



Article

Integration of Urinary EN2 Protein & Cell-Free RNA Data in the Development of a Multivariable Risk Model for the Detection of Prostate Cancer Prior to Biopsy

Shea P. Connell ¹, Robert Mills ², Hardev Pandha ³, Richard Morgan ⁴, Colin S. Cooper ¹, Jeremy Clark ¹, Daniel S. Brewer ^{1,5,*} and The Movember GAP1 Urine Biomarker Consortium ^{1,†}

¹ Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK; sheaconnell@gmail.com (S.P.C.); colin.cooper@uea.ac.uk (C.S.C.); Jeremy.Clark@uea.ac.uk (J.C.)

² Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, Norfolk NR4 7UY, UK; robert.mills@nnuh.nhs.uk

³ Faculty of Health and Medical Sciences, The University of Surrey, Guildford GU2 7XH, UK; h.pandha@surrey.ac.uk

⁴ School of Pharmacy and Medical Sciences, University of Bradford, Bradford BD7 1DP, UK; r.morgan3@bradford.ac.uk

⁵ The Earlham Institute, Norwich Research Park, Norwich, Norfolk NR4 7UZ, UK

* Correspondence: d.brewer@uea.ac.uk; Tel.: +44-(0)-1603-593761

† 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.



Citation: Connell, S.P.; Mills, R.; Pandha, H.; Morgan, R.; Cooper, C.S.; Clark, J.; Brewer, D.S.; Integration of Urinary EN2 Protein & Cell-Free RNA Data in the Development of a Multivariable Risk Model for the Detection of Prostate Cancer Prior to Biopsy. *Cancers* **2021**, *13*, 2102. <https://doi.org/10.3390/cancers13092102>

Academic Editor: Jonas Cicenias

Received: 15 February 2021

Accepted: 14 April 2021

Published: 27 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Prostate cancer is a disease responsible for a large proportion of all male cancer deaths but there is a high chance that a patient will die with the disease rather than from. Therefore, there is a desperate need for improvements in diagnosing and predicting outcomes for prostate cancer patients to minimise overdiagnosis and overtreatment whilst appropriately treating men with aggressive disease, especially if this can be done without taking an invasive biopsy. In this work we develop a test that predicts whether a patient has prostate cancer and how aggressive the disease is from a urine sample. This model combines the measurement of a protein-marker called EN2 and the levels of 10 genes measured in urine and proves that integration of information from multiple, non-invasive biomarker sources has the potential to greatly improve how patients with a clinical suspicion of prostate cancer are risk-assessed prior to an invasive biopsy.

Abstract: The objective is to develop a multivariable risk model for the non-invasive detection of prostate cancer prior to biopsy by integrating information from clinically available parameters, Engrailed-2 (EN2) whole-urine protein levels and data from urinary cell-free RNA. Post-digital-rectal examination urine samples collected as part of the Movember Global Action Plan 1 study which has been analysed for both cell-free-RNA and EN2 protein levels were chosen to be integrated with clinical parameters ($n = 207$). A previously described robust feature selection framework incorporating bootstrap resampling and permutation was applied to the data to generate an optimal feature set for use in Random Forest models for prediction. The fully integrated model was named ExoGrail, and the out-of-bag predictions were used to evaluate the diagnostic potential of the risk model. ExoGrail risk (range 0–1) was able to determine the outcome of an initial trans-rectal ultrasound guided (TRUS) biopsy more accurately than clinical standards of care, predicting the presence of any cancer with an area under the receiver operator curve (AUC) = 0.89 (95% confidence interval(CI): 0.85–0.94), and discriminating more aggressive Gleason $\geq 3 + 4$ disease returning an AUC = 0.84 (95% CI: 0.78–0.89). The likelihood of more aggressive disease being detected significantly increased as ExoGrail risk score increased (Odds Ratio (OR) = 2.21 per 0.1 ExoGrail increase, 95% CI: 1.91–2.59). Decision curve analysis of the net benefit of ExoGrail showed the potential to reduce the numbers of unnecessary biopsies by 35% when compared to current standards of care. Integration of information from multiple, non-invasive biomarker sources has the potential to greatly improve how patients with a clinical suspicion of prostate cancer are risk-assessed prior to an invasive biopsy.

Keywords: prostate cancer; biomarker; urine; machine learning; TRIPOD; liquid biopsy

1. Introduction

Prostate cancer is responsible for 13% of all male cancer deaths in the UK, yet this is contrasted by 10-year survival rates approaching 84% [1]. This dichotomy has led to uncertainty for clinicians in how best to diagnose and predict the outcome for prostate cancer patients to minimise overdiagnosis and overtreatment whilst appropriately treating men with aggressive disease [2]. More accurate discrimination of disease state in biopsy naïve men would mark a significant development compared to current standards and impact large numbers of patients suspected of harbouring prostate cancer. The development of such a pre-biopsy screening test would provide a convenient checkpoint along the clinical pathway for patients to exit without the need for further invasive and stressful follow-up.

Under current guidelines patients are selected for further clinical investigations for prostate cancer if they have an elevated prostate specific antigen (PSA) (≥ 4 ng/mL) and/or an adverse finding on digital rectal examination (DRE) or lower urinary tract symptoms, whilst other factors such as age and ethnicity are also considered alongside patient preference [3–5]. More recently multiparametric MRI (mpMRI) has been used as a triage tool to reduce negative biopsy rates since its validation in the PROMIS clinical trial [6]. However, as it has gained more widespread adoption, mpMRI has shown a higher rate of inter-operator and inter-machine variability than reported in controlled clinical trials; up to 28% of clinically significant disease is missed in practice [5,7–9]. Coupled with the relative expense, time and expertise required to undertake an mpMRI meeting the current clinical guidelines, there is a need to improve on current clinical practices.

Biomarkers utilising tissue samples taken at the time of diagnosis for the detection of aggressive or significant prostate cancer requiring clinical attention are relatively plentiful [10–13]. Many of these markers are good tests, whether that be for discerning the most aggressive disease [11,14], or for predicting disease-free survival following radical prostatectomy [15]. However, requiring tissue means a biopsy must already have been performed, making these tests incompatible with reducing the rates of unnecessary biopsy that come at considerable economic, psychological and societal cost to patients and healthcare systems alike [2,16,17].

As a secretory organ directly interacting with the male urinary tract, the prostate is well-placed as a candidate for non-invasive liquid biopsy from urine samples [18]. Single- or few-biomarker panels such as Engrailed-2 (EN2) protein expression [19], the SelectMDx [20] and ExoDx Prostate (IntelliScore) [21] tests have published promising results for the non-invasive detection of significant disease (Gleason score (Gs) ≥ 7). However, they are in various stages of clinical validation and none are currently implemented in the UK healthcare system [5]. Most urinary biomarkers developed to date for the prediction of biopsy outcome are unimodal; considering a singular fraction of urine (such as the cell-pellet or cell-free fractions) or biological aspect of cancer to appraise disease status. Whilst these tests have shown promising clinical use and accuracy, for the majority it has not yet been explored whether extra predictive value could be derived by integrating multiple streams of information from other sources.

Since initial development, the SelectMDx model has been updated to include clinically available parameters of serum PSA, patient age and DRE alongside urinary *HOXC6* and *DLX1* mRNA, adding significant predictive ability for patients with a PSA < 10 ng/mL [22]. We have also recently shown the benefit of such a holistic approach, presenting the development of the multivariable ExoMeth risk prediction model integrating clinical parameters, hypermethylation within the urinary cell pellet and urinary cell-free RNA expression data that displayed improved clinical utility over any single mode [23].

EN2 is a homeodomain-containing transcription factor that has an essential function in early development, which in mammals includes the delineation of the midbrain/hindbrain

border [24]. For a transcription factor it has a number of unusual properties, including the ability to be secreted from cells and taken up by others [25]. Indeed, a recent study indicated that prostate cancer cells can secrete EN2 protein through vesicles which are then taken up by other non-EN2 expressing cells, where it can directly influence the transcription of target genes [25].

This secretory behaviour of EN2 makes it a potential biomarker for prostate cancer, and indeed EN2 protein can be detected in the urine of men with prostate tumours [19]. The original and subsequent studies have generally supported a diagnostic role for urinary EN2, including a relationship between urinary EN2 concentration and tumour volume [19,26]. More recently, a lateral flow-based test for EN2 has been described that could potentially allow point-of-care testing [27].

In this study, we report the utility of a predictive model produced by the integration of clinically available parameters, urinary EN2 protein levels and targeted cell-free RNA transcriptomics. The data were collected within the Movember Global Action Plan 1 (GAP1) study that explored a range of biomarkers in urine for PCa diagnosis and prognosis. The clinical utility of this model is determined by the ability to predict the presence of $G_s \geq 7$ and $G_s \geq 4 + 3$ disease on biopsy, both critical distinctions in clinical settings, where patients with $G_s \geq 7$ are recommended radical therapy [5], whilst patients with $G_s 4 + 3$ have significantly worse outcomes than $G_s 3 + 4$ patients [28]. Aware that most cancer biomarkers and predictive models fail to reach clinical adoption, we have adhered to the guidelines for the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) whilst developing the models and results presented here [29].

2. Materials and Methods

2.1. Patient Population and Characteristics

The full Movember GAP1 urine cohort comprises 1257 first-catch post-DRE urine samples collected between 2009 and 2015 from urology clinics at multiple sites, as described in Connell et al. (2019). As a diverse range of techniques was applied to samples from this cohort and restricted amounts of urine, the number of experiments that could be performed on any one sample was limited. Samples within the Movember cohort that were quantified for both EN2 levels by ELISA and cell-free-RNA (cf-RNA) expression by NanoString (Seattle, WA, USA) were eligible for selection for model development in the current study ($n = 218$).

Exclusion criteria for model development included a recent prostate biopsy or transurethral resection of the prostate (<6 weeks) and metastatic disease (confirmed by a positive bone-scan or PSA > 100 ng/mL), resulting in a cohort of 207 samples, deemed the ExoGrail cohort (Table 1). All samples analysed in the ExoGrail cohort were collected from the Norfolk and Norwich University Hospital (NNUH, Norwich, UK). Sample collections and processing were ethically approved by the East of England REC.

2.2. Sample Processing and Analysis

Urine samples were processed according to the Movember GAP1 standard operating procedure (Supplementary Methods). In brief, within 30 min of collection, urine was centrifuged ($1200 \times g$ 10 min, $6^\circ C$) to remove cellular material. Supernatant extracellular vesicles were harvested by microfiltration and cell-free mRNA extracted (RNeasy micro kit, #74004, Qiagen, Hilden, Germany) on the same day that they were provided by the patient. RNA was amplified as cDNA with an Ovation PicoSL WTA system V2 (Nugen, Redwood City, CA, USA, #3312-48). Urinary EN2 protein concentration was quantified by ELISA from whole urine using a monoclonal anti-mouse EN2 antibody, as described by Morgan et al. (2011) [19]. Cell-free mRNA was quantified from urinary extracellular vesicles using NanoString technology, with 167 gene-probes (Table S1), as described in Connell et al. (2019), with the modification that NanoString data were normalised according to NanoString guidelines using NanoString internal positive controls, and \log_2 transformed.

Clinical variables serum PSA, age at sample collection, DRE finding, and urine volume collected were considered.

Table 1. Characteristics of the ExoGrail development cohort, stratified according to a record of cancer or not, either on an initial biopsy or for a no cancer finding if a biopsy was not performed.

| | No Cancer Finding: | Biopsy Positive Cancer Finding |
|-------------------------------|----------------------|--------------------------------|
| Collection Centre: | | |
| NNUH, n (%) | 77 (100) | 130 (100) |
| Age: | | |
| minimum | 45.00 | 53.00 |
| median (IQR) | 65.00 (59.00, 71.00) | 68.50 (65.00, 76.00) |
| mean (sd) | 65.22 ± 8.10 | 69.71 ± 7.67 |
| maximum | 82.00 | 91.00 |
| PSA: | | |
| minimum | 0.30 | 4.10 |
| median (IQR) | 6.10 (3.70, 8.80) | 10.35 (6.82, 16.48) |
| mean (sd) | 7.89 ± 8.72 | 17.08 ± 18.33 |
| maximum | 63.80 | 95.90 |
| Prostate Size (DRE Estimate): | | |
| Small, n (%) | 13 (17) | 13 (10) |
| Medium, n (%) | 34 (44) | 64 (49) |
| Large, n (%) | 21 (27) | 38 (29) |
| Unknown, n (%) | 9 (12) | 15 (12) |
| Gleason Score: | | |
| 0, n (%) | 77 (100) | 0 (0) |
| 6, n (%) | 0 (0) | 30 (23) |
| 3 + 4, n (%) | 0 (0) | 48 (37) |
| 4 + 3, n (%) | 0 (0) | 24 (18) |
| ≥8, n (%) | 0 (0) | 28 (22) |
| Biopsy Outcome: | | |
| No Biopsy, n (%) | 25 (32) | 0 (0) |
| Biopsy Negative, n (%) | 52 (68) | 0 (0) |
| Biopsy Positive, n (%) | 0 (0) | 130 (100) |

2.3. Statistical Analysis

All analyses, model construction and data preparation were undertaken in R version 3.5.3 [30], and unless otherwise stated, utilised base R and default parameters. All data and code required to reproduce these analyses can be found at the UEA Cancer Genetic GitHub repository [31].

2.4. Feature Selection

In total, 172 variables were available for prediction (cf-RNA ($n = 167$), clinical variables ($n = 4$) and urinary EN2 ($n = 1$); for full list see Table S1), making feature selection a key task for minimising model overfitting and increasing the robustness of trained models. To avoid dataset-specific features being positively selected [32], we implemented a robust feature selection workflow utilising the Boruta algorithm [33] and bootstrap resampling. Boruta is a random forest-based algorithm that iteratively compares feature importance against random predictors, deemed “shadow features.” Features that perform significantly worse compared to the maximally performing shadow feature at each permutation, ($p \leq 0.01$, calculated by Z-score difference in mean accuracy decrease) are consecutively dropped until only confirmed, stable features remain.

Boruta was applied on 1000 datasets generated by resampling with replacement. Features were only positively selected for model construction when confirmed as stable features in $\geq 90\%$ of resampled Boruta runs.

2.5. Comparator Models

To evaluate potential clinical utility, additional models were trained as comparators using subsets of the available variables across the patient population: a clinical standard of care (SoC) model was trained by incorporating age, PSA, T-staging and clinician DRE impression; a model using only the values from the EN2 ELISA (EN2, $n = 1$); and a model only using NanoString gene-probe information (NanoString, $n = 167$). The fully integrated ExoGrail model was trained by incorporating information from all of the above variables ($n = 177$). Each set of variables for comparator models were independently selected via the bootstrapped Boruta feature selection process described above to select the most optimal subset of variables possible for each predictive model.

2.6. Model Construction

All models were trained via the random forest algorithm [34], using the *randomForest* package [35] with default parameters except for resampling without replacement and 401 trees being grown per model. Risk scores from trained models are presented as the out-of-bag predictions; the aggregated outputs from decision trees within the forest where the sample in question has not been included within the resampled dataset [34]. Bootstrap resamples were identical for feature selection and model training for all models and used the same seed for random number generator.

Models were trained on a modified continuous label, based on biopsy outcome and constructed as follows: samples were scored on a continuous scale (range: 0–1) according to the dominant Gleason pattern: where 0 represented no evidence of cancer, Gleason scores 6 & 3 + 4 were assigned to 0.5 and Gleason scores $\geq 4 + 3$ are set to 1. Following this categorisation, the score is treated as a continuous variable by the Random Forest algorithm described above. This process was designed to recognise that two patients with the same TRUS-biopsy Gleason score will not share the exact same proportions of tumour pattern, or overall disease burden. This scale was solely used for model training and was not represented in any endpoint measurements, or for determining the predictive ability and clinical utility.

2.7. Statistical Evaluation of Models

Area Under the Receiver-Operator Characteristic curve (AUC) metrics were produced using the *pROC* package [36], with confidence intervals calculated via 1000 stratified bootstrap resamples. Density plots of model risk scores, and all other plots were created using the *ggplot2* package [37]. Partial dependency plots were calculated using the *pdp* package [38]. Cumming estimation plots and calculations were produced using the *dabestr* package [39] and 1000 bootstrap resamples were used to visualise robust effect size estimates of model predictions.

Decision curve analysis (DCA) [40] examined the potential net benefit of using PUR-signatures in the clinic. Standardised net benefit (sNB) was calculated with the *rmda* package [41] and presented throughout our decision curve analyses as it is a more directly interpretable metric compared to net benefit [42]. In order to ensure DCA was representative of a more general population, the prevalence of Gleason scores within the ExoGrail cohort were adjusted via bootstrap resampling to match those 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 [43], as described in Connell et al. (2019). Briefly, of the biopsied men within this CAP cohort, 23.6% were Gs 6, 8.7% Gs 7 and 7.1% Gs ≥ 8 , with 60.6% of biopsies showing no evidence of cancer. These ratios were used to perform stratified bootstrap sampling with a replacement of the Movember cohort to produce a “new” dataset of 197 samples with risk scores from each comparator model. sNB was then calculated for this resampled dataset, and the process repeated for a total of 1000 resamples with replacement. The mean sNB for each risk score and the “treat-all” options over all of the iterations were used to produce the presented figures to account for variance in

resampling. Net reduction in biopsies, based on the adoption of models versus the default treatment option of undertaking biopsy in all men with PSA ≥ 4 ng/mL was calculated as:

$$Biopsy_{NetReduction} = (NB_{Model} - NB_{All}) \times \frac{1 - Threshold}{Threshold} \quad (1)$$

where the decision threshold (*Threshold*) is determined by accepted patient/clinician risk [40]. For example, a clinician may accept up to a 25% perceived risk of cancer before recommending biopsy to a patient, equating to a decision threshold of 0.25.

3. Results

3.1. The ExoGrail Development Cohort

Urinary EN2 protein and cf-RNA data were available for 207 patients within the Movember GAP1 cohort, with all samples originating from the NNUH to form the ExoGrail development cohort (Table 1). The proportion of Gleason ≥ 7 disease in the ExoGrail cohort was 48%, whilst 25 patients were deemed to have no evidence of cancer (NEC, PSA < 4 ng/mL), and did not receive a biopsy.

3.2. Feature Selection and Model Development

Using the robust feature selection framework, four models were produced in total: a standard of care (SoC) model incorporating only clinically available parameters (age and PSA), a model using urinary EN2 protein levels as the sole predictor variable (Engrailed), a model using only cf-RNA information (ExoRNA, 11 gene-probes) and the integrated model, named ExoGrail that incorporated variables from all three sources (12 variables) (Table 2). The ExoGrail model is a multivariable risk prediction model incorporating clinical parameters, urinary EN2 protein levels and cf-RNA expression information. When the resampling strategy was applied for feature reduction using Boruta, 12 variables were selected for the ExoGrail model. Each of the retained variables were positively selected in every resample and notably included information from clinical and cf-RNA variables, as well as urinary EN2 (Figure 1; full resample-derived Boruta variable importance for the SoC, Engrailed and ExoRNA comparator models can be seen in Figures S1–S3, respectively).

Table 2. Features positively selected for each model by bootstrap resampling and the Boruta algorithm. Features are selected for each model by being confirmed as important for predicting biopsy outcome, categorised as a modified ordinal variable (see Methods) by Boruta in $\geq 90\%$ of bootstrap resamples.

| | SoC | Engrailed | ExoRNA | ExoGrail |
|---------------------------|------------------|-------------|--|--|
| Clinical Parameters | Serum PSA Age | - - | - - | Serum PSA - |
| ELISA Targets | | EN2 (ELISA) | - | EN2 (ELISA) |
| | | | ERG exons 4-5 ERG exons 6-7 | ERG exons 4-5 ERG exons 6-7 |
| | | | GJB1 HOXC6 HPN | GJB1 HOXC6 HPN |
| NanoString cf-RNA targets | | | NKAIN1 PCA3 PPFIA2 RPLP2 - | - PCA3 PPFIA2 - |
| | | | TMEM45B TMPRSS2/ERG fusion | SLC12A1 TMEM45B TMPRSS2/ERG fusion |

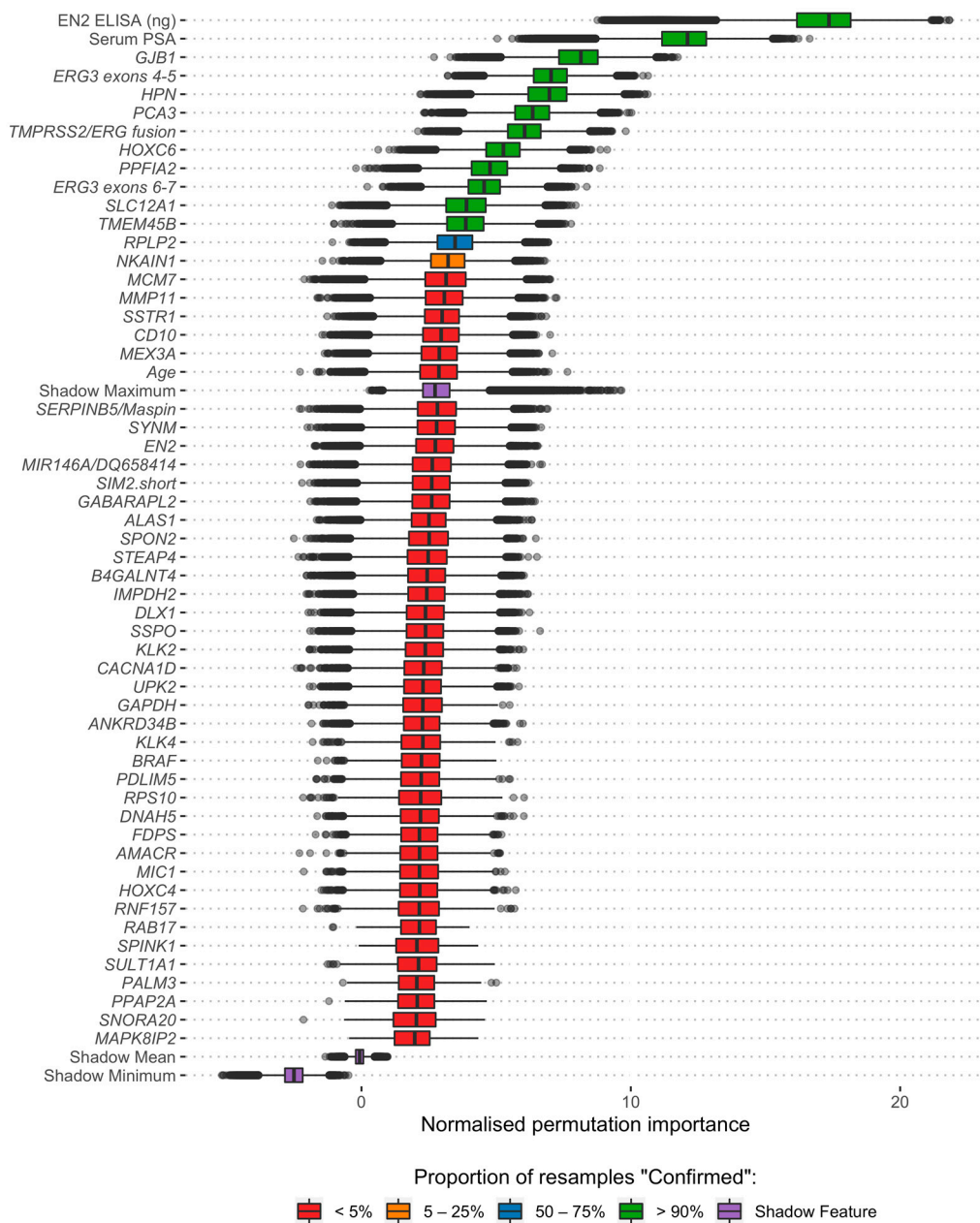


Figure 1. Analysis of variables available for the training of the ExoGrail model through the application of the Boruta algorithm via bootstrap resampling. 1000 resamples with replacement of the available data were made, with the normalised permutation importance of each variable recorded at each iteration, along with the decision of Boruta within that resample. Fill colour shows the proportion of resamples that a feature was positively retained by Boruta. Those features selected in $\geq 90\%$ of resamples were selected for fitting predictive models. Variables rejected in all of the 1000 resamples are not shown here but are fully detailed in Table S1.

In the SoC comparator model, only PSA and age were selected as important predictors. Urinary EN2 levels were confirmed as important in the independent Engrailed model as the sole variable, and also within the ExoGrail model (Table 2). For the cf-RNA model, ExoRNA, 11 transcripts were selected, notably including both variants of the ERG gene-probe and *TMPRSS2/ERG* fusion gene-probe. ExoGrail incorporated an additional cf-RNA transcript, *SLC12A1*, which was not previously selected in the ExoRNA comparator model. When this was examined by partial dependency plots, an additive interaction effect was observed between quantified levels of urinary EN2 and counts of *SLC12A1* on the predicted ExoGrail risk signature output (Figure S4).

3.3. ExoGrail Predictive Ability

As ExoGrail Risk Score (range 0–1) increased, the likelihood of high-grade disease detection on TRUS-biopsy was significantly greater (Proportional odds ratio = 2.21 per 0.1 ExoGrail increase, 95% CI: 1.91–2.59; ordinal logistic regression, Figure 2). The median ExoGrail risk score for metastatic patients was 0.76 ($n = 11$). These patients were excluded from model training and can be considered as a positive control for model calibration.

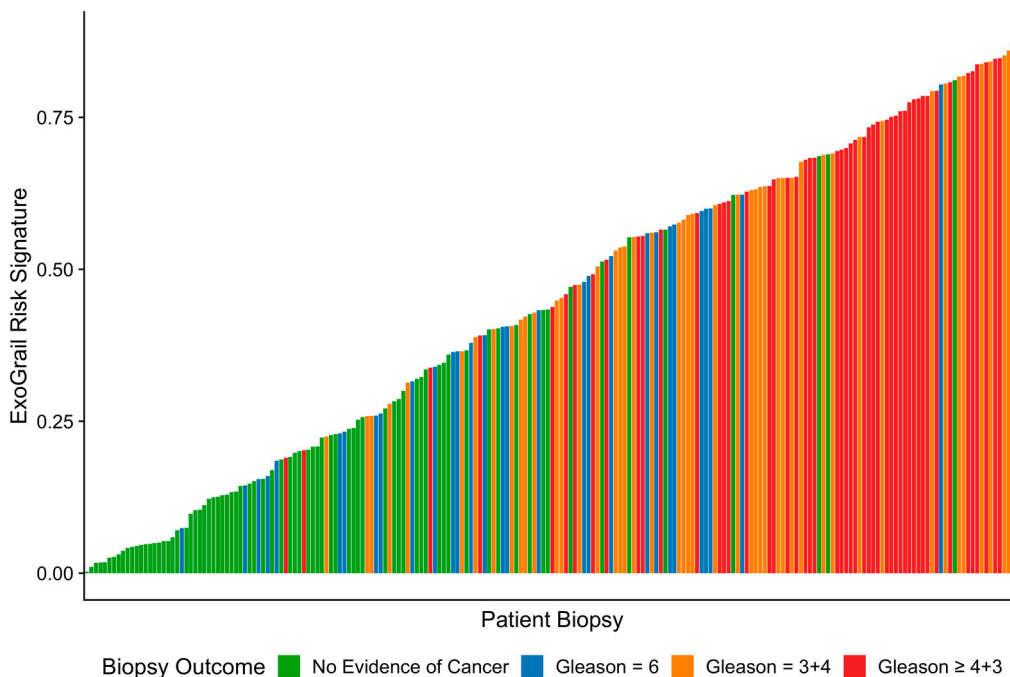


Figure 2. Representation of the ExoGrail risk score for each patient within a waterfall plot, where each coloured bar represents an individual’s biopsy outcome (fill colour) and predicted ExoGrail risk score (bar height). Green—No evidence of cancer, Blue—Gs 6, Orange—Gs 3 + 4, Red—Gs $\geq 4 + 3$.

ExoGrail was superior to all other models for the detection of Gleason $\geq 3 + 4$ (AUC = 0.90 (95% CI: 0.86–0.94), $p < 0.001$, bootstrap test with 1000 resamples) and for any cancer (AUC = 0.89 (95% CI: 0.85–0.94), $p < 0.001$, bootstrap test with 1000 resamples) (Table 3). When Gleason $\geq 4 + 3$ was considered, ExoGrail returned an AUC = 0.84 (95% CI: 0.78–0.89), outperforming the SoC and cf-RNA models ($p < 0.001$, bootstrap test with 1000 resamples), whilst the Engrailed model displayed similar performance by AUC metrics (Table 3). A model consisting of the combination of EN2 and PSA showed a similar ability in the detection of Gleason $\geq 4 + 3$ compared to ExoGrail (AUCs of 0.85 compared to 0.84), whilst ExoGrail showed a small improvement in the detection of Gleason $\geq 3 + 4$ disease and any cancer (Table S2).

Table 3. Area under the receiver operator curve (AUC) of all trained models for detecting outcomes of an initial biopsy for varying clinically significant thresholds. Numbers within brackets detail 95% confidence intervals of the AUC, calculated from 1000 stratified bootstrap resamples. Input variables for each model are detailed in Table 1.

| Initial Biopsy Outcome: | SoC | Engrailed | ExoRNA | ExoGrail |
|-------------------------|------------------|------------------|------------------|------------------|
| Gleason $\geq 4 + 3$: | 0.77 (0.69–0.84) | 0.81 (0.74–0.88) | 0.67 (0.59–0.75) | 0.84 (0.78–0.89) |
| Gleason $\geq 3 + 4$: | 0.72 (0.65–0.79) | 0.83 (0.77–0.88) | 0.77 (0.70–0.83) | 0.90 (0.86–0.94) |
| Any Cancer | 0.75 (0.68–0.82) | 0.81 (0.74–0.86) | 0.81 (0.74–0.87) | 0.89 (0.85–0.94) |

As revealed by the distributions of risk scores and AUC, ExoGrail achieved clearer discrimination of disease status Gleason $\geq 3 + 4$ disease from other biopsy outcomes when compared to any of the other models (ExoGrail all comparisons $p < 0.01$ bootstrap test, 1000 resamples, Figure 3).

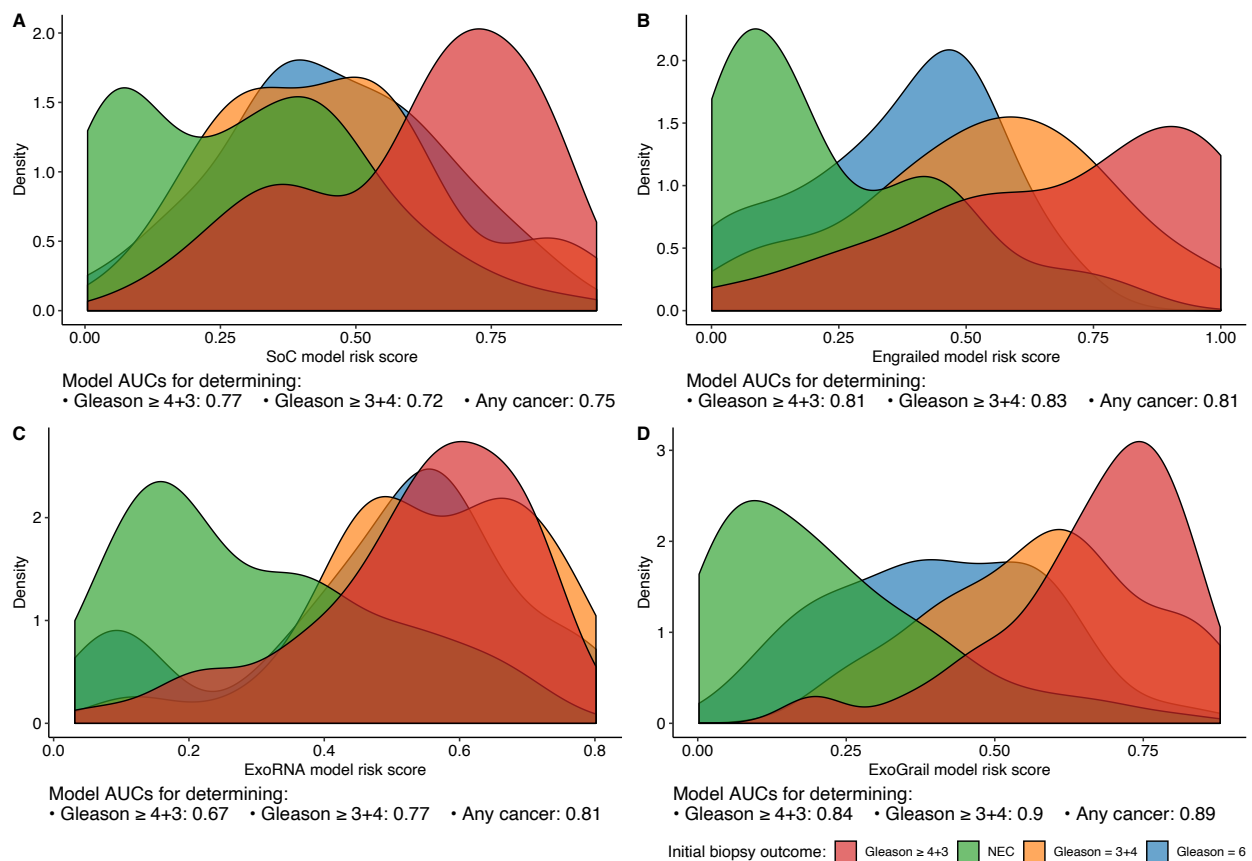


Figure 3. Risk score distributions of the four trained models, calculated as the out-of-bag predictions and represented as density plots. Area under the receiver operator curves (AUCs) for each model's predictive ability for clinically relevant outcomes are detailed underneath each panel. Each random forest model was fit using different input variables; (A) SoC clinical risk model, including Age and serum prostate specific antigen (PSA), (B) Engrailed model, (C) ExoRNA model and (D) ExoGrail model, combining predictors from all three modes of analysis. The full list of variables in each model is available in Table 1. Fill colour shows the risk score distribution of patients with respect to biopsy outcome: No evidence of cancer (Green), Gleason 6 (Blue), Gleason 3 + 4 (Orange), Gleason $\geq 4 + 3$ (Red).

Investigation of risk score distributions found that whilst the SoC model returned respectable AUCs and detection of the higher grade disease (Gleason $\geq 3 + 4$), it displayed a relative inability to clearly stratify intermediate disease states. This uncertainty would cause large numbers of patients to be inappropriately selected for further investigation (Figure 3A). For example, to classify 90% of patients with Gleason 7 disease correctly, an SoC risk score of 0.251 would misclassify 64.5% of men with less significant, or no disease. The Engrailed model detailed clearer discrimination, though featured a bimodal distribution of patients without prostate cancer (Figure 3B, green density plot), misidentifying 51.4% of patients with low-grade disease as similar to those with more clinically significant disease (Figure 3B). A similar bimodal distribution was seen for the EN2 plus PSA model (Figure S5). Whilst the AUCs returned for the ExoRNA model were lower, the distribution of risk scores shows that ExoRNA could more accurately discriminate cancer from non-cancer than either the SoC or EN2 models, a key clinical step in the triage of patients prior to biopsy (Figure 3C).

Examination of ExoGrail scores displayed similar distributions for NEC patients as the ExoRNA model whilst also being able to more accurately separate different cancer outcomes from biopsy, resulting in fewer misclassifications of patients without cancer if binary detection of 95% of Gleason $\geq 3 + 4$ were considered (28% of NEC patients misclassified). The greater discriminatory ability of the ExoGrail model when biopsy outcomes are considered as a binary Gleason $\geq 3 + 4$ threshold can also be seen in Figure S6.

Comparisons of mean ExoGrail scores between groups were performed with re-sampling and Cumming estimation plots (1000 bias-corrected and accelerated bootstrap resamples, Figure 4). The mean ExoGrail differences between patients with no evidence of cancer on biopsy were: Gleason 6 = 0.3 (95% CI: 0.22–0.37), Gleason 3 + 4 = 0.48 (95% CI: 0.41–0.53) and Gleason $\geq 4 + 3$ = 0.56 (95% CI: 0.51–0.61). Of note, patients with no evidence of cancer had a lower ExoGrail risk score (mean difference = 0.17 (95% CI: 0.11–0.24)) than those with a raised PSA but no findings of cancer on biopsy (Figure 4).

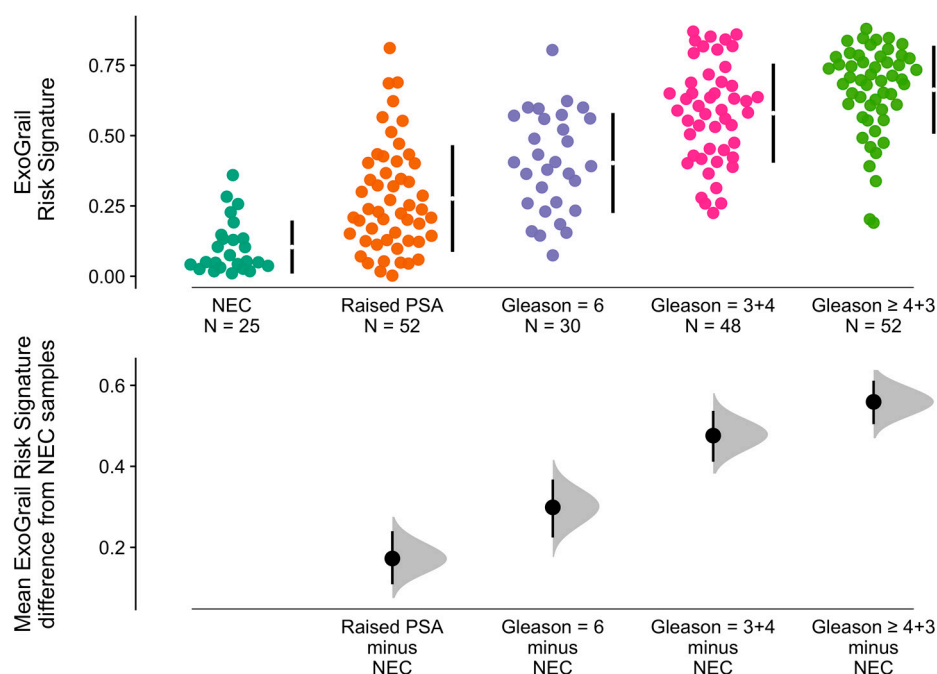


Figure 4. Mean ExoGrail risk score differences between biopsy outcomes, as represented by Cumming estimation plots. Individual patient risk scores (y-axis) are presented as points in the top panel, separated according to Gleason score (x-axis) with gapped vertical lines detailing the mean and standard deviation of each clinical group's ExoGrail risk score. Mean ExoGrail risk score differences relative to the no evidence of cancer (NEC) group are shown in the bottom panel. Mean difference and 95% confidence intervals are shown as a point estimate and vertical bar, respectively, with density plots generated from 1000 bias-corrected and accelerated bootstrap resamples.

Decision curve analyses examined the net benefit of ExoGrail adoption in a population of patients with a clinical suspicion of prostate cancer and a PSA level suitable to trigger biopsy (≥ 4 ng/mL). The biopsy of men based upon their ExoGrail risk score provided a net benefit over current standards of care across all decision thresholds examined and was the most consistent amongst all comparator models across a range of clinically relevant endpoints for biopsy (Figure 5).

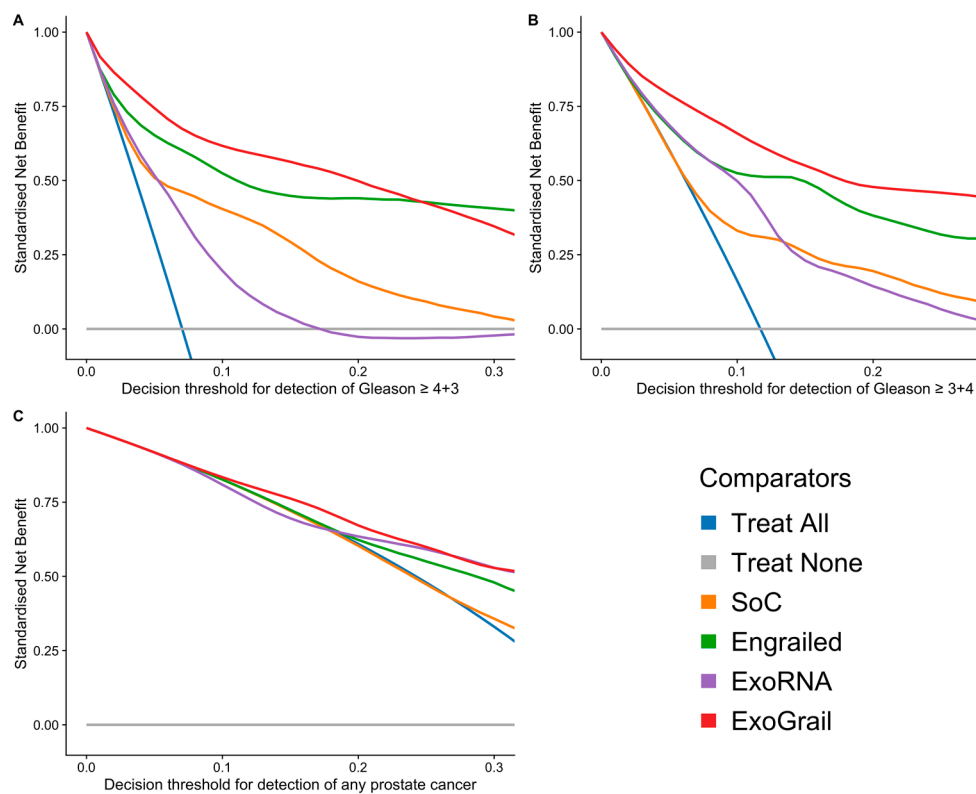


Figure 5. Exploration of the standardised net benefit (sNB) by decision curve analysis (DCA) for adopting risk models to aid the decision to undertake an initial biopsy for patients presenting with a serum PSA ≥ 4 ng/mL, where current clinical practice is to biopsy all patients. The accepted patient/clinician risk threshold for accepting biopsy is detailed on the x-axis. Different biopsy outcomes are shown in each of the three panels; (A) detection of Gleason $\geq 4 + 3$, (B) detection of Gleason $\geq 3 + 4$, (C) any cancer; Blue—biopsy all patients with a PSA >4 ng/mL, Orange—biopsy patients according to the SoC model, Green—biopsy patients based on the Engrailed model, Purple—biopsy patients based on the exoRNA model, Red—biopsy patients based on a the ExoGrail model. To assess the benefit of adopting these risk models in a clinically relevant population, we used data available from the control arm of the The Cluster Randomized Trial of PSA Testing for Prostate Cancer (CAP) study [42] for proportionally resampling the ExoGrail cohort. DCA curves were calculated from 1000 bootstrap resamples of the available data to match the distribution of disease reported in the CAP trial population. Mean sNB from these resampled DCA results are plotted here. See Methods for full details.

Using the SoC model as the baseline with which to compare the potential for biopsy reduction of each model, we found that ExoGrail could reduce unnecessary biopsy rates by upwards of 40%, depending on accepted patient–clinician risk. For example, if a decision threshold of 0.1 were accepted, representing a perceived risk of 1 in 10 for Gleason $\geq 3 + 4$ on biopsy, ExoGrail could result in up to a 35% reduction in unnecessary biopsies of men presenting with a suspicion of prostate cancer, whilst also correctly identifying patients with more aggressive disease. If Gleason $\geq 4 + 3$ were considered the threshold of clinical significance, a more conservative decision threshold of 0.05 could save 32% of men from receiving an unnecessary biopsy (Figure 6).

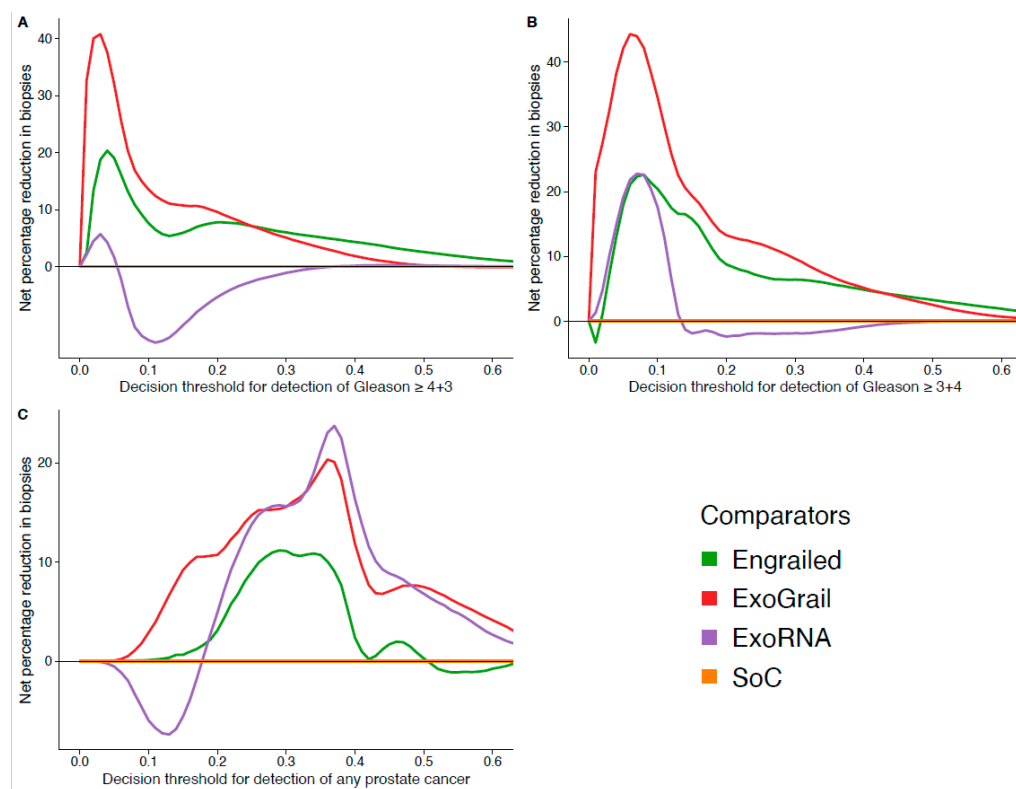


Figure 6. Estimation of biopsy reduction, as calculated by comparing the DCA-calculated net benefit of each risk model to the net benefit of the standard of care (SoC) model. The accepted patient/clinician risk threshold for accepting biopsy is detailed on the x-axis. Different biopsy outcomes are shown in each of the three panels: (A) detection of Gleason $\geq 4 + 3$, (B) detection of Gleason $\geq 3 + 4$ and (C) any cancer. Coloured lines show differing comparator models; Orange—biopsy patients according to the SoC model, Green—biopsy patients based on the Engrailed model, Purple—biopsy patients based on the ExoRNA model, Red—biopsy patients based on the ExoGrail model. To assess the benefit of adopting these risk models in a clinically relevant population we used data available from the control arm of the CAP study [42] for proportionally resampling the ExoGrail cohort. DCA curves were calculated from 1000 bootstrap resamples of the available data to match the distribution of disease reported in the CAP trial population. Net benefit averaged over all resamples was used to calculate the potential reductions in biopsy rates here. See Methods for full details.

4. Discussion

Discriminating disease status in patients before a diagnostic biopsy with higher accuracy than current standards could bring about a sizeable change in treatment pathways and reduce the number of men sent forward for ultimately unnecessary biopsy. Given that up to 75% of patients are negative for prostate cancer when presenting with serum PSA levels ≥ 4 ng/mL [5,43,44], a concentration of research efforts has been made to address this problem. To date, several biomarker panels have been successfully developed to non-invasively detect prostate cancer using urine samples, Gleason $\geq 3 + 4$ disease with superior accuracy to current clinically implemented methods, including the PUR model developed by ourselves [20,21,45,46]. However, as only a single aspect of urine, assay method or biological process are assessed by these examples, the heterogeneity of prostate cancer may not be entirely accounted for [47], requiring an approach to be taken that provides a more holistic insight into disease status.

Recent analyses, including those presented here, have demonstrated the added value of integrating multiple prognostic biomarkers within the process of fitting risk models for determining patient risk upon an initial biopsy [23,48]. Urine clearly contains a wealth of useful information concerning the disease status of the prostate through the quantification

of cf-RNA transcripts, circulating and cell-free DNA, hypermethylation of DNA, and protein biomarker levels [19,46,49–52].

Our results show that an improved multivariable risk prediction model can be developed from the careful consideration of information from multiple different urine fractions in men suspected to have prostate cancer. Urinary levels of EN2 protein were quantified by ELISA, whilst the transcript levels of 167 cell-free mRNAs were quantified using NanoString technology. The final model integrating information from those assays with serum PSA levels was deemed ExoGrail. Markers selected for the model include well-known genes associated with prostate cancer and proven in other diagnostic tests, such as *PCA3* [45], *HOXC6* [20], and the *TMPRSS2/ERG* gene fusion [53]. An interaction between urinary EN2 protein levels and quantified transcripts of *SLC12A1* was observed, further demonstrating the benefit of considering information from multiple biological sources (Figure S4).

ExoGrail was able to accurately predict the presence of significant ($G_s \geq 7$) prostate cancer on biopsy with an AUC of 0.89, comparing favourably to other published tests (AUCs for $G_s \geq 7$: PUR = 0.77 [46], ExoMeth = 0.89 [23], ExoDX Prostate IntelliScore = 0.77 [21], SelectMDX = 0.78 [20], epiCaPture $G_s \geq 4 + 3$ AUC = 0.73 [49]). Furthermore, ExoGrail resulted in accurate predictions even when serum PSA levels alone proved inaccurate; patients with a raised PSA but negative biopsy result possessed ExoGrail scores significantly different from both clinically benign patients and those with low-grade Gleason 6 disease, whilst still able to discriminate between more clinically significant Gleason ≥ 7 cancers (Figure 4). The adoption of ExoGrail into current clinical pathways for reducing unnecessary biopsies was considered, showing the potential for up to 32% of patients to safely forgo an invasive biopsy without incurring excessive risk (Figure 6).

ExoGrail was developed with the explicit goal of being robust to potential overfitting and bias, using strong internal validation methods in bootstrap resampling and out-of-bag predictions. Nonetheless, ExoGrail was developed in a relatively small dataset and so requires external validation in an independent cohort before it can be considered for use as a clinical risk model. To this end, we are currently collecting samples from multiple sites in the UK, EU and Canada using an updated 'At-Home' Collection system [54]. The At-Home collection system enables biomarker analysis to be performed on urine samples provided by patients at home, which they send in the post to a centralised laboratory. This collection and analysis system will sidestep the need for a visit to the clinic and lead to a postal screening system for prostate cancer diagnosis and prognosis. In this study, we will also assess the potential utility of supplementing MP-MRI with ExoGrail, as MP-MRI can misrepresent disease status, even with rigorous controls in place [6]. The NanoString expression analysis system used in the ExoGrail signature is a rapid and cost-effective analysis system that is also used in the FDA-approved Prosigna Pam50 test for breast cancer aggressiveness [55], making ExoGrail well-positioned for implementation for patient benefit.

5. Conclusions

ExoGrail was able to accurately predict the presence of significant ($G_s \geq 7$) prostate cancer on biopsy and showed the potential for an important number of patients to safely forgo an invasive biopsy. If validated in future studies, ExoGrail has the potential to positively impact the clinical experience of patients being investigated for prostate cancer that ultimately have no disease or indolent prostate cancer.

6. Patents

A patent application has been filed by the authors for the present work and work related to this.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/cancers13092102/s1>, Figure S1: Boruta analysis of variables available for the training of the SoC model, Figure S2: Boruta analysis of variables available for the training of the Engrailed model, Figure S3: Boruta analysis of variables available for the training of the ExoRNA model, Figure S4:

Partial dependency plots detailing the marginal effects and interactions of SLC12A1 and urinary EN2 on predicted ExoGrail Risk Score, Figure S5: Risk score distributions of the trained models including the EN2 and PSA model, Figure S6: Density plots detailing risk score distributions generated from four trained models. Table S1: List of all features available for selection as input variables for each model prior to bootstrapped Boruta feature selection, Table S2: AUC of all trained models, including a combination of EN2 and PSA, for detecting outcomes of an initial biopsy for varying clinically significant thresholds. Supplementary Methods.

Author Contributions: S.P.C. and D.S.B. drafted the manuscript and conceived, designed, and performed the statistical analyses. H.P. and R.M. (Richard Morgan) were involved in sample collection and ELISA analyses at their respective institutes. J.C. and R.M. were involved in sample collection and NanoString analyses, as well as the development of clinical methodologies. D.S.B., J.C., R.M. (Robert Mills), H.P., and C.S.C. had joint and equal contributions to senior authorship and were contributors in writing the manuscript. C.S.C., J.C., H.P. and D.S.B. provided the original idea for this study. All authors have read and agreed to the published version of the manuscript. All authors critiqued the manuscript for intellectual content.

Funding: This research was funded by 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.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approval was granted for the collection and processing of samples by the Ethics Committees at the East of England REC.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data and code required to reproduce these analyses can be found at https://github.com/UEA-Cancer-Genetics-Lab/ExoGrail_paper.

Acknowledgments: The research presented in this paper was carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia.

Conflicts of Interest: A patent application has been filed by the authors for the present work and work related to this. There are no other conflicts of interest to disclose.

References

1. Cancer Research UK Prostate Cancer Incidence Statistics. 2019. Available online: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence> (accessed on 29 June 2019).
2. Loeb, S.; Bjurlin, M.A.; Nicholson, J.; Tammela, T.L.; Penson, D.F.; Carter, H.B.; Carroll, P.; Etzioni, R. Overdiagnosis and Overtreatment of Prostate Cancer. *Eur. Urol.* **2014**, *65*, 1046–1055. [[CrossRef](#)]
3. Sanda, M.G.; Cadeddu, J.A.; Kirkby, E.; Chen, R.C.; Crispino, T.; Fontanarosa, J.; Freedland, S.J.; Greene, K.; Klotz, L.H.; Makarov, D.V.; et al. Clinically Localized Prostate Cancer: AUA/ASTRO/SUO Guideline. Part I: Risk Stratification, Shared Decision Making, and Care Options. *J. Urol.* **2018**, *199*, 683–690. [[CrossRef](#)]
4. Cornford, P.; Bellmunt, J.; Bolla, M.; Briers, E.; De Santis, M.; Gross, T.; Henry, A.M.; Joniau, S.; Lam, T.B.; Mason, M.D.; et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part II: Treatment of Relapsing, Metastatic, and Castration-Resistant Prostate Cancer. *Eur. Urol.* **2017**, *71*, 630–642. [[CrossRef](#)]
5. National Institute for Health and Care Excellence. *Prostate Cancer: Diagnosis and Management (Update)*; NICE: London, UK, 2015.
6. Ahmed, H.U.; Bosaily, A.E.-S.; Brown, L.C.; Gabe, R.; Kaplan, R.; Parmar, M.K.; Collaco-Moraes, Y.; Ward, K.; Hindley, R.G.; Freeman, A.; et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): A paired validating confirmatory study. *Lancet* **2017**, *389*, 815–822. [[CrossRef](#)]
7. Pepe, P.; Pennisi, M. Gleason score stratification according to age at diagnosis in 1028 men. *Współczesna Onkol.* **2015**, *19*, 471–473. [[CrossRef](#)]
8. Sonn, G.A.; Fan, R.E.; Ghanouni, P.; Wang, N.N.; Brooks, J.D.; Loening, A.M.; Daniel, B.L.; To'O, K.J.; Thong, A.E.; Leppert, J.T. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. *Eur. Urol. Focus* **2019**, *5*, 592–599. [[CrossRef](#)] [[PubMed](#)]
9. Walz, J. The “PROMIS” of Magnetic Resonance Imaging Cost Effectiveness in Prostate Cancer Diagnosis? *Eur. Urol.* **2018**, *73*, 31–32. [[CrossRef](#)] [[PubMed](#)]
10. Moschini, M.; Spahn, M.; Mattei, A.; Cheville, J.; Karnes, R.J. Incorporation of tissue-based genomic biomarkers into localized prostate cancer clinics. *BMC Med.* **2016**, *14*, 67. [[CrossRef](#)] [[PubMed](#)]

11. Luca, B.-A.; Brewer, D.S.; Edwards, D.R.; Edwards, S.; Whitaker, H.C.; Merson, S.; Dennis, N.; Cooper, R.A.; Hazell, S.; Warren, A.Y.; et al. DESNT: A Poor Prognosis Category of Human Prostate Cancer. *Eur. Urol. Focus* **2018**, *4*, 842–850. [[CrossRef](#)]
12. Knezevic, D.; Goddard, A.D.; Natraj, N.; Cherbavaz, D.B.; Clark-Langone, K.M.; Snable, J.; Watson, D.; Falzarano, S.M.; Magi-Galluzzi, C.; Klein, E.A.; et al. Analytical validation of the Oncotype DX prostate cancer assay—A clinical RT-PCR assay optimized for prostate needle biopsies. *BMC Genom.* **2013**, *14*, 690. [[CrossRef](#)] [[PubMed](#)]
13. Cuzick, J.; Berney, D.M.; Fisher, G.J.; Mesher, D.; Moller, H.; Reid, J.; Perry, M.B.A.; Park, J.; Younus, A.; on behalf of the Transatlantic Prostate Group; 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*, 1095–1099. [[CrossRef](#)]
14. Luca, B.-A.; Moulton, V.; Ellis, C.; Edwards, D.R.; Campbell, C.; Cooper, R.A.; Clark, J.; Brewer, D.S.; Cooper, C.S. A novel stratification framework for predicting outcome in patients with prostate cancer. *Br. J. Cancer* **2020**, *122*, 1467–1476. [[CrossRef](#)] [[PubMed](#)]
15. Cooperberg, M.R.; Davicioni, E.; Crisan, A.; Jenkins, R.B.; Ghadessi, M.; Karnes, R.J. Combined Value of Validated Clinical and Genomic Risk Stratification Tools for Predicting Prostate Cancer Mortality in a High-risk Prostatectomy Cohort. *Eur. Urol.* **2015**, *67*, 326–333. [[CrossRef](#)] [[PubMed](#)]
16. Eklund, M.; Nordström, T.; Aly, M.; Adolfsson, J.; Wiklund, P.; Brandberg, Y.; Thompson, J.; Wiklund, F.; Lindberg, J.; Presti, J.C.; et al. The Stockholm-3 (STHLM3) Model can Improve Prostate Cancer Diagnostics in Men Aged 50–69 yr Compared with Current Prostate Cancer Testing. *Eur. Urol. Focus* **2018**, *4*, 707–710. [[CrossRef](#)] [[PubMed](#)]
17. Tosoian, J.J.; Carter, H.B.; Lepor, A.; Loeb, S. Active surveillance for prostate cancer: Current evidence and contemporary state of practice. *Nat. Rev. Urol.* **2016**, *13*, 205–215. [[CrossRef](#)] [[PubMed](#)]
18. Frick, J.; Aulitzky, W. Physiology of the prostate. *Infection* **1991**, *19* (Suppl. 3), S115–S118. [[CrossRef](#)] [[PubMed](#)]
19. Morgan, R.; Boxall, A.; Bhatt, A.; Bailey, M.; Hindley, R.; Langley, S.; Whitaker, H.C.; Neal, D.E.; Ismail, M.; Whitaker, H.; et al. Engrailed-2 (EN2): A Tumor Specific Urinary Biomarker for the Early Diagnosis of Prostate Cancer. *Clin. Cancer Res.* **2011**, *17*, 1090–1098. [[CrossRef](#)]
20. Van Neste, L.; Hendriks, R.J.; Dijkstra, S.; Trooskens, G.; Cornel, E.B.; Jannink, S.A.; de Jong, H.; Hessels, D.; Smit, F.P.; Melchers, W.J.; et al. Detection of High-grade Prostate Cancer Using a Urinary Molecular Biomarker-Based Risk Score. *Eur. Urol.* **2016**, *70*, 740–748. [[CrossRef](#)] [[PubMed](#)]
21. McKiernan, J.; Donovan, M.J.; O'Neill, V.; Bentink, S.; Noerholm, M.; Belzer, S.; Skog, J.; Kattan, M.W.; Partin, A.; Andriole, G.; et al. A Novel Urine Exosome Gene Expression Assay to Predict High-grade Prostate Cancer at Initial Biopsy. *JAMA Oncol.* **2016**, *2*, 882–889. [[CrossRef](#)] [[PubMed](#)]
22. Haese, A.; Trooskens, G.; Steyaert, S.; Hessels, D.; Brawer, M.; Vlaeminck-Guillem, V.; Ruffion, A.; Tilki, D.; Schalken, J.; Groskopf, J.; et al. Multicenter Optimization and Validation of a 2-Gene mRNA Urine Test for Detection of Clinically Significant Prostate Cancer before Initial Prostate Biopsy. *J. Urol.* **2019**, *202*, 256–263. [[CrossRef](#)]
23. Connell, S.P.; O'Reilly, E.; Tuzova, A.; Webb, M.; Hurst, R.; Mills, R.; Zhao, F.; Bapat, B.; Cooper, C.S.; Perry, A.S.; et al. Development of a multivariable risk model integrating urinary cell DNA methylation and cell-free RNA data for the detection of significant prostate cancer. *Prostate* **2020**, *80*, 547–558. [[CrossRef](#)] [[PubMed](#)]
24. Morgan, R. Engrailed: Complexity and economy of a multi-functional transcription factor. *FEBS Lett.* **2006**, *580*, 2531–2533. [[CrossRef](#)]
25. Punia, N.; Primon, M.; Simpson, G.R.; Pandha, H.S.; Morgan, R. Membrane insertion and secretion of the Engrailed-2 (EN2) transcription factor by prostate cancer cells may induce antiviral activity in the stroma. *Sci. Rep.* **2019**, *9*, 5138. [[CrossRef](#)] [[PubMed](#)]
26. Pandha, H.; Sorensen, K.D.; Orntoft, T.F.; Langley, S.; Hoyer, S.; Borre, M.; Morgan, R. Urinary engrailed-2 (EN2) levels predict tumour volume in men undergoing radical prostatectomy for prostate cancer. *BJU Int.* **2012**, *110*, E287–E292. [[CrossRef](#)] [[PubMed](#)]
27. Elamin, A.A.; Klunkelfuß, S.; Kämpfer, S.; Oehlmann, W.; Stehr, M.; Smith, C.; Simpson, G.R.; Morgan, R.; Pandha, H.; Singh, M. A Specific Blood Signature Reveals Higher Levels of S100A12: A Potential Bladder Cancer Diagnostic Biomarker Along with Urinary Engrailed-2 Protein Detection. *Front. Oncol.* **2020**, *9*, 1484. [[CrossRef](#)]
28. Stark, J.R.; Perner, S.; Stampfer, M.J.; Sinnott, J.A.; Finn, S.; Eisenstein, A.S.; Ma, J.; Fiorentino, M.; Kurth, T.; Loda, M.; et al. Gleason Score and Lethal Prostate Cancer: Does $3 + 4 = 4 + 3$? *J. Clin. Oncol.* **2009**, *27*, 3459–3464. [[CrossRef](#)]
29. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Eur. Urol.* **2015**, *67*, 1142–1151. [[CrossRef](#)]
30. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2019.
31. UEA Cancer Genetic GitHub Repository. Available online: <https://github.com/UEA-Cancer-Genetics-Lab/ExoGrail> (accessed on 21 April 2021).
32. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
33. Kursu, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13. [[CrossRef](#)]
34. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
35. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
36. Robin, X.A.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M.J. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [[CrossRef](#)] [[PubMed](#)]
37. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; ISBN 978-3-319-24277-4.

38. Greenwell, B.M. Pdp: An r Package for Constructing Partial Dependence Plots. *R J.* **2017**, *9*, 421–436. [[CrossRef](#)]
39. Ho, J.; Tumkaya, T.; Aryal, S.; Choi, H.; Claridge-Chang, A. Moving beyond P values: Data analysis with estimation graphics. *Nat. Methods* **2019**, *16*, 565–566. [[CrossRef](#)] [[PubMed](#)]
40. Vickers, A.J.; Elkin, E.B. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med. Decis. Mak.* **2006**, *26*, 565–574. [[CrossRef](#)] [[PubMed](#)]
41. Brown, M. *rmda: Risk Model Decision Analysis*; Fred Hutchinson Cancer Research Center: Seattle, WA, USA, 2018.
42. Kerr, K.F.; Brown, M.D.; 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*, 2534–2540. [[CrossRef](#)]
43. Martin, R.M.; Donovan, J.L.; Turner, E.L.; Metcalfe, C.; Young, G.J.; Walsh, E.I.; Lane, J.A.; Noble, S.; Oliver, S.E.; Evans, S.; et al. Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality: The CAP randomized clinical trial. *JAMA* **2018**, *319*, 883–895. [[CrossRef](#)]
44. Lane, J.A.; Donovan, J.L.; Davis, M.; Walsh, E.; Dedman, D.; Down, L.; Turner, E.L.; Mason, M.D.; Metcalfe, C.; Peters, T.J.; 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*, 1109–1118. [[CrossRef](#)]
45. Hessels, D.; Gunnewiek, J.M.K.; van Oort, I.; Karthaus, H.F.; van Leenders, G.J.; van Balken, B.; Kiemeny, L.A.; Witjes, J.; Schalken, J.A. DD3PCA3-based Molecular Urine Analysis for the Diagnosis of Prostate Cancer. *Eur. Urol.* **2003**, *44*, 8–16. [[CrossRef](#)]
46. Connell, S.P.; Yazbek-Hanna, M.; McCarthy, F.; Hurst, R.; Webb, M.; Curley, H.; Walker, H.; Mills, R.; Ball, R.Y.; Sanda, M.G.; et al. A four-group urine risk classifier for predicting outcomes in patients with prostate cancer. *BJU Int.* **2019**, *124*, 609–620. [[CrossRef](#)]
47. Ciccamese, C.; Massari, F.; Iacovelli, R.; Fiorentino, M.; Montironi, R.; Di Nunno, V.; Giunchi, F.; Brunelli, M.; Tortora, G. Prostate cancer heterogeneity: Discovering novel molecular targets for therapy. *Cancer Treat. Rev.* **2017**, *54*, 68–73. [[CrossRef](#)]
48. Strand, S.H.; Bavafaye-Haghighi, E.; Kristensen, H.; Rasmussen, A.K.; Hoyer, S.; Borre, M.; Mouritzen, P.; Besenbacher, S.; Orntoft, T.F.; Sorensen, K.D. A novel combined miRNA and methylation marker panel (miMe) for prediction of prostate cancer outcome after radical prostatectomy. *Int. J. Cancer* **2019**, *145*, 3445–3452. [[CrossRef](#)]
49. O'Reilly, E.; Tuzova, A.V.; Walsh, A.L.; Russell, N.M.; O'Brien, O.; Kelly, S.; Ni Dhomhnaillain, O.; DeBarra, L.; Dale, C.M.; Brugman, R.; et al. epiCaPture: A Urine DNA Methylation Test for Early Detection of Aggressive Prostate Cancer. *JCO Precis. Oncol.* **2019**, *2019*, 1–18. [[CrossRef](#)]
50. Zhao, F.; Olkhov-Mitsel, E.; Kamdar, S.; Jeyapala, R.; Garcia, J.; Hurst, R.; Hanna, M.Y.; Mills, R.; Tuzova, A.V.; O'Reilly, E.; et al. A urine-based DNA methylation assay, ProCUrE, to identify clinically significant prostate cancer. *Clin. Epigenet.* **2018**, *10*, 147. [[CrossRef](#)]
51. Xia, Y.; Huang, C.-C.; Dittmar, R.; Du, M.; Wang, Y.; Liu, H.; Shenoy, N.; Wang, L.; Kohli, M. Copy number variations in urine cell free DNA as biomarkers in advanced prostate cancer. *Oncotarget* **2016**, *7*, 35818–35831. [[CrossRef](#)]
52. Killick, E.; Morgan, R.; Launchbury, F.; Bancroft, E.; Page, E.; Castro, E.; Kote-Jarai, Z.; Aprikian, A.; Blanco, I.; Clowes, V.; et al. Role of Engrailed-2 (EN2) as a prostate cancer detection biomarker in genetically high risk men. *Sci. Rep.* **2013**, *3*, 2059. [[CrossRef](#)]
53. Tomlins, S.A.; Day, J.R.; Lonigro, R.J.; Hovelson, D.H.; Siddiqui, J.; Kunju, L.P.; Dunn, R.L.; Meyer, S.; Hodge, P.; Grosskopf, J.; et al. Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment. *Eur. Urol.* **2016**, *70*, 45–53. [[CrossRef](#)] [[PubMed](#)]
54. Webb, M.; Manley, K.; Olivan, M.; Guldvik, I.; Palczynska, M.; Hurst, R.; Connell, S.P.; Mills, I.G.; Brewer, D.S.; Mills, R.; et al. Methodology for the at-home collection of urine samples for prostate cancer detection. *Biotechniques* **2020**, *68*, 65–71. [[CrossRef](#)] [[PubMed](#)]
55. Wallden, B.; Storhoff, J.; Nielsen, T.; Dowidar, N.; Schaper, C.; Ferree, S.; Liu, S.; Leung, S.; Geiss, G.; Snider, J.; et al. Development and Verification of the PAM50-Based Prosigna Breast Cancer Gene Signature Assay. *BMC Med. Genom.* **2015**, *8*, 54. [[CrossRef](#)] [[PubMed](#)]