73–0.82; training) and 0.76 (95% CI 0.69–0.83; test) and GS ≥4 + 3 with AUCs of 0.76 (95% CI 0.70–0.81; training) and 0.72 (95% CI 0.63–0.81; test; Fig. 3). The ability to predict GS ≥3 + 4 was particularly relevant because this was previously

Table 2 NanoString gene probes incorporated by LASSO regularization in the final optimal model used to produce the prostate urine risk signatures

Gene targets of NanoString probes in PUR model	
AMACR	***MEX3A***
AMH	*MEMO1*
ANKRD34B	*MME*
APOC1	*MMP11*
AR (exons 4–8)	*MMP26*
DPP4	*NKAIN1*
ERG (exons 4–5)	*PALM3*
GABARAPL2	*PCA3*
GAPDH	*PPFIA2*
GDF15	*SIM2 (short)*
HOXC6	*SMIM1*
HPN	*SSPO*
IGFBP3	*SULT1A1*
IMPDH2	*TDRD1*
ITGBL1	*TMPRSS2/ERG fusion*
KLK4	*TRPM4*
MARCH5	*TWIST1*
MED4	*UPK2*

PUR, prostate urine risk.

chosen as an endpoint for aggressive disease in other urine biomarker studies, where AUCs of 0.77, 0.78 and 0.74 were reported by McKiernan et al. [14], Tomlins et al. [13] and Van Neste et al. [16], respectively.

Decision-curve analysis [26] examined the potential net benefit of using PUR signatures in a non-PSA-screened population. Biopsy of men based on their PUR-4 score provided a net benefit over biopsy of men based on current clinical practice across all thresholds (Fig. 4). When DCA was also undertaken within the context of a PSA-screened population, PUR continued to provide a net benefit (Fig. S4).

Active Surveillance Cohort

Within the Movember cohort, 87 men were enrolled in AS at the RMH. The median (range) follow-up time from initial urine sample collection was 5.7 (5.1–7.0) years (Table S2). The median (range) time from initial urine sample collection to progression or final follow-up was 503 days (0.1–7.4 years). The PUR profiles from these men were used to investigate the prognostic utility of PUR beyond categorizing D'Amico risk. The PUR profiles were significantly different among the 23 men who progressed within 5 years of urine sample collection, and the 49 men who did not progress ($P < 0.001$, Wilcoxon rank sum test; Fig. 5A). Twenty-two men progressed by the criteria detailed above, with an additional nine men progressing based solely on mpMRI criteria. Further AS cohort characteristics are available in Table S2.

Calculation of Kaplan–Meier estimators with samples divided on the basis of 1°, 2° and 3° PUR-1 and PUR-4 signatures showed significant differences in clinical outcome ($P < 0.001$, log-rank test; Fig. 5B) and was robust (log-rank test $P < 0.05$ in 93.6% of 100 000 cohort resamples with replacement; see Methods for full details). Proportion of PUR-4, a continuous variable, had a significant association with clinical outcome ($P < 0.001$; IQR HR 5.87, 95% CI 1.68–20.46; Cox Proportional hazards model). A robust optimal threshold of PUR-4 was determined to dichotomize patients on AS (PUR-4 = 0.174). The two groups had a large difference in time to progression: 60% progression within 5 years of urine sample collection in the poor prognosis group compared to 10% in the good prognosis group ($P < 0.001$, log-rank test, HR 8.23; 95% CI 3.26–20.81; Fig. 5C). This result is robust ($P < 0.05$ in 99.8% of 100 000 cohort resamples with replacement; see Methods for full details).

When mpMRI criteria for progression were also included, both primary PUR status and dichotomized PUR threshold remained a significant predictors of progression ($P < 0.001$ log-rank test; Fig. S5). When the AS cohort was split by D'Amico risk category at initial urine collection, PUR-4 remained a significant predictor of progression in men with low-risk disease, but not in men with intermediate-risk disease ($P < 0.001$ log-rank test; Fig. S6).

Multiple urine specimens had been collected for 20 of the patients entered into the AS trial, allowing us to assess the stability of urine profiles over time (Fig. S7). In patients who had not progressed, samples were found to be stable compared to a null model generated by randomly selected samples from the whole Movember cohort ($P = 0.011$; bootstrap analysis with 100 000 iterations). Samples from patients deemed to have progressed failed this stability test ($P = 0.059$).

Discussion

The variation in clinical outcome for prostate cancer, even within risk-stratified groups such as D'Amico, is well established. Many attempts have been made to address this problem, including the subcategorization of intermediate-risk disease into favourable and unfavourable groups [31] and the development of the CAPRA classification system [8]. Other approaches include the development of an unsupervised classification framework [32] and of biomarkers of aggressive disease, as illustrated by Cuzick et al. [33], Knezevic et al. [34] and Robert et al. [35]. In each of the examples given above, analyses are performed on cancerous tissue, usually taken at the time of diagnosis via needle biopsy.

Urine biomarkers offer the prospect of a more holistic assessment of cancer status prior to invasive tissue biopsy and may also be used to supplement standard clinical stratification. Previous urine biomarker models have been designed specifically for single purposes, such as the detection of prostate cancer on re-biopsy (*PCA3* test), or to detect GS ≥3 + 4 [13,14,16,36]. In the present study, we constructed the

Fig. 1 (**A**) Prostate urine risk (PUR) profiles (PUR-1, green; PUR-2 , blue; PUR-3, yellow; PUR-4, red) for the training cohort, grouped by D'Amico risk group and ordered by ascending PUR-4 score. Horizontal lines indicate where the PUR thresholds lie for: 1° PUR-1 (green, 0.342), 2° PUR-1 (purple, 0.297), 1° PUR-4 (red, 0.476), 2° PUR-4 (orange, 0.219) and the crossover point between PUR-1 and PUR-4 (black, 0.123 both PUR-1 and 4). (**B**) PUR profiles in the test cohort. (**C**) Examples of samples with primary PUR signatures, where coloured circles indicate the primary PUR signal for that sample: 1° PUR-1 (green), 1° PUR-2 (blue), 1° PUR-3 (yellow), 2° PUR-4 (orange) and 1° PUR-4 (red). The sum of all four PUR signatures in any individual sample is 1, i.e. PUR-1 + PUR-2 + PUR-3 + PUR-4 = 1. (**D**) The outline of the four PUR signatures for all samples, ordered in ascending PUR-4 (red) to illustrate where 1°, 2° and the 3° crossover point of PUR-1 and PUR-4 lies.

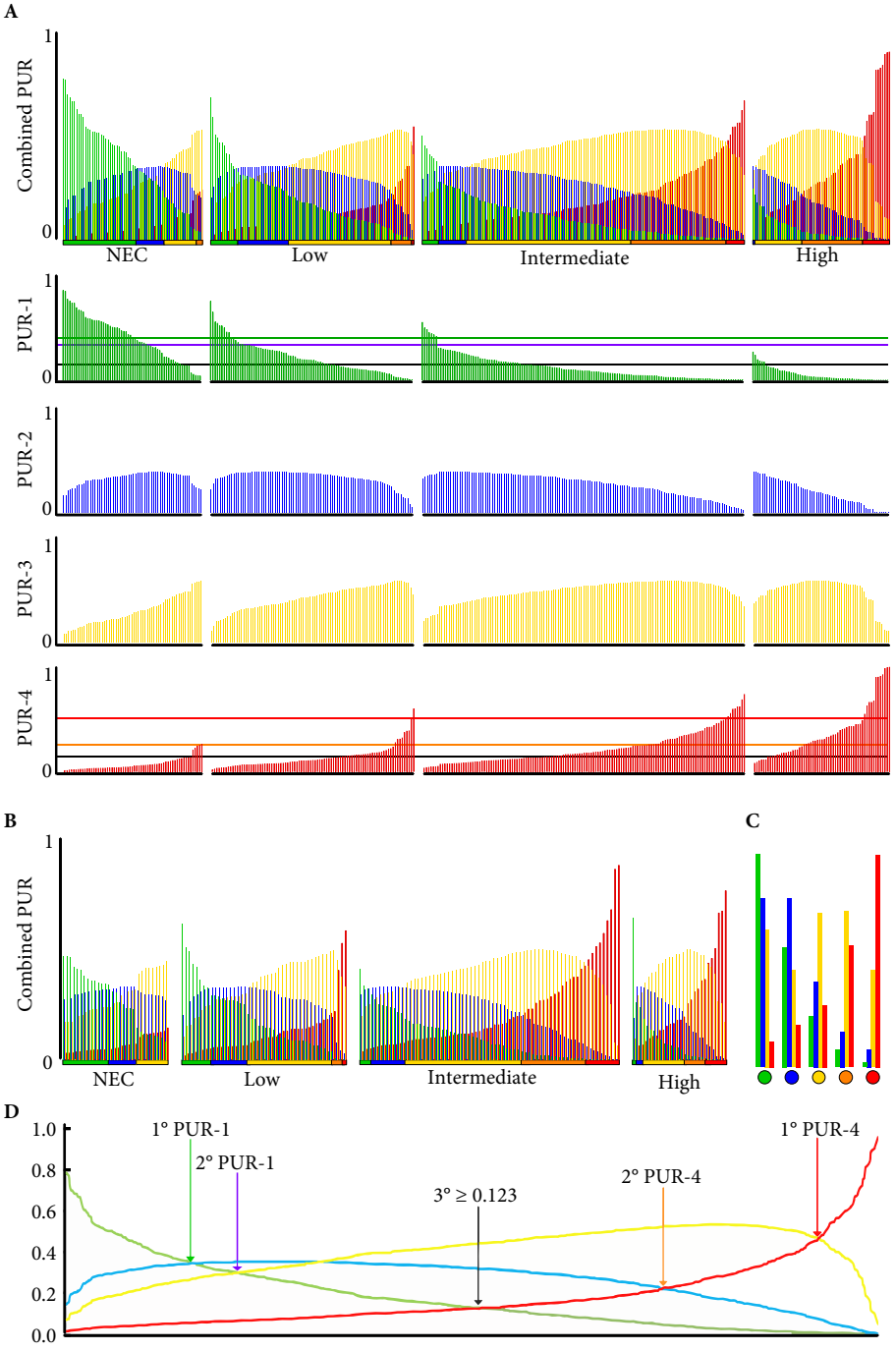
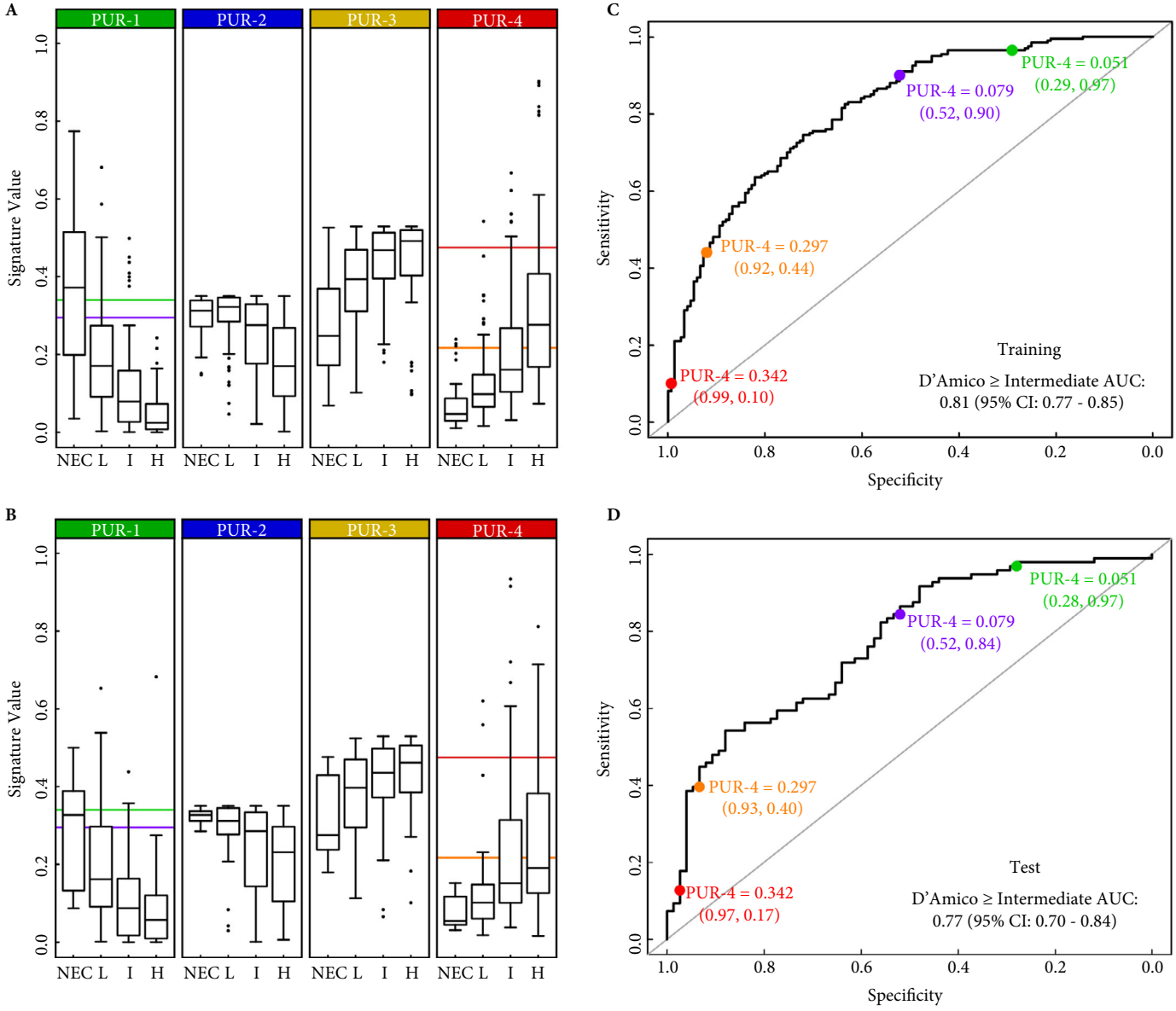


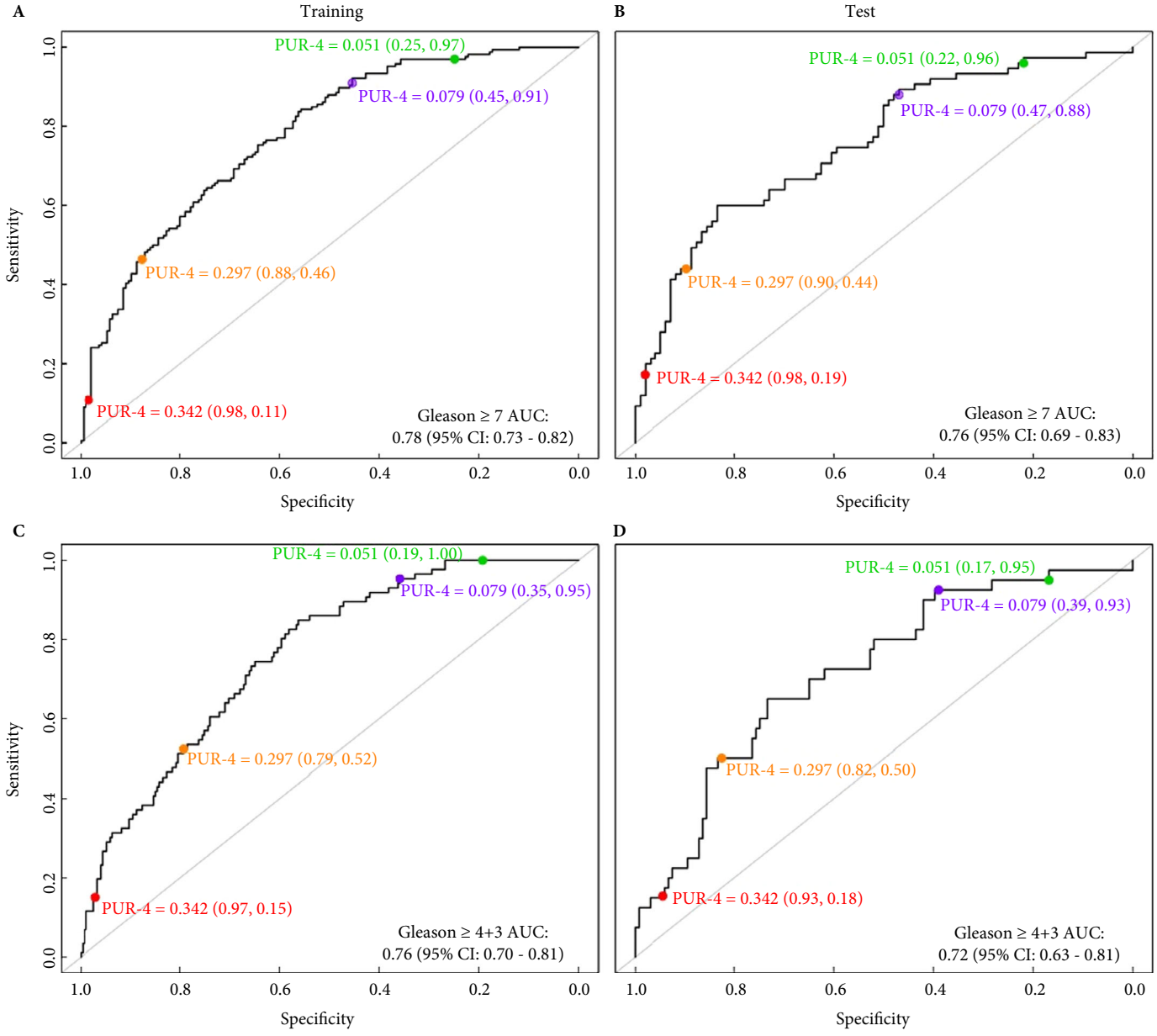
Fig. 2 (**A** and **B**) Boxplots of prostate urine risk (PUR) signatures in samples categorized as no evidence of cancer (NEC [*n* = 62, training; *n* = 30, test]) and D'Amico risk categories; (L – low-: *n* = 89, training, *n* = 45, test; I – intermediate-: *n* = 131, training, *n* = 69, test; and H – high-risk, *n* = 61, training, *n* = 27, test) in (**A**) the training and **B**) the test cohort. Horizontal lines indicate where the PUR thresholds lie for: 1° PUR-1 (green), 2° PUR-1 (purple), 1° PUR-4 (red), 2° PUR-4 (orange). (**C** and **D**) Receiver-operating characteristic (ROC) curves of PUR-4 and PUR-1 predicting the presence of significant (D'Amico intermediate- or high-risk) prostate cancer prior to initial biopsy in (**C**) the training and (**D**) the test cohort. Coloured circles indicate the specificity and sensitivity, respectively, of thresholds along the ROC curve that correspond to the indicated PUR-4 thresholds, equivalent to: red – 1° PUR-4, orange – 2° PUR-4, purple – equivalent to 2° PUR-1, green – equivalent to 1° PUR-1.



four PUR signatures to provide a non-invasive and simultaneous assessment of non-cancerous tissue and D'Amico low-, intermediate- and high-risk prostate cancer in individual prostates. The use of individual signatures for the three D'Amico risk types is unique and could significantly aid the deconvolution of complex cancerous states into more readily identifiable forms for monitoring the development of high-risk disease in, for example, men on AS.

For the detection of significant prostate cancer, PUR compares favourably to other published biomarkers which have used simpler transcript expression systems involving low numbers of probes [13,14,16,36]. In the present study, we show that the PUR classifier, based on the RNA expression levels of 36 gene probes, can be used as a versatile predictor of cancer aggression. Notably *PCA3*, *TMPRSS2-ERG* and *HOXC6* were all included within the optimal PUR model

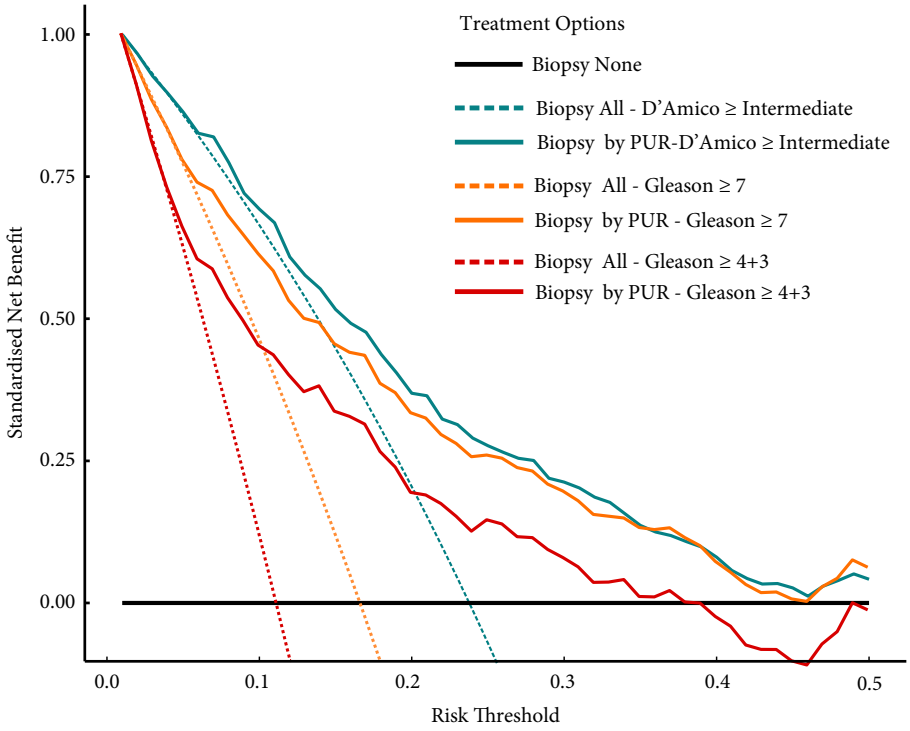
Fig. 3 Area under the curve (AUC) for prostate urine risk (PUR)-4 predicting the presence/absence of Gleason score (GS) ≥7 on initial biopsy in the training and test cohorts (**A** and **B**, respectively) and GS ≥4 + 3 in the training and test cohorts (**C** and **D**, respectively). Coloured circles indicate the specificity and sensitivity, respectively, of thresholds along the receiver-operating characteristic curve that correspond to the indicated PUR-4 thresholds, equivalent to: red – 1° PUR-4, orange – 2° PUR-4, purple – equivalent to 2° PUR-1, green – equivalent to 1° PUR-1.



defined by the LASSO criteria, while DLX1 was not. We first showed that the ability of PUR-4 status to predict TRUS-detected GS ≥3 + 4 was similar (AUC 0.76; 95% CI 0.69–0.83, test dataset) to these published models using *PCA3*/*TMPRSS2-ERG* (AUC 0.74–0.78) [13,14] and *HOXC6*/*DLX1* (AUC 0.77) [16].

Current clinical practice assesses patient's disease using PSA, needle biopsy of the prostate and mpMRI; however, up to 75% of men with a raised PSA level (≥3 ng/mL) are negative for prostate cancer on biopsy [6,37], whilst in absence of a raised PSA level, 15% of men are found to have prostate cancer, with a further 15% of these cancers being high-grade [38]. This illustrates the considerable need for additional biomarkers that can make pre-biopsy assessment of prostate cancer more accurate. In this respect, we show that both PUR-4 and PUR-1 are each equally good at predicting the

Fig. 4 Decision-curve analysis (DCA) plot depicting the standardized net benefit of adopting prostate urine risk (PUR)-4 as a continuous predictor for detecting significant cancer on initial biopsy, when significant is defined as: D'Amico risk group of intermediate or greater (teal), Gleason score (GS) ≥3 + 4 (orange) or GS ≥4 + 3 (red). To assess benefit in the context of cancer arising in a non-PSA screened population of men we used data from the control arm of the Cluster Randomized Trial of PSA Testing for Prostate Cancer (CAP) study [29]. Bootstrap analysis with 100 000 resamples was used to adjust the distribution of Gleason grades in the Movember cohort to match that of the CAP population. For full details see Methods.



presence of intermediate- or high-risk prostate cancer as defined by D'Amico criteria or by CAPRA status, while in DCA analysis we found that PUR provided a net benefit in both a PSA-screened and non-PSA-screened population of men. With the increased adoption of mpMRI, it would be useful in future studies to correlate PUR and other urine-based markers with MRI findings and radical prostatectomy outcomes.

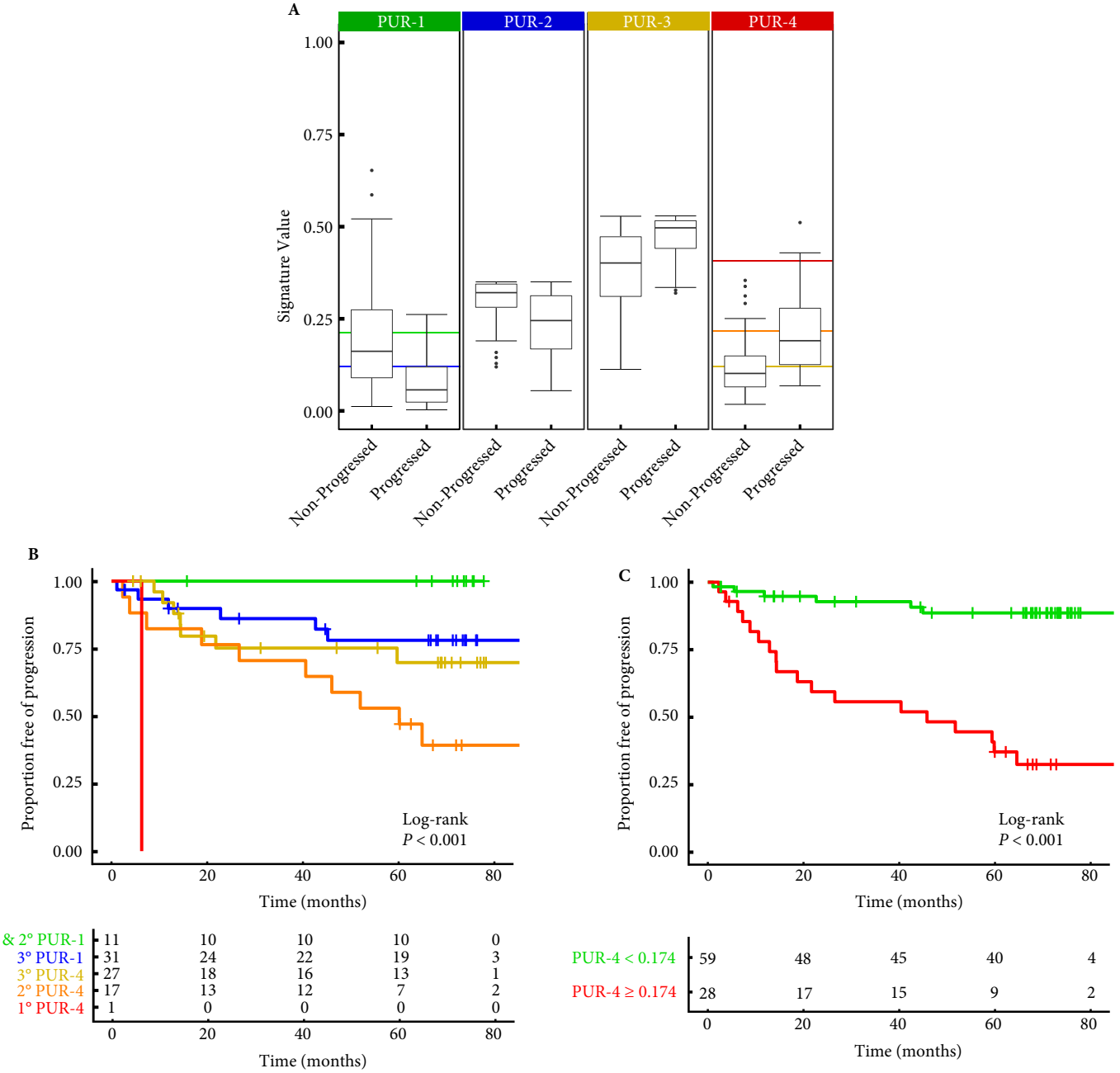
Variation in clinical outcomes is also well recognized for patients entered onto AS [39]. We found that the PUR framework worked well when applied to patients on AS monitored by PSA and biopsy, and also in patients monitored by mpMRI. A potential limitation of the present study is that we were not able to test the PUR stratification in an independent and more conservatively managed AS cohort; however, based on our observations, ~13% of the RMH AS cohort could have been safely removed from AS monitoring for a minimum of 5 years. An interesting feature is that, in some patients, the PUR signature predicted disease progression up to 5 years before it was detected by standard clinical methods. This prognostic information could potentially also aid the reduction of patient-elected radical intervention in men on AS, which in some cohorts can be as high as 75% within 3 years of enrolment [39]. Indeed, we would view the use of PUR within the context of AS as its major potential clinical application. Repeated longitudinal measurements of PUR status could help correctly assess and track a patient's risk over time in a non-invasive manner. A future priority is to further validate the utility of PUR within AS using other previously described longitudinal cohorts.

In conclusion, we have shown that PUR represents a new and versatile urine biomarker system capable of detecting aggressive prostate cancer and predicting the need for therapeutic intervention in men enrolled on AS. The dramatic differences in RNA expression profiles across the spectrum from high-risk cancer to patients with NEC, confirmed in a test cohort, can leave no doubt that the presence of cancer is substantially influencing the RNA transcripts found in urine EVs. We also provide evidence that the majority of post-DRE urine-derived EVs are derived from the prostate and that urine signatures are longitudinally stable.

Acknowledgements

This study was possible thanks to the Movember Foundation GAP1 Urine Biomarker project, The Masonic Charitable Foundation, The Bob Champion Cancer Trust, the King family, The Andy Ripley Memorial Fund and the Stephen Hargrave Trust.

Fig. 5 (**A**) Prostate urine risk (PUR) profiles of patients on active surveillance (AS) who met the clinical criteria, not including multiparametric MRI criteria, for progression (n = 23) or not (n = 49) 5 years after urine sample collection. Progression criteria were either: PSA velocity >1 ng/mL per year or Gleason score ≥4 + 3 or ≥50% cores positive for cancer on repeat biopsy. PUR signatures for progressed vs non-progressed samples were significantly different for all PUR signatures (P < 0.001, Wilcoxon rank-sum test). Horizontal line colour indicates the thresholds for PUR categories described in: (**B**) Kaplan–Meier plot of progression in AS patients with respect to PUR categories described by the corresponding colours (green – 1° and 2° PUR-1, blue – 3° PUR-1, yellow – 3° PUR-4, orange – 2° PUR-4, red – 1° PUR-4), and the number of patients within each PUR category at the given time intervals in months from urine collection. (**C**) Kaplan–Meier plot of progression with respect to the dichotomized PUR thresholds described by the corresponding colours (green – PUR-4 < 0.174, red – PUR-4 ≥0.174) and the number of patients within each group at the given time intervals in months from urine collection.



Conflict of Interest

A patent application has been filed by the authors for the present work. There are no other conflicts of interest to disclose.

References

1 **D'Amico AV, Moul J, Carroll PR, Sun L, Lubeck D, Chen MH.** Cancer-specific mortality after surgery or radiation for patients with clinically localized prostate cancer managed during the prostate-specific antigen era. *J Clin Oncol* 2003; 21: 2163–72

2 **D'Amico AV, Whittington R, Bruce Malkowicz S et al.** Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *J Am Med Assoc* 1998; 280: 969–74

3 **Gleason DFMG.** Prediction of prognosis for prostatic staging, adenocarcinoma by combined histological grading and clinical. *J Urol* 1974; 111: 58–64

4 **Sanda MG, Cadeddu JA, Kirkby E et al.** Clinically localized prostate cancer: AUA/ASTRO/SUO guideline. Part I: risk stratification, shared decision making, and care options. *J Urol* 2018; 199: 683–90

5 **Mottet N, Bellmunt J, Bolla M et al.** EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. *Eur Urol* 2017; 71: 618–29

6 **National Institute for Health and Care Excellence**. *Prostate Cancer: Diagnosis and Treatment*. London: National Institute for Health and Care Excellence, 2014

7 **Selvadurai ED, Singhera M, Thomas K et al.** Medium-term outcomes of active surveillance for localised prostate cancer. *Eur Urol* 2013; 64: 981–7

8 **Cooperberg MR, Freedland SJ, Pasta DJ et al.** Multiinstitutional validation of the UCSF cancer of the prostate risk assessment for prediction of recurrence after radical prostatectomy. *Cancer* 2006; 107: 2384–91

9 **Brajtbord JS, Leapman MS, Cooperberg MR.** The CAPRA score at 10 years: contemporary perspectives and analysis of supporting studies. *Eur Urol* 2017; 71: 705–9

10 **Andreoiu M, Cheng L.** Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. *Hum Pathol* 2010; 41: 781–93

11 **Corcoran NM, Hovens CM, Hong MKH et al.** Underestimation of Gleason score at prostate biopsy reflects sampling error in lower volume tumours. *BJU Int* 2012; 109: 660–4

12 **Ahmed HU, El-Shater Bosaily A, Brown LC et al.** Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet* 2017; 389: 815–22

13 **Tomlins SA, Day JR, Lonigro RJ et al.** Urine TMPRSS2:ERG Plus PCA3 for individualized prostate cancer risk assessment. *Eur Urol* 2016; 70: 45–53

14 **McKiernan J, Donovan MJ, O'Neill V et al.** A novel urine exosome gene expression assay to predict high-grade prostate cancer at initial biopsy. *JAMA Oncol* 2016; 2: 882–9

15 **Donovan MJ, Noerholm M, Bentink S et al.** A molecular signature of PCA3 and ERG exosomal RNA from non-DRE urine is predictive of initial prostate biopsy result. *Prostate Cancer Prostatic Dis* 2015; 18: 370–5

16 **Van Neste L, Hendriks RJ, Dijkstra S et al.** Detection of high-grade prostate cancer using a urinary molecular biomarker-based risk score. *Eur Urol* 2016; 70: 740–8

17 **Deantoni EP, Crawford ED, Oesterling JE et al.** Age- and race-specific reference ranges for prostate-specific antigen from a large community-based study. *Urology* 1996; 48: 234–9

18 **Miranda KC, Bond DT, McKee M et al.** Nucleic acids within urinary exosomes/microvesicles are potential biomarkers for renal disease. *Kidney Int* 2010; 78: 191–9

19 **Geiss GK, Bumgarner RE, Birditt B et al.** Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* 2008; 26: 317–25

20 **Johnson WE, Li C, Rabinovic A.** Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; 8: 118–27

21 **R Core Team**. *R: A Language and Environment for Statistical Computing. R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2018

22 **Archer KJ, Williams AAA.** L1 penalized continuation ratio models for ordinal response prediction using high-dimensional datasets. *Stat Med* 2012; 31: 1464–74

23 **Tibshirani R.** Regression shrinkage and selection via the lasso. *J Roy Statist Soc Ser B (Methodol)* 1996; 58: 267–88

24 **Christensen RHB.** Ordinal regression models for ordinal data, 2018

25 **Robin X, Turck N, Hainard A et al.** pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12: 77

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

27 **Brown M.** rmda: Risk Model Decision Analysis, 2018

28 **Kerr KF, Brown MD, Zhu K, Janes H.** Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol* 2016; 34: 2534–40

29 **Martin RM, Donovan JL, Turner EL et al.** Effect of a low-intensity PSA-based screening intervention on prostate cancer mortality. *JAMA* 2018; 319: 883

30 **Pellegrini KL, Patil D, Douglas KJS et al.** Detection of prostate cancer-specific transcripts in extracellular vesicles isolated from post-DRE urine. *Prostate* 2017; 77: 990–9

31 **Aghazadeh MA, Frankel J, Belanger M et al.** National Comprehensive Cancer Network® favorable intermediate risk prostate cancer—is active surveillance appropriate? *J Urol* 2018; 199: 1196–201

32 **Luca BA, Brewer DS, Edwards DR et al.** DESNT: a poor prognosis category of human prostate cancer. *Eur Urol Focus* 2018; 4: 842–50

33 **Cuzick J, Berney DM, Fisher G et al.** Prognostic value of a cell cycle progression signature for prostate cancer death in a conservatively managed needle biopsy cohort. *Br J Cancer* 2012; 106: 1095–9

34 **Knezevic D, Goddard AD, Natraj N et al.** Analytical validation of the Onco type DX prostate cancer assay – a clinical RT-PCR assay optimized for prostate needle biopsies. *BMC Genom* 2013; 14: 690

35 **Robert G, Jannink S, Smit F et al.** Rational basis for the combination of PCA3 and TMPRSS2:ERG gene fusion for prostate cancer diagnosis. *Prostate* 2013; 73: 113–20

36 **Hessels D, Klein Gunnewiek JMT, Van Oort I et al.** DD3PCA3-based molecular urine analysis for the diagnosis of prostate cancer. *Eur Urol* 2003; 44: 8–16

37 **Lane JA, Donovan JL, Davis M et al.** Active monitoring, radical prostatectomy, or radiotherapy for localised prostate cancer: study design and diagnostic and baseline results of the ProtecT randomised phase 3 trial. *Lancet Oncol* 2014; 15: 1109–18

38 **Thompson IM, Pauler DK, Goodman PJ et al.** Prevalence of prostate cancer among men with a prostate-specific antigen level ≤4.0 ng per milliliter. *N Engl J Med* 2004; 350: 2239–46

39 **Simpkin AJ, Tilling K, Martin RM et al.** Systematic review and meta-analysis of factors determining change to radical treatment in active surveillance for localized prostate cancer. *Eur Urol* 2015; 67: 993–1005

Appendix

The Movember GAP1 Urine Biomarker Consortium: Bharati Bapat, Rob Bristow, Andreas Doll, Jeremy Clark, Colin Cooper, Hing Leung, Ian Mills, David Neal, Mireia Olivan, Hardev Pandha, Antoinette Perry, Chris Parker, Martin Sanda, Jack Schalken, Hayley Whitaker.

Correspondence: Shea P. Connell, Norwich Medical School, University of East Anglia, Norwich NR4 7UQ, UK.

e-mail: s.connell@uea.ac.uk

Abbreviations: EV, extracellular vesicle; AS, active surveillance; PUR, Prostate Urine Risk; AUC, area under the curve; GS, Gleason score; mpMRI, multiparametric MRI; NNUH, Norfolk and Norwich University Hospital; RMH, Royal Marsden Hospital; NEC, no evidence of cancer; DCA, decision-curve analysis; CAP, Cluster Randomized Trial of PSA Testing for Prostate Cancer.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Extracellular vesicle RNA yields from samples of different clinical categories collected at the NNUH.
Fig. S2. Boxplots of prostate-urine-risk signatures relative to no evidence of cancer and CAPRA scores 1–10 in the Training and Test cohorts.
Fig. S3. Area under the curves for each of the four prostate-urine-risk (PUR) signatures (green – PUR-1, blue – PUR-2, yellow – PUR-3, red – PUR-4) predicting presence of D'Amico intermediate- or high-risk cancers on initial biopsy in both training and test cohorts.
Fig. S4. Decision curve analysis plot depicting the standardised net benefit of adopting PUR-4 as a continuous predictor for detecting significant cancer on initial biopsy, when significant is defined as: D'Amico risk group of Intermediate or greater (teal), Gleason ≥7 (orange) or Gleason ≥4 + 3 (red).
Fig. S5. (**A**) Kaplan-Meier plot of active surveillance progression, including mpMRI criteria over time in days with respect to prostate-urine-risk (PUR) thresholds described by the corresponding colours green – 1° and 2° PUR-1, blue – 3° PUR-1, yellow – 3° PUR-4, orange – 2° PUR-4, red – 1° PUR-4. Table underneath details the number of patients still at risk of progression within each group. (**B**) Kaplan-Meier plot of progression, including mpMRI criteria, with respect to the dichotomised PUR thresholds described by the corresponding colours green – PUR-4 <0.174, red – PUR-4 ≥0.174 and the number of patients within each group at the given time intervals in months from urine collection.
Fig. S6. Kaplan-Meier plots and risk tables of active surveillance (AS) progression with optimised PUR-4 thresholds when the AS cohort is split based on D'Amico risk categories.
Fig. S7. Prostate-urine-risk (PUR) signatures from Active Surveillance longitudinal samples: 1° PUR-1 (green), 2° PUR-1 (purple), 1° PUR-2 (blue), 1° PUR-3 (yellow), 2° PUR-4 (orange), 1° PUR-4 (red).
Table S1. Coefficients of the 36 gene probes included as variables in the final PUR model and the intercepts.
Table S2. Active surveillance cohort characteristics.

Data S1. Urine collection and processing – the Movember protocol.
Data S2. EV RNA harvest and extraction.