A Four-Group Urine Risk Classifier for Predicting Outcome in Prostate Cancer Patients

Shea P. Connell[#1], Marcel Hanna[#1], Frank McCarthy[2#], Rachel Hurst[1], Martyn Webb[1], Helen Curley[1], Helen Walker[3], Rob Mills[3], Richard Y. Ball[3], Martin G. Sanda[4], Kathryn L. Pellegrini[4], Dattatraya Patil[4], Antoinette S. Perry[5], Jack Schalken[6], Hardev Pandha[7], Hayley Whitaker[8], Nening Dennis[2], Christine Stuttle[2], Ian G. Mills[9, 10,11], Ingrid Guldvik[10], Movember GAP1 Urine Biomarker Consortium, Chris Parker[12], Daniel S. Brewer*[1,13], Colin S. Cooper*[1], Jeremy Clark*[1]

[#] Connell, Hanna & McCarthy are joint first authors

* Brewer, Clark & Cooper are joint senior authors.

[1] Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, UK.

[2] The Institute of Cancer Research, Sutton, Surrey, UK

[3] Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, Norfolk, UK

[4] Department of Urology, Winship Cancer Institute, Emory University School of Medicine, Atlanta, Georgia, USA.

[5] Cancer Biology and Therapeutics Laboratory, School of Biology and Environmental Science, Conway Institute, University College Dublin, Dublin 4, Ireland.

[6] Nijmegen Medical Centre, Radboud University Medical Centre, Nijmegen, The Netherlands.

[7] Faculty of Health and Medical Sciences, The University of Surrey, Guildford, UK

[8] Molecular Diagnostics and Therapeutics Group, University College London, Gower Street, London, UK.

[9] School of Medicine, Dentistry and Biomedical Sciences, Institute for Health Sciences, Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, UK.

[10] Centre for Molecular Medicine, University of Oslo, Oslo, Norway.

[11] Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK

[12] The Royal Marsden Hospital, Sutton, Surrey, UK;

[13] The Earlham Institute, Norwich Research Park, Norwich, Norfolk, UK.

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,

Corresponding Author:

Shea P. Connell,

Norwich Medical School,

University of East Anglia,

Norwich,

NR4 7UQ,

UK.

s.connell@uea.ac.uk

+44 (0) 7587 103339

Abstract

Objectives:

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

Patients and Methods:

Post-digital rectal examination UEV-RNA expression profiles from urine (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) PCa. 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, AUC = 0.77 (95% 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 (p<0.001; IQR HR = 2.86, 95% CI:1.83–4.47). PUR-4, when utilised continuously, dichotomised patient groups with differential progression rates of 10% and 60% five years post-urine collection (p<0.001, HR = 8.23, 95% CI:3.26–20.81).

Conclusion:

UEV-RNA can provide diagnostic information of aggressive PCa prior to biopsy, and prognostic information for men on AS. PUR represents a new & versatile biomarker that could result in substantial alterations to current treatment of PCa patients.

Keywords:

Prostate Cancer; Biomarker; Urine; Machine Learning; Active Surveillance; Liquid Biopsy

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 localised disease. D'Amico stratification(2), which classifies patients as 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 USA(4–6). Low-, and some favourable Intermediate-risk, patients are generally offered active surveillance(4,7) (AS) while unfavourable Intermediate-, and High-risk patients are considered for radical therapy(7). Other classification systems, such as CAPRA score(8), use additional clinical information, assigning simple numeric values based on age, pre-treatment PSA, Gleason score, 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 (MP-MRI), 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 $\geq$ 7 on initial biopsy: Tomlins *et al*. (2016), and McKiernan *et al*. (2016) used *PCA3* and *TMPRSS2-ERG* transcript expression levels, whilst Van Neste *et al.* (2016) used *HOXC6* and *DLX1* in combination with traditional clinical markers(14,16,17). The objectives of the current study were to develop a urine classifier that can predict D'Amico & CAPRA risk group, and additionally test its utility as a predictor of disease progression, triggering the requirement for therapeutic

intervention, within an AS cohort with five 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, deemed Prostate Urine Risk (PUR).

Methods:

Patient samples and clinical criteria:

The Movember cohort comprised of first-catch post-digital rectal examination (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, 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, clinical stage T1/T2, PSA < 15 ng/mL, Gs ≤ 3+3 (Gs ≤ 3+4 if age > 65), and < 50% percent positive biopsy cores. Progression was defined as the detection of disease by clinical criteria that typically triggers 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. MP-MRI criteria for progression were either: detection of >1 cm^3 prostate tumour, an increase in volume >100% for lesions between 0.5-1 cm^3, or T3/4 disease(7).

D'Amico classification used Gleason and PSA criteria as per D'Amico *et al*. (1998)(2). CAPRA classification used the criteria as described by Cooperberg *et al.*(2006)(8). Sample collections and processing were ethically approved in their country of origin: NNUH samples by the East of England REC, Dublin samples by St. James's Hospital. iii) RMH by the local ethics committee, iv) Emory Healthcare samples by the Institutional Review board of Emory University. Trans-rectal ultrasound (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 to have no evidence of cancer (NEC) with a PSA normal for their age or lower(18) and as such, were not

subjected to biopsy. Metastatic disease was defined by a PSA >100 ng/mL and were excluded from analyses.

Sample processing:

For the full Movember protocol see Supplementary Methods. 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 Miranda *et al.* (2010)(19) and RNA extracted (RNeasy micro kit, #74004, Qiagen). RNA was amplified as cDNA with an Ovation PicoSL WTA system V2 (Nugen #3312-48). 5-20 ng of total RNA was amplified where possible, down to 1 ng input in 10 samples. cDNA yields were mean 3.83 µg (1-6 µg).

Expression analyses:

NanoString expression analysis (167 probes, 164 genes, Supplementary Data) of 100 ng cDNA was performed at the Human Dendritic Cell Laboratory, Newcastle University, UK. 137 probes were selected based on previously proposed controls plus prostate cancer diagnostic and prognostic biomarkers within tissue and control probes (Supplementary Data). 30 additional probes were selected as overexpressed in prostate cancer samples when next generation sequence data generated from 20 urine derived EV RNA (UEV-RNA) samples were analysed (unpublished). Target gene sequences were provided to NanoString, who designed the probes according to their protocols(20). 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(21). Data were adjusted by means of a correction factor (CF) for input amount by normalisation to two invariant and highly expressed housekeeping gene-probes, *GAPDH* and *RPLP2*. The CF for a given sample *i*, was calculated as the mean sum of average *GAPDH* and *RPLP2* expression, divided by the sample-specific mean of *GAPDH* and *RPLP2*:

$$CF_i = \frac{\sum_j \bar{x}_{GAPDH_j, RPLP2_j}}{n \times \bar{x}_{GAPDH_i, RPLP2_i}}$$

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 normalised by adjusting the median of each

probe across all samples to 1, with the interquartile range adjusted to that of *KLK2*. More formally, for each sample *i* and gene-probe *j*, the *KLK2* normalised value, $\hat{y}_{i,j}$ was calculated as:

$$\hat{y}_{i,j} = \frac{\left(\left(\frac{y_{i,j} - median_j}{IQR_j}\right) \times IQR_{KLK2}\right) + Median_{KLK2}}{y_{i,KLK2}}$$

No correlation was seen with respect to patient's drugs, cohort site, urine pH, colour or sample volume ($p > 0.05$; Chi-square 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(22), and unless otherwise stated utilised base R and default parameters.

The Prostate Urine Risk (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 penalisation was fitted to the training dataset using the *glmnetcr* library(23), an adaption of the LASSO method(24). Default parameters were applied using the LASSO penalty and values from all probes selected by the univariate analysis used as input. The model with the minimum Akaike information criterion was selected. Ordinal logistic regression was undertaken using the *ordinal* library(25).

Bootstrap resampling of ROC analyses used the *pROC* library(26) for calculation, statistical tests and production of figures, with 2,000 resamples used. Random predictors were generated by randomly sampling from a uniform distribution between 0 and 1.

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 AS patient, the sample with the highest PUR-4 signature was used in survival analyses and Kaplan-Meier (KM) plots. No multiple samples from AS patients appeared simultaneously in both training and test datasets, minimising overfitting and bias of the model.

Decision curve analysis (DCA)(27) examined the potential net benefit of using PUR-signatures in the clinic. Standardised net benefit was calculated with the *rmda* library(28) and presented throughout our decision curve analyses, as it is more interpretable when compared to net benefit(29). In order to ensure DCA was representative of a more general population, the prevalence of Gleason grades within the Movember cohort were adjusted via bootstrap resampling to match that observed in a population of 219,439 men that were in the control arm of the Cluster Randomised Trial of PSA Testing for Prostate Cancer (CAP) Trial(30). For the biopsied men within this CAP cohort, 23.6% were Gs 6, 8.7% Gs 7 and 7.1% Gs 8 or greater, with 60.6% of biopsies being PCa negative. This was used to perform stratified random sampling with replacement of the Movember cohort to produce a "new" dataset of 300 samples. Standardised net benefit was calculated on the resampled dataset, and the process repeated for a total of 1,000 resamples. The mean standardised 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:

The Clinical Cohort

The Movember cohort comprised of 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 program at the RMH. The median time between collection of multiple samples was 185 days (IQR: 122-252 days) and were treated independently from one another. Samples originated from men categorised as having either No Evidence of Cancer (NEC, n = 92) or localised 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 (L), n = 134; Intermediate (I), n = 208; and High-risk (H), 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 EV 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*.(31), EVs were selected for further study. Median UEV-RNA yields for the NNUH cohort were similar for NEC (204 ng), Low- (180 ng) and Intermediate-risk (221 ng) patients, and lower in High-risk (108 ng) (Supplementary Figure 1). Yields from three patients post-radical prostatectomy were 0.8-2 ng, suggesting that most UEV-RNA originates from the prostate.

Development of the Prostate Urine Risk Signatures

Samples in D'Amico categories Low, Intermediate and High-risk, 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. Age, Stage, PSA, and Gleason scores were not significantly different across the two sets ($p > 0.05$; Wilcoxon rank sum test/Fisher's Exact Test; Table 1).

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 Supplementary Table 1) and was applied to both training and test datasets (Figure 1A, B). For each sample the 4-signature PUR-model defined the probability of containing NEC (PUR-1), L (PUR-2), I (PUR-3) and H (PUR-4) material within samples (Figure 1A, B). The sum of all four PUR-signatures in any individual sample was 1 (PUR1 + PUR2 + PUR3 + PUR4 = 1). The strongest PUR-signature for a sample was termed the primary (1°) signature while the second highest was called the secondary (2°) signature (Figure 1C, D).

Pre-biopsy Prediction of D'Amico risk, CAPRA score and Gleason:

Primary PUR-signatures (PUR-1 to 4) were found to significantly associate with clinical category (NEC, L, I, H respectively) in both training and test sets ($p < 0.001$, Wald test for ordinal logistic regression in both Training and Test datasets, Figure 2A, B). A similar association was observed with

CAPRA score ($p < 0.001$, Wald test for ordinal logistic regression in both Training and Test datasets; Supplementary Figure 2).

Based on recommended guidelines(4–6), the distinction between D'Amico low and intermediate risk 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 H or I disease from L or NEC upon 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; Supplementary Figure 3), and were significantly better than a random predictor ($p < 0.001$, bootstrap test, 2,000 resamples). However, PUR-1 and PUR-4 were best at discerning significant disease and were equally effective; AUCs for both PUR-4 and for PUR-1 in the Training and Test cohorts were respectively 0.81 (95% CI: 0.77 - 0.85) and 0.77 (95% CI: 0.70 - 0.84), (Figure 2C & D).

When Gleason score 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) (Figure 3). The ability to predict Gs ≥ 3+4 was particularly relevant because this was previously 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*., 2016; Tomlins *et al.*, 2016 and Van Neste *et al.*, 2016, respectively.

Decision curve analysis (DCA)(27) examined the potential net benefit of using PUR-signatures in a non-PSA screened population. Biopsy of men based upon their PUR-4 score provided a net benefit over biopsy of men based on current clinical practice across all thresholds (Figure 4). When DCA was also undertaken within the context of a PSA-screened population, PUR continued to provide a net benefit (Supplementary Figure 4).

Active surveillance cohort:

Within the Movember cohort were 87 men enrolled in AS at the Royal Marsden Hospital, UK. The median follow-up time from initial urine sample collection was 5.7 years (range 5.1 – 7.0 years) (Supplementary Table 2). The median time from initial urine sample collection to progression or final

follow up was 503 days (range 0.1 – 7.4 years). The PUR profiles from these men were used to investigate the prognostic utility of PUR beyond categorising D'Amico Risk. The PUR profiles were significantly different between the 23 men who progressed within five years of urine sample collection, and the 49 men who did not progress ($p < 0.001$, Wilcoxon rank sum test; Figure 5A). Twenty-two men progressed by the criteria detailed above, with an additional nine men progressing based solely on MP-MRI criteria. Further AS cohort characteristics are available in Supplementary Table 2.

Calculation of the Kaplan-Meier plots 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, Figure 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 dichotomise AS patients (PUR-4 = 0.174, based on the median optimal threshold to minimise Log-rank test *p*-value from 10,000 resamples of the cohort with replacement). 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, Figure 5C, HR = 8.23; 95% CI: 3.26 – 20.81). This result is robust ($p < 0.05$ in 99.8% of 100,000 cohort resamples with replacement, see Methods for full details).

When MP-MRI criteria for progression was also included, both primary PUR-status and dichotomised PUR threshold remained a significant predictor of progression ($p < 0.001$ log–rank test, Supplementary Figure 5). When the AS cohort were 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 for men with Intermediate-risk disease ($p < 0.001$ log–rank test, Supplementary Figure 6).

Multiple urine specimens had been collected for 20 of the men entered into the AS trial, allowing us to assess the stability of urine profiles over time (Supplementary Figure 7). In patients that 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 men 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 subcategorisation of intermediate risk disease into favourable and unfavourable groups(32) and the development of the CAPRA classification system(8). Other approaches include the development of an unsupervised classification framework(33) and of biomarkers of aggressive disease, as illustrated by Cuzick *et al.* (2012), Knezevic *et al.* (2013) and Robert *et al.* (2013)(34,36,37). 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(13,14,17,38). Here we have constructed the 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 AS men.

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(13,14,17,38). Here 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 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) to these published models using *PCA3*/*TMPRSS2-ERG* (AUC, 0.74 - 0.78)(13,14) and *HOXC6*/*DLX1* (AUC, 0.77)(17).

Current clinical practice assesses patient's disease using PSA, needle biopsy of the prostate and MP-MRI. However, up to 75% of men with a raised PSA (≥3 ng/ml) are negative for prostate cancer on biopsy(6,39), whilst in absence of a raised PSA, 15% of men are found to have prostate cancer, with a further 15% of these cancers being high-grade(40). 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 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 MP-MRI 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 are also well recognised for patients entered onto AS surveillance(41). We found that the PUR framework worked well when applied to men on AS monitored by PSA and biopsy, and also in patients monitored by MP-MRI. A potential limitation of this study is that we have not been able to test the PUR stratification in an independent and more conservatively managed active surveillance cohort. However, based on our observations approximately 13% of the RMH AS cohort could have been safely removed from AS monitoring for a minimum of five years. An interesting feature is that in some patients the PUR urine signature predicted disease progression up to five years before it was detected by standard clinical methods. This prognostic information could potentially also aid the reduction of patient-elected radical intervention in active surveillance men which in some cohorts can be as high as 75% within three years of enrolment(41). Indeed, we would view the use of PUR within the context of active surveillance 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 active surveillance using other previously described longitudinal cohorts.

In conclusion, we have shown that PUR represents a new & versatile urine biomarker system capable of detecting aggressive prostate cancer and predicting the need for therapeutic intervention in AS men. The dramatic differences in RNA expression profiles across the spectrum from high risk cancer to patients with no evidence of cancer, 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.

Funding Sources:

This study was possible thanks to the Movember Foundation GAP1 Urine Biomarker project.

Conflicts of Interest:

The authors hold patent pending status on the work from this manuscript. There are no other conflicts of interest to disclose.

Author Contributions:

DSB, CSC & JC had joint and equal contributions to senior authorship. SPC, JC, CSC & DSB drafted the manuscript. SPC, HC, DP, and DSB performed NanoString data analyses. SPC, and DSB performed the statistical analyses. MH, FC, RM, and CP setup clinical collection and developed clinical methodologies. MGS, ASP, JS, HP, HW & IGM all conceived gene-probes for NanoString interrogation. RH, MW, HC, HW, KLP, ASP, ND, MGS, CP & CS were involved in sample collection, extraction and preparation at their respective institutes. RWB oversaw histopathological analysis of biopsies. CSC, JC, JS, FC, MH, IG, SPC, MH, FC, ND, CS, CP, and DSB conceived and designed the studies. The Movemember GAP1 Urine Biomarker Consortium oversaw project management and experimental design.

References:

1. D'Amico A V., 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 Jun;21(11):2163–72.

2. D'Amico A V., Whittington R, Bruce Malkowicz S, Schultz D, Blank K, Broderick GA, 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 Sep 16;280(11):969–74.

3. Gleason DF MG. Prediction of prognosis for prostatic Staging, adenocarcinoma by combined histological grading and slinical. J Urol. 1974 Jan 1;111(1):111:58–64.

4. Sanda MG, Cadeddu JA, Kirkby E, Chen RC, Crispino T, Fontanarosa J, et al. Clinically Localized Prostate Cancer: AUA/ASTRO/SUO Guideline. Part I: Risk Stratification, Shared Decision Making, and Care Options. J Urol. 2018;199(3):683–90.

5. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol. 2017;71(4):618–29.

6. National Institute for Health and Care Excellence. Prostate Cancer: diagnosis and treatment. 2014.

7. Selvadurai ED, Singhera M, Thomas K, Mohammed K, Woode-Amissah R, Horwich A, et al. Medium-term outcomes of active surveillance for localised prostate cancer. Eur Urol. 2013 Dec;64(6):981–7.

8. Cooperberg MR, Freedland SJ, Pasta DJ, Elkin EP, Presti JC, Amling CL, et al. Multiinstitutional validation of the UCSF cancer of the prostate risk assessment for prediction of recurrence after radical prostatectomy. Cancer. 2006;107(10):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 May 1;71(5):705–9.

10. Andreoiu M, Cheng L. Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. Vol. 41, Human Pathology. W.B. Saunders; 2010. p. 781–93.

11. Corcoran NM, Hovens CM, Hong MKH, Pedersen J, Casey RG, Connolly S, et al. Underestimation of Gleason score at prostate biopsy reflects sampling error in lower volume tumours. BJU Int. 2012 Mar;109(5):660–4.

12. Ahmed HU, El-Shater Bosaily A, Brown LC, Gabe R, Kaplan R, Parmar MK, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet. 2017 Feb 25;389(10071):815–22.

13. Tomlins SA, Day JR, Lonigro RJ, Hovelson DH, Siddiqui J, Kunju LP, et al. Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment. Eur Urol. 2016 Jul 1;70(1):45–53.

14. McKiernan J, Donovan MJ, O'Neill V, Bentink S, Noerholm M, Belzer S, et al. A novel urine exosome gene expression assay to predict high-grade prostate cancer at initial biopsy. JAMA Oncol. 2016 Jul 1;2(7):882–9.

15. Donovan MJ, Noerholm M, Bentink S, Belzer S, Skog J, O'Neill V, 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(4):370–5.

16. Tomlins SA, Day JR, Lonigro RJ, Hovelson DH, Siddiqui J, Kunju LP, et al. Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment. Eur Urol. 2016;70(1):45–53.

17. Van Neste L, Hendriks RJ, Dijkstra S, Trooskens G, Cornel EB, Jannink SA, et al. Detection of High-grade Prostate Cancer Using a Urinary Molecular Biomarker–Based Risk Score. Eur Urol. 2016;70(5):740–8.

18. Deantoni EP, Crawford ED, Oesterling JE, Ross CA, Berger ER, McLeod DG, et al. Age- and

race-specific reference ranges for prostate-specific antigen from a large community-based study. Urology. 1996 Aug;48(2):234–9.

19. Miranda KC, Bond DT, McKee M, Skog J, Punescu TG, Da Silva N, et al. Nucleic acids within urinary exosomes/microvesicles are potential biomarkers for renal disease. Kidney Int. 2010 Jul;78(2):191–9.

20. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 2008 Mar;26(3):317–25.

21. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007 Jan 1;8(1):118–27.

22. R Core Team. R: A Language and Environment for Statistical Computing. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing , Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. Vienna, Austria: R Foundation for Statistical Computing; 2018.

23. Archer KJ, Williams AAA. L1 penalized continuation ratio models for ordinal response prediction using high-dimensional datasets. Stat Med. 2012 Jun 30;31(14):1464–74.

24. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Vol. 58, Journal of the Royal Statistical Society. Series B (Methodological). WileyRoyal Statistical Society; 1996. p. 267–88.

25. Christensen RHB. ordinal Regression Models for Ordinal Data. 2018.

26. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.

27. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction

Models. Med Decis Mak. 2006;26(6):565–74.

28. Brown M. rmda: Risk Model Decision Analysis. 2018.

29. 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(21):2534–40.

30. Martin RM, Donovan JL, Turner EL, Metcalfe C, Young GJ, Walsh EI, et al. Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality. JAMA. 2018 Mar 6;319(9):883.

31. Pellegrini KL, Patil D, Douglas KJS, Lee G, Wehrmeyer K, Torlak M, et al. Detection of prostate cancer-specific transcripts in extracellular vesicles isolated from post-DRE urine. Prostate. 2017;77(9):990–9.

32. Aghazadeh MA, Frankel J, Belanger M, McLaughlin T, Tortora J, Staff I, et al. National Comprehensive Cancer Network® Favorable Intermediate Risk Prostate Cancer—Is Active Surveillance Appropriate? J Urol. 2018 May 1;199(5):1196–201.

33. Luca BA, Brewer DS, Edwards DR, Edwards S, Whitaker HC, Merson S, et al. DESNT: A Poor Prognosis Category of Human Prostate Cancer. European Urology Focus. 2017 Mar 6;

34. Robert G, Jannink S, Smit F, Aalders T, Hessels D, Cremers R, et al. Rational basis for the combination of PCA3 and TMPRSS2:ERG gene fusion for prostate cancer diagnosis. Prostate. 2013 Jan;73(2):113–20.

35. Cuzick J, Berney DM, Fisher G, Mesher D, Møller H, Reid JE, 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 Mar 13;106(6):1095–9.

36. Knezevic D, Goddard AD, Natraj N, Cherbavaz DB, Clark-Langone KM, Snable J, et al.

Analytical validation of the Oncotype DX prostate cancer assay - a clinical RT-PCR assay optimized for prostate needle biopsies. BMC Genomics. 2013 Oct 8;14(1):690.

37. Cuzick J, Berney DM, Fisher G, Mesher D, Møller H, Reid JE, 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 Mar;106(6):1095–9.

38. Hessels D, Klein Gunnewiek JMT, Van Oort I, Karthaus HFM, Van Leenders GJL, Van Balken B, et al. DD3PCA3-based molecular urine analysis for the diagnosis of prostate cancer. Eur Urol. 2003 Jul 1;44(1):8–16.

39. Lane JA, Donovan JL, Davis M, Walsh E, Dedman D, Down L, 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 Sep 1;15(10):1109–18.

40. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level ≤4.0 ng per Milliliter. N Engl J Med. 2004 May 27;350(22):2239–46.

41. Simpkin AJ, Tilling K, Martin RM, Lane JA, Hamdy FC, Holmberg L, et al. Systematic review and meta-analysis of factors determining change to radical treatment in active surveillance for localized prostate cancer. Vol. 67, European Urology. 2015. p. 993–1005.

Figures & Tables:

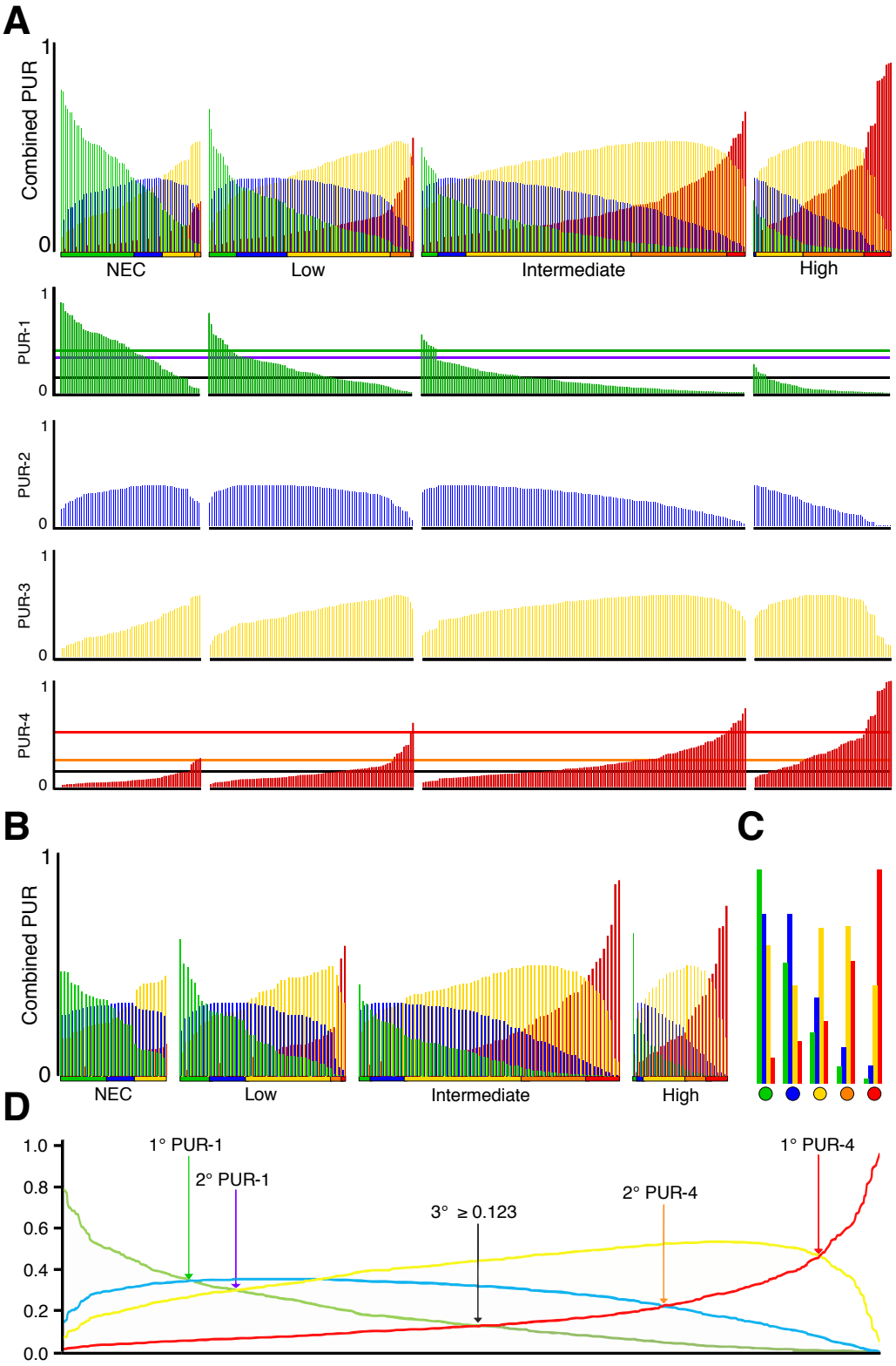

Figure 1. A) PUR profiles (PUR-1 – green, PUR-2 – blue, PUR-3 – yellow, PUR-4 – red) for the Training cohort, grouped by D'Amico risk group and ordered by ascending PUR-4 score. Horizontal lines indicate where the PUR thresholds lie for: 1° PUR-1 (Green, 0.342), 2° PUR-1 (Purple, 0.297), 1° PUR-4 (Red, 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 lie.

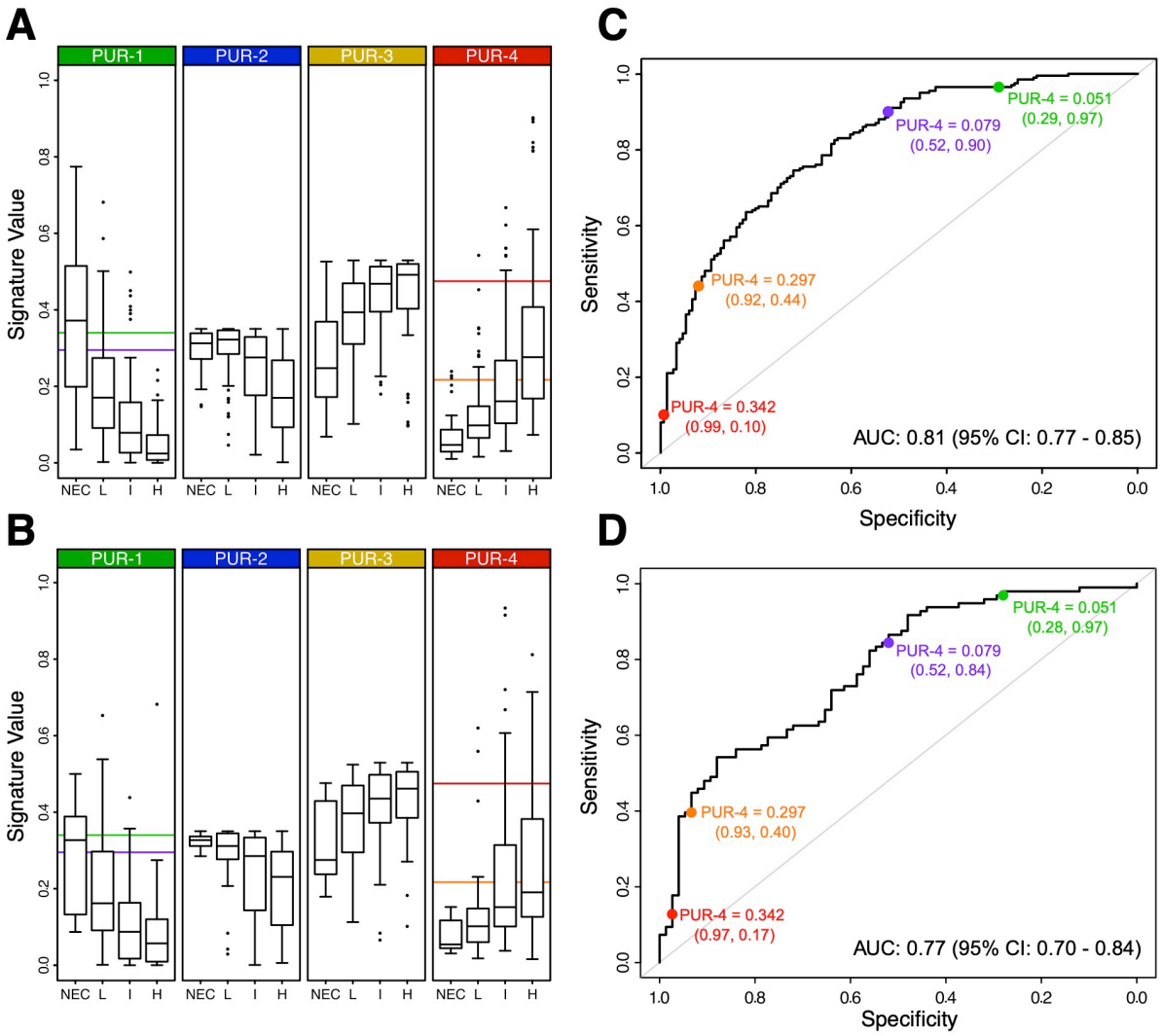

Figure 2. A & B) Boxplots of PUR signatures in samples categorised as no evidence of cancer (NEC, *n* = 62 (Training), *n* = 30 (Test)) and D'Amico risk categories; (L – Low, *n* = 89 (Training), *n* = 45 (Test), I – Intermediate, *n* = 131 (Training), *n* = 69 (Test) and H – High risk, *n* = 61 (Training), *n* = 27 (Test)) in **A)** the Training and **B)** Test cohorts. 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 & 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)** Training and **D)** Test cohorts. 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

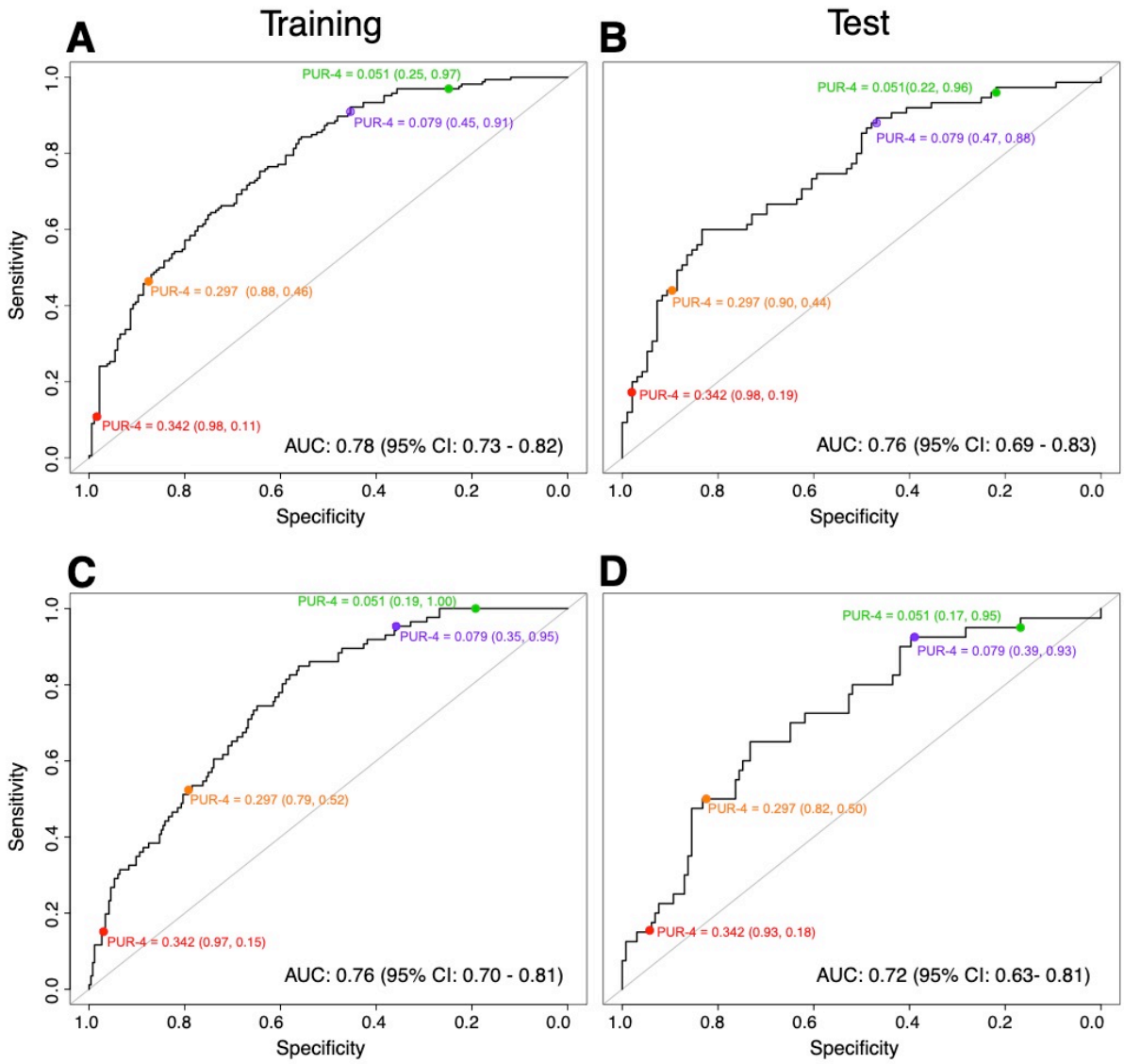

Figure 3. AUC curves for PUR-4 predicting the presence/absence of Gleason ≥ 7 on initial biopsy in Training and Test cohorts (A and B, respectively) and Gleason ≥ 4+3 in Training and Test cohorts (C and D, respectively). Coloured circles indicate the specificity and sensitivity, respectively, of thresholds 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

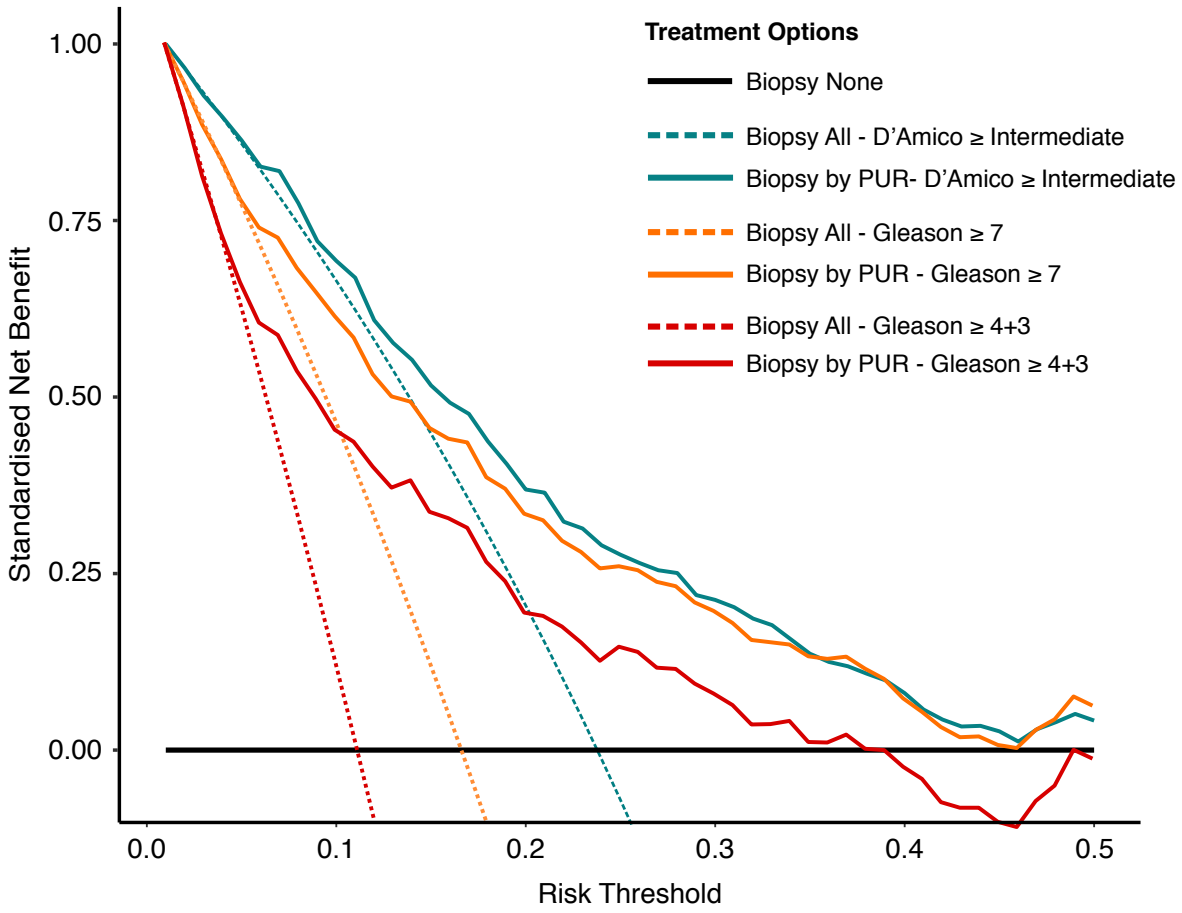

Figure 4. DCA plot depicting the standardised net benefit of adopting PUR-4 as a continuous predictor for detecting significant cancer on initial biopsy, when significant is defined as: D'Amico risk group of Intermediate or greater (teal), Gs ≥ 3+4 (orange) or Gs ≥ 4+3 (red). To assess benefit in the context of cancer arising in a non-PSA screened population of men we used data from the control arm of the CAP study(30). 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.

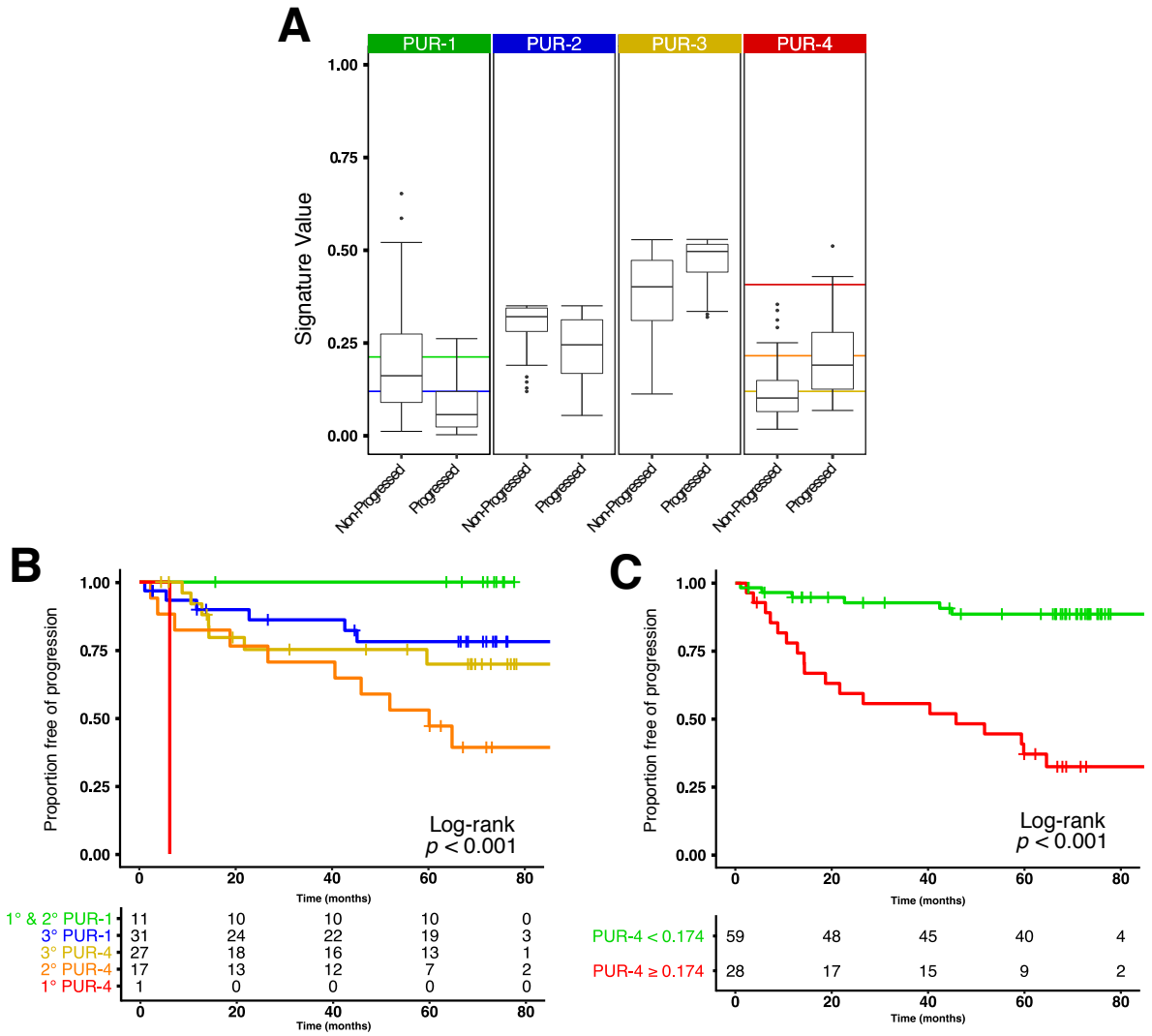

Figure 5. A) PUR profiles of patients on active surveillance that had met the clinical criteria, not including MP-MRI criteria, for progression (n = 23) or not (n = 49) at five years post urine sample collection. Progression criteria were either: PSA velocity >1 ng/ml per year or Gs ≥ 4+3 or ≥ 50% cores positive for cancer on repeat biopsy. PUR signatures for progressed vs non-progressed samples were significantly different for all PUR signature ($p < 0.001$, Wilcoxon rank sum test). Horizontal line colour indicates the thresholds for PUR categories described in: **B)** Kaplan-Meier plot of progression in active surveillance 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 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.

Tables:

Table 1. Characteristics of the Training and Test cohorts.

Characteristic	Training	Test	*p*-value
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)	0.85
Age, yr, mean (median; IQR)	65.8 (67, 11)	67.2 (67, 11)	0.71
Family history of PCa, %; no, yes, NA	3.0, 6.1, 90.8	0.6, 6.2, 93.3	1
First biopsy, *n* (%)	298 (82.78)	145 (81.46)	1
Prostate volume, ml; mean (median; IQR)	59.2 (49.8, 30.4)	61.1 (49.2, 32.8)	0.95
PSAD, ng/ml; ml, mean (median; IQR)	0.29 (0.19, 0.16)	0.29 (0.18, 0.17)	0.95
Suspicious DRE, *n*	107	52	1
Diagnosis, *n*:	358	177	0.9
NEC, n (%)	62 (17.3)	30 (17.0)	-
D'Amico Low *n* (%)	89 (24.9)	45 (25.4)	-
D'Amico Intermediate *n* (%)	139 (38.8)	69 (39.0)	-
D'Amico High *n* (%)	61 (17.0)	27 (15.3)	-
Metastatic (bone scan) *n* (%)*	7 (2.0)	6 (3.3)	-
CAPRA, *n*:	288	145	1
Low (0-2) *n* (%)	97 (33.7)	49 (33.7)	-
Intermediate (3-5) *n* (%)	108 (37.5)	53 (36.6)	-
High (≥6) *n* (%)	83 (28.8)	43 (29.7)	-
Gleason, *n*:	292	144	0.5
Gs = 6, *n* (%)	119 (40.8)	64 (44.4)	-
Gs = 7, *n* (%)	131 (44.9)	56 (38.9)	-
Gs ≥ 8 *n* (%)	42 (14.4)	24 (16.7)	-

DRE = digital rectal examination; Gs = Gleason score; IQR = interquartile range; NA = not available; PCa = prostate cancer; PSA = prostate-specific antigen; PSAD = prostate-specific antigen density; TRUS = transrectal ultrasound. NEC=No Evidence of Cancer/PSA normal for age or <1ng/ml. *Metastatic men were diagnosed as High risk at time of urine collection. Percentages reported for CAPRA and Gleason headings are calculated with the data available for that heading; out of the 535 samples, only 467 data available for

CAPRA groupings and 436 Gleason scores available, where no biopsy took place, was refused or information was incomplete.

Table 2. Gene probes incorporated by LASSO regularisation in the optimal model.

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	*TDRD*
ITGBL1	*TMPRSS2/ERG fusion*
KLK4	*TRPM4*
MARCH5	*TWIST1*
MED4	*UPK2*