Non-coding RNA Research 9 (2024) 612–623

Contents lists available at ScienceDirect

Non-coding RNA Research

journal homepage: www.keaipublishing.com/en/journals/non-coding-rna-research

Predicting prostate cancer progression with a Multi-lncRNA expression-based risk score and nomogram integrating ISUP grading

Sabrina Ledesma-Bazan [a,b], Florencia Cascardo [a,b], Juan Bizzotto [a,b,c], Santiago Olszevicki [a,b], Elba Vazquez [a,b], Geraldine Gueron [a,b], Javier Cotignola [a,b,*]

[a] Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Química Biológica, Laboratorio de Inflamación y Cáncer, C1428EGA, CABA, Buenos Aires, Argentina
[b] CONICET - Universidad de Buenos Aires, Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), C1428EGA, CABA, Buenos Aires, Argentina
[c] Universidad Argentina de la Empresa (UADE), Instituto de Tecnología (INTEC), Buenos Aires C1073AAO, Argentina

A B S T R A C T

Prostate cancer is a highly heterogeneous disease; therefore, estimating patient prognosis accurately is challenging due to the lack of biomarkers with sufficient specificity and sensitivity. One of the current challenges lies in integrating genomic and transcriptomic data with clinico-pathological features and in incorporating their application in everyday clinical practice. Therefore, we aimed to model a risk score and nomogram containing long non-coding RNA (lncRNA) expression and clinico-pathological data to better predict the probability of prostate cancer progression. We performed bioinformatics analyses to identify lncRNAs differentially expressed across various prostate cancer stages and associated with progression-free survival. This information was further integrated into a prognostic risk score and nomogram containing transcriptomic and clinico-pathological features to estimate the risk of disease progression. We used RNA-seq data from 5 datasets from public repositories (total n = 178) comprising different stages of prostate cancer: pre-treatment primary prostate adenocarcinomas, post-treatment tumors and metastatic castration resistant prostate cancer. We found 30 lncRNAs with consistent differential expression in all comparisons made using two R-based packages. Multivariate progression-free survival analysis including the ISUP group as covariate, revealed that 7/30 lncRNAs were significantly associated with time-to-progression. Next, we combined the expression of these 7 lncRNAs into a multi-lncRNA score and dichotomized the patients into low- or high-score. Patients with a high-score showed a 4-fold risk of disease progression (HR = 4.30, 95 %CI = 2.66–6.97, p = 3.1e-9). Furthermore, we modelled a combined risk-score containing information on the multi-lncRNA score and ISUP group. We found that patients with a high-risk score had nearly 8-fold risk of progression (HR = 7.65, 95 %CI = 4.05–14.44, p = 3.4e-10). Finally, we created and validated a nomogram to help uro-oncologists to better predict patient's risk of progression at 3- and 5-years post-diagnosis. In conclusion, the integration of lncRNA expression data and clinico-pathological features of prostate tumors into predictive models might aid in tailored disease risk assessment and treatment for patients with prostate cancer.

1. Introduction

Prostate cancer is the second most incident cancer and the fifth leading cause of cancer-related deaths among men worldwide [1]. The course of the disease is aggressive and life-threatening for some patients, while others have an indolent tumor with a low risk of progression. In standard clinical practice, several clinico-pathological characteristics are used to predict patient outcomes, which include: Gleason score, clinical and pathological tumor stage (TNM), serum PSA (Prostatic Specific Antigen) and the groups defined by the International Society of Urological Pathology (ISUP), among others [2]. However, estimating patient prognosis accurately is challenging due to the high heterogeneity of prostate tumors [3] and the lack of biomarkers with sufficient specificity and sensitivity [4]. These limitations restrict the effectiveness in predicting different outcomes among patients with similar clinico-pathological characteristics. Hence, there is an urgent need to discover novel markers highly associated with disease progression and develop prognostic models to improve the risk stratification of patients with prostate cancer.

One promising field to discover new biomarkers involves studying changes in gene expression in the tumoral tissue. While expression analyses of individual genes (mainly protein-coding) have been extensively evaluated, they have not yet demonstrated sufficient specificity and sensitivity to be used as reliable biomarkers for prostate cancer diagnosis and progression [5]. Therefore, tumor whole transcriptomic information has been under consistent investigation in order to discover

* Corresponding author. CONICET - Universidad de Buenos Aires, Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), 1428, CABA, Buenos Aires, Argentina.
E-mail address: jcotignola@qb.fcen.uba.ar (J. Cotignola).

https://doi.org/10.1016/j.ncrna.2024.01.014
Received 28 August 2023; Received in revised form 11 January 2024; Accepted 23 January 2024
Available online 23 January 2024

gene-expression profiles that may define tumor subgroups with different outcomes, which may constitute a novel tool to improve tumor classification and to model gene-risk scores to predict disease progression [6–13]. To date, there are some RNA-based commercial kits that may predict prostate cancer outcomes and help to improve clinical decision-making. For example, *Oncotype Dx Genomic Prostate Score*® analyzes the expression of 17 genes and can predict an aggressive tumor behavior after radical prostatectomy [8,11]. *Decipher*® is another commercial kit that evaluates the expression of 22 genes to classify patients into risk groups of recurrence and distant metastases development after biopsy or radical prostatectomy [6,7,9]. However, there is still conflicting data and the predictive values are still uncertain; therefore, further validation is required and these tests are not widely used in daily clinical settings yet.

Long non-coding RNAs (lncRNAs) are RNAs longer than 200 nucleotides which do not encode for proteins. Historically, they were considered as by-products of the transcriptional process, but subsequent studies have changed this paradigm emphasizing their central role in cell biology. These lncRNAs are distributed throughout the genome and perform various cellular functions, including the regulation of gene transcription, splicing and post-translational modification, assembly of protein complexes acting as scaffold molecules, and sponge (capture) miRNAs to modulate their functions, among other activities [14,15]. Consequently, they can be considered as molecular hubs for the modulation of biochemical pathways. An increasing body of evidence has demonstrated aberrant expression of lncRNAs in various diseases, including cancer; and has highlighted the role of these molecules as key regulators of signaling pathways involved in tumor development and progression [16–19]. Due to their specific expression and biological functions, there is growing interest in these molecules as potential biomarkers. Several reports have investigated the utility of lncRNAs in the clinical setting, and proposed that they can serve as biomarkers in liquid biopsies for diagnosis, monitoring disease progression and therapeutic response [20–24].

One of the current challenges in oncology lies in integrating genomic and transcriptomic data with clinico-pathological features into mixed genetic and clinical models to enhance their individual prognostic values and, subsequently, incorporate their application in everyday clinical practice. Therefore, in this study, we aimed to model a risk score containing lncRNA expression and clinico-pathological data to better predict the probability of prostate cancer progression. We developed a prognostic score containing information based on a 7-lncRNA expression profile combined with the tumor ISUP group. In addition, with this information, we constructed a nomogram, which is a graphical scoring tool used in medicine to aid in predicting the likelihood of an individual to develop a specific event or outcome. Nomograms are particularly useful when complex statistical models and traits are involved, as they simplify the prediction process for healthcare professionals. Consequently, they allow for a more personalized approach to decision-making by considering multiple factors simultaneously.

2. Methods

2.1. Selection of datasets

We searched the public repositories Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) and European Bioinformatics Institute (EBI, https://www.ebi.ac.uk/) to identify transcriptomic studies that included tumor tissue samples from patients with different stages of prostate cancer. Our search included the following keywords: "prostate cancer" AND "RNA assay" OR "sequencing assay". We selected the following datasets:

- *GSE51005* [25]: 4 paired advanced prostate tumor biopsies (TRUS-guided) before and after treatment with androgen deprivation therapy (ADT) combined with docetaxel. Post-ADT samples were obtained ~22 weeks after treatment initiation and all biopsy cores had >60 % of tumoral cells.
- *GSE48403* [26]: 7 paired tumor biopsies (TRUS-guided) from locally advanced or metastatic prostate cancer before and after ADT-alone treatment. Post-ADT specimens were obtained ~22 weeks after treatment initiation and all biopsy cores had >60 % of tumoral cells.
- *GSE54460* [27]: 100 Formalin-Fixed Paraffin-Embedded radical prostatectomy samples from patients without hormonal or radiation treatment.
- *E-MTAB-6525* [28]: 98 paired RNA-seq of prostate biopsies from patients on active surveillance over 12 months on different diets (ESCAPE trial). We included only the 49 samples taken at baseline prior to the initiation of the special diets.
- *GSE31528* [29]: 8 snap frozen samples from patients with metastatic castration-resistant prostate cancer (mCRPC) obtained from the posterior iliac crest and with >90 % of tumoral cells.
- *TCGA-PRAD*: 498 primary prostate tumor samples. Pre-processed RNA-seq data was downloaded from https://portal.gdc.cancer.gov/repository using the following filters: project id "TCGA-PRAD", data category "transcriptome profiling", data type "gene expression quantification", experimental strategy "RNA-seq", and workflow type "STAR – counts". The clinico-pathological information was downloaded from UCSC Xena (last accession: March 2022, [30]).
- *GSE46602* (validation dataset) [31]: 36 prostate cancer samples from patients undergoing radical prostatectomy. This dataset used a gene expression microarray platform.

All datasets were downloaded from public repositories and, therefore, this study did not require ethical approval.

2.2. Data pre-processing

We downloaded the raw RNA-seq data (.fastq files) from GSE51005, GSE48403, GSE54460, E-MTAB-6525 and GSE31528 datasets. Sequencing quality was assessed with the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We trimmed adapters and removed poor quality sequences (quality score <20) with Trimmomatic [32]. Sequence reads were aligned to the reference genome GRCh38 using the STAR aligner [33].

2.3. Differential gene expression analysis

Because our main goal was to investigate lncRNAs as potential biomarkers for prostate cancer progression, we first filtered the transcriptome data by lncRNA biotype, as reported in Ensembl. We then performed pairwise differential gene expression analysis for different groups of samples using two R-base packages: *DESeq2* (v1.32.0) [34] and *edgeR* (v3.34.0) [35]. We then filtered out the lncRNAs with an adjusted p-value (p_{adj}) > 0.05 or absolute Log_2Fold-Change ($|Log_2FC|$) < 1.60 (corresponding to approximately a 3-fold change in gene expression). We established this Log_2FC value as a cutoff based on the differential expression analysis of *PCA3* (supplemental Table A1), a very well-known lncRNA associated with prostate cancer. Finally, we selected the lncRNAs that were consistently identified by both algorithms in all pairwise comparisons as potential candidate genes for modeling a lncRNA expression-based risk score.

2.4. Modeling of a multi-lncRNA expression-based score

For the modeling of the multi-gene score we used the Cutoff Finder tool that fits Cox proportional hazard models to a dichotomized independent variable and a survival outcome; defining the optimal cutoff as the point with the smallest Log-rank p split [36]. After establishing the optimal cutoff, we dichotomized the individual lncRNA expressions based on these cut points and classified the patients into low- or

high-expression groups. Subsequently, we performed a univariate Progression–Free Survival analysis (PFS) comparing both groups with the Log-rank test; and estimated the Hazard Ratios (HR) and 95 % Confidence Intervals (95 %CI) using univariate Cox proportional hazard models. PFS (Progression-Free Interval or PFI, as called by TCGA) was defined as the time between the date of diagnosis and the date of the first occurrence of a new tumor event, which includes progression of the disease, locoregional recurrence, distant metastasis, new primary tumor, or death with tumor.

Then, we combined all significant lncRNAs identified in the univariate analysis in a multivariate Cox proportional hazard model to estimate the coefficients and construct a multi-lncRNA score as follows:

$$multi\ lncRNA\ \ score = \sum_{i=1}^{n} \beta_i \times Exp\ lncRNA_i \qquad \text{Equation 1}$$

Where the β coefficients were obtained from the multivariate Cox proportional hazard model, and *Exp lncRNA* is the dichotomized expression of the lncRNA.

Finally, we stratified the patients into low- or high-multi-lncRNA score using the median score as the cutoff value and performed a PFS analysis comparing these two groups. PFS was evaluated by Kaplan-Meier curves, the Log-rank test and the Cox proportional hazard model regression using the R package *survival* (v3.5-5) [37].

2.5. Construction of a combined risk score based on lncRNA and clinico-pathological data

We first performed a multivariate Cox proportional hazard model for PFS including the multi-lncRNA score (previously calculated), pathologic T stage, pre-operative PSA and ISUP group as covariates, defined as: ISUP grading categorized as ISUP1 (Gleason score $\leq$ 6), ISUP2 (Gleason score 3 + 4 = 7), ISUP3 (Gleason score 4 + 3 = 7), ISUP4 (Gleason score = 8), ISUP5 (Gleason scores = 9–10); pathological T-stage classified as pT1, pT2, pT3 and pT4; pre-operative PSA considered as a numerical continuous variable; and the multi-lncRNA score dichotomized into 0 (low-score) or 1 (high-score) by the median value.

All variables were then combined into a new score as follows:

$$combined\ risk\ score = \sum_{i=1}^{n} \beta_i \times covariable_i \qquad \text{Equation 2}$$

The β coefficients were obtained from the multivariate Cox proportional hazard model, and *covariable* is the significant variable selected from the multivariate model.

PFS was evaluated by Kaplan-Meier curves, the Log-rank test and the Cox proportional hazard model regression using the R package *survival* (v3.5-5) [37].

2.6. Receiver Operating Characteristic (ROC) curves

We constructed ROC curves to analyze the performance of the combined risk score in predicting prostate cancer progression. We calculated the AUC (Area Under Curve) for the individual variables and the combined risk score. The AUC is a parameter that measures the performance of a classification method; the higher the AUC, the better the model at classifying the samples into the assigned groups. ROC curves and AUCs were calculated using the R package *pROC* (v1.18.0) [38].

2.7. Construction of a nomogram

We constructed a nomogram to estimate the risk of prostate cancer progression at 3 and 5 years using the R package *rms* (v6.5-0) (https://cran.r-project.org/package=rms).

2.8. R-packages for graphics

The following R packages were used for the different graphics: *ggplot2* (v3.4.2) [39], *venn* (v1.11) (https://CRAN.R-project.org/package=venn), *survminer* (v0.4.9) (https://CRAN.R-project.org/package=survminer), and *rms* (v6.5-0) (https://cran.r-project.org/package=rms).

3. Results

We conducted a bioinformatics study using transcriptomic data from public repositories. We performed differential gene expression analyses to identify candidate lncRNAs associated with different stages of prostate cancer, evaluated their association with disease progression and constructed a nomogram integrating lncRNA expression data and clinico-pathological features to assist uro-oncologists to better predict patients' risk of progression. The overall study workflow is shown in Fig. 1.

3.1. LncRNAs exhibited differential expression in prostate tumor samples across various stages of the disease

First, we aimed at identifying lncRNAs with differential expression in tissues representing different stages of prostate cancer progression (untreated primary tumors through mCRPC). To achieve this, we downloaded RNA-seq data from 148 primary tumors without treatment (100 radical prostatectomies and 48 biopsies), 11 paired pre- and post-ADT primary tumors (total = 22 samples) and 8 mCRPC. We then performed the following comparisons: a) paired post- vs pre-ADT samples, b) post- vs pre-ADT primary tumors, and c) mCRPC vs post-ADT. To increase the robustness and confidence of our findings, we used two R-base packages to calculate differential gene expression: *DESeq2* and *edgeR*. Subsequently, we filtered out all lncRNAs with a $p_{adj} > 0.05$ or $|Log_2FC| < 1.60$. For downstream analysis, we selected lncRNAs that were consistently found dysregulated by both packages across all three comparisons (Fig. 2A–C). This approach allowed us to identify 30 lncRNAs that responded to ADT and exhibited further dysregulation during progression to mCRPC (Fig. 2C and supplemental Table A1). Interestingly, among the 30 identified lncRNAs, we found *PCA3* and *PCGEM1* which have previously been reported to be associated with prostate cancer development and progression [40–42]. This finding allowed us to validate the workflow used.

Six lncRNAs were excluded for further analyses because they were reported as uncharacterized lncRNAs in different databases: *CH507–42P11.6*, *LOC100506474*, *LOC101929532*, *LOC101929563*, *LOC105373682*, and *LOC105375341*.

3.2. The lncRNAs selected were associated with progression-free survival (PFS)

To study the clinical relevance of the 24 lncRNAs identified in the previous step, we used samples from TCGA-PRAD (n = 498 primary tumors) as a validation dataset. We excluded 16 samples with incomplete lncRNA expression, phenotypic or survival data; resulting in the analysis of 482 prostate tumor samples.

We first dichotomized the expression of all 24 individual lncRNAs into low- or high-expression. To establish the best cutoff value for each lncRNA we used the Cutoff Finder tool. After stratifying the patients according to their lncRNA expression, we performed univariate PFS analyses using Kaplan-Meier curves and the Log-rank test to determine statistical differences between the groups; and estimated the Hazard Ratios (HR) using Cox proportional hazard models. We found that 18/24 lncRNAs were statistically associated with PFS in this cohort (Fig. 3A–G, and supplemental Figure A1 and Table A.2). Then, we performed a multivariate model including the expression of these 18 lncRNAs; and we found that 7 lncRNAs remained significant after adjusting the results

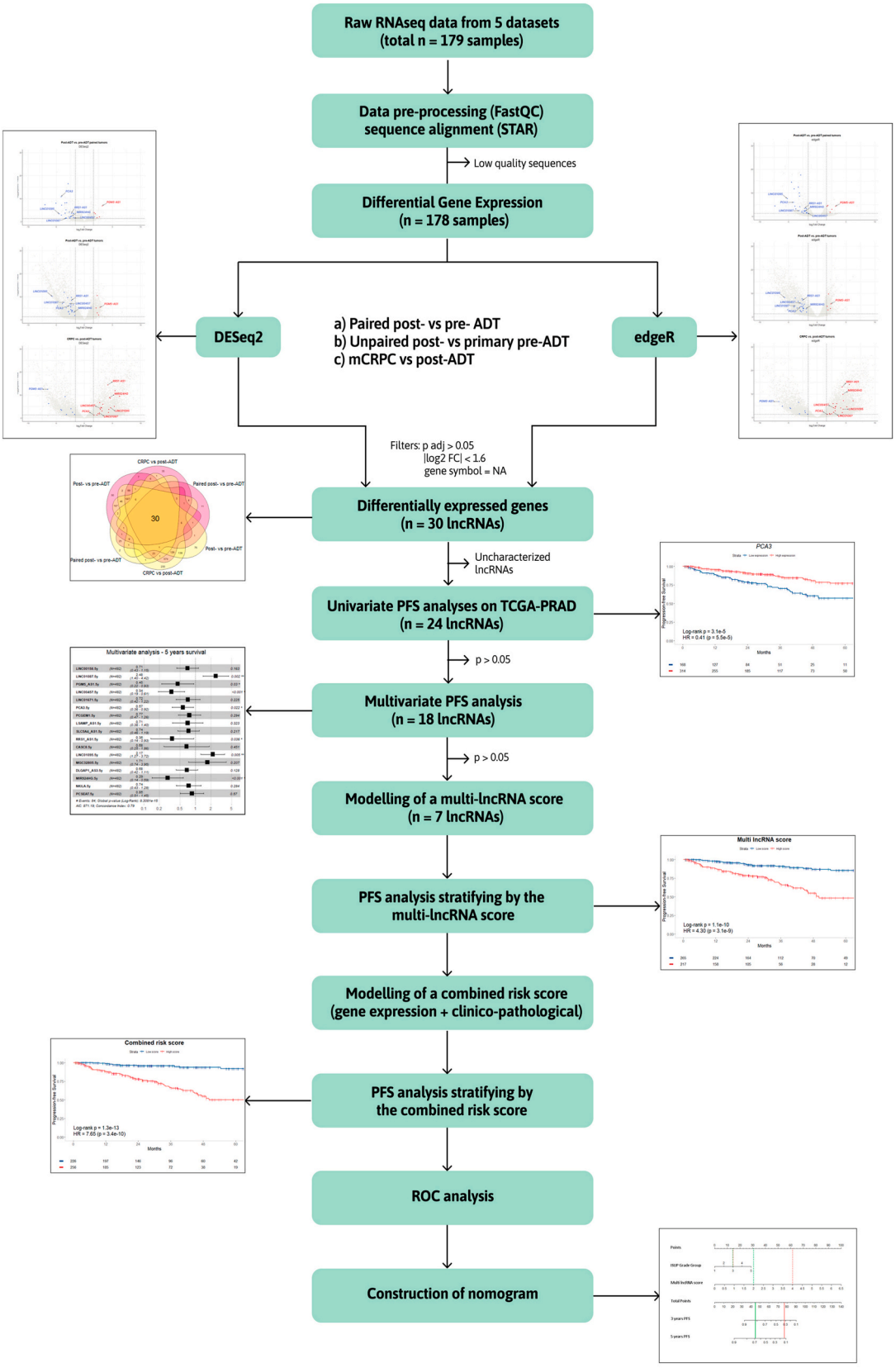


Fig. 1. Overall study workflow.
Abbreviations: ADT: androgen-deprivation therapy; mCRPC: metastatic castration-resistant prostate cancer; PFS: progression-free survival; ROC: receiver operating characteristic.

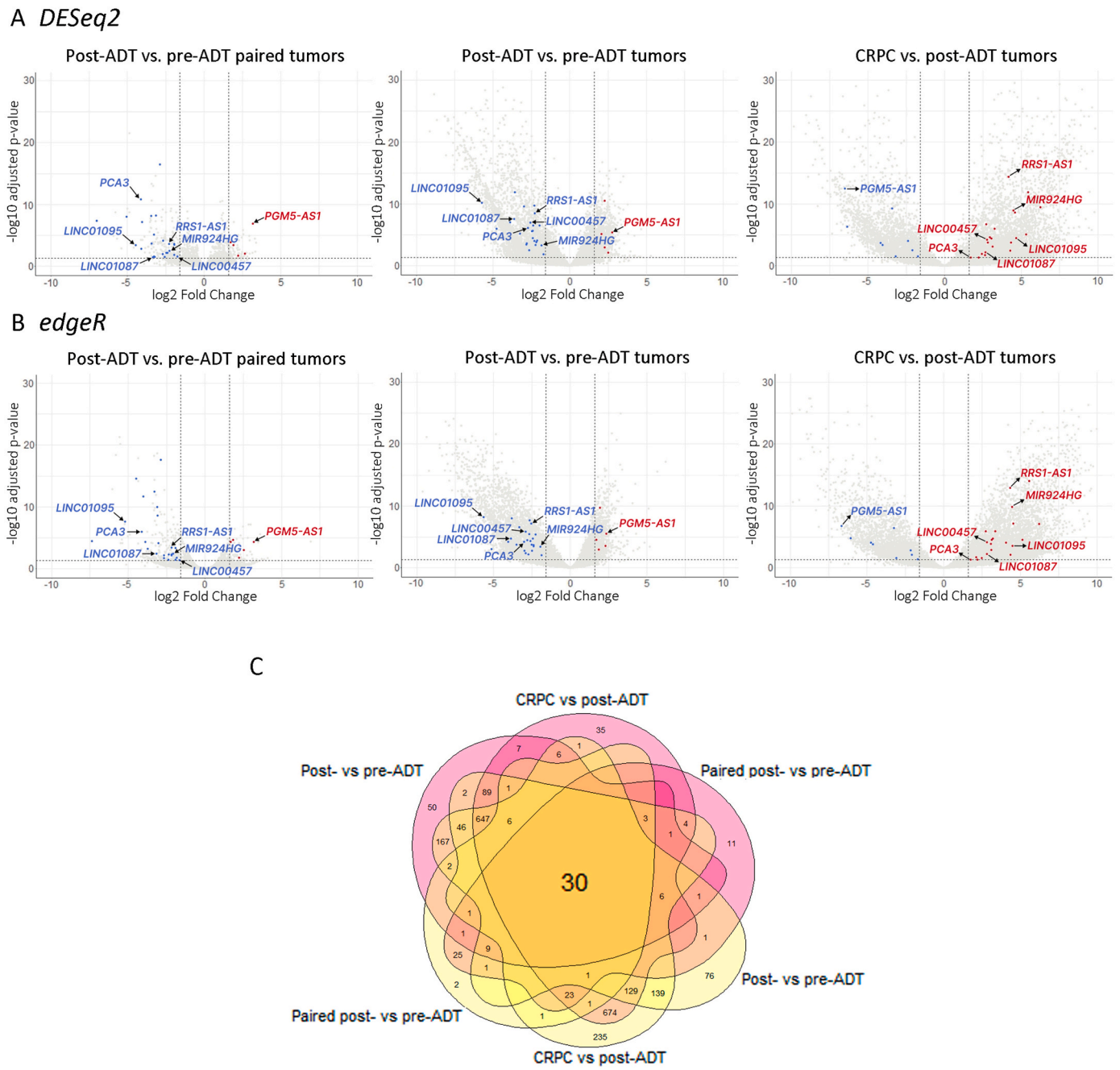


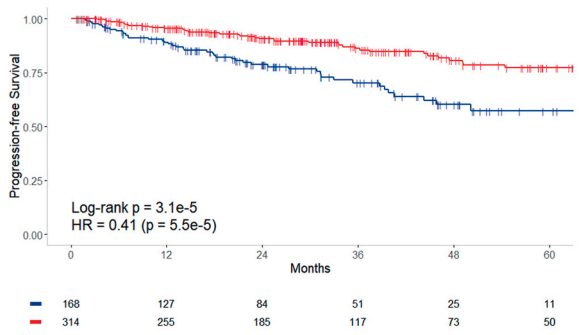
Fig. 2. Differential lncRNA expression in tissues representing different stages of prostate cancer progression.
The figure depicts the differential lncRNA expression analyses conducted with two R-based packages: *DESeq2* (panel A) and *edgeR* (panel B). In both panels, the volcano plots illustrate the results for the three comparisons made, with dashed lines indicating the cutoff values used for the selection of differentially expressed lncRNAs ($p_{adj} = 0.05$ and $|Log_2FC| = 1.60$). The 30 lncRNAs selected for further analysis are represented by blue dots for down-regulated genes and red dots for up-regulated genes (see supplemental Table A1 for detailed results). The names of the 7 lncRNAs used in the final risk score are provided within the volcano plots. Panel C features Venn diagrams displaying the number of dysregulated lncRNAs in each comparison as well as the intersection among them. The Venn diagrams colored in pink and yellow correspond to the analyses conducted by *DESeq2* and *edgeR*, respectively.

for the expression of the other lncRNAs (Fig. 3H). Overall, we observed that high expression of *PCA3, MIR924HG, PGM5-AS1, RRS1-AS1* (current gene symbol *RRS1-DT*) and *LINC00457* was associated with lower risk of progression compared to patients with low expression of these genes (HR < 0.50 and p ≤ 5.3e-3 for all genes; Fig. 3H). For *LINC01087* and *LINC01095*, we found that higher expression was associated with shorter PFS (HR = 2.17, p = 4.5e-3 and HR = 1.90, p = 1.3e-2; respectively; Fig. 3H).

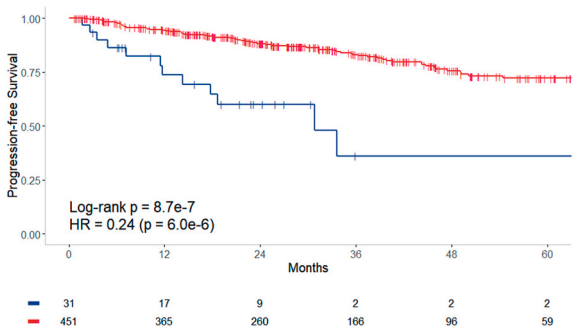
3.3. Modeling a polygenic risk score for prostate cancer progression

After confirming that the expression of these 7 lncRNAs were independent risk factors for prostate cancer progression, we created a multi-lncRNA score that captures the information about the expression of all 7 lncRNAs (Equation (1)). To do this, we coded the expression of the lncRNAs as 0 (reference) or 1 (poor-prognosis expression level) as determined by the multivariate Cox proportional hazard model for PFS. The multi-lncRNA score equation was: (0.73 x Exp *PCA3*) + (1.47 x Exp *MIR924HG*) + (1.32 x Exp *PGM5-AS1*) + (1.32 x Exp *RRS1-AS1*) + (1.26

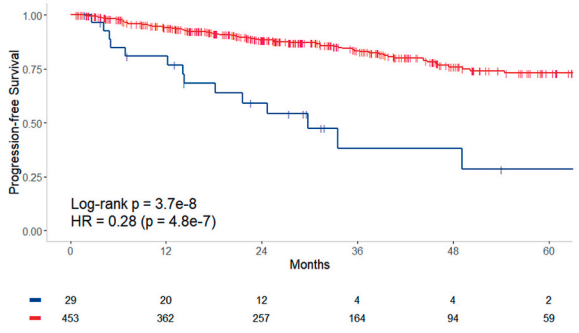
A



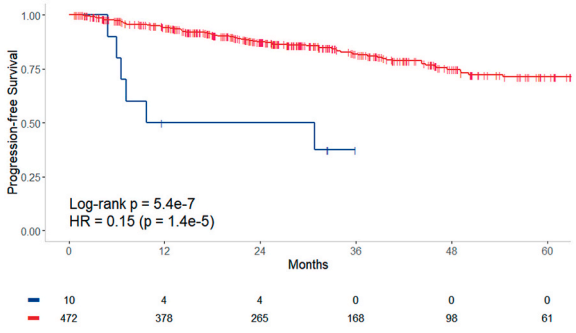
B



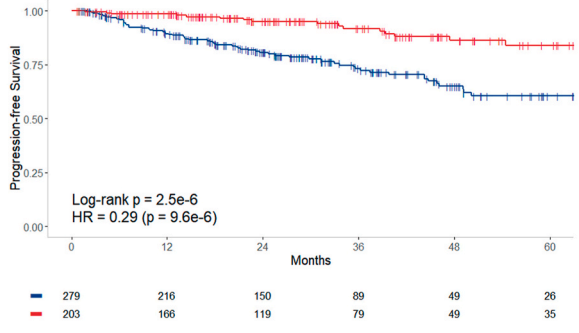
C



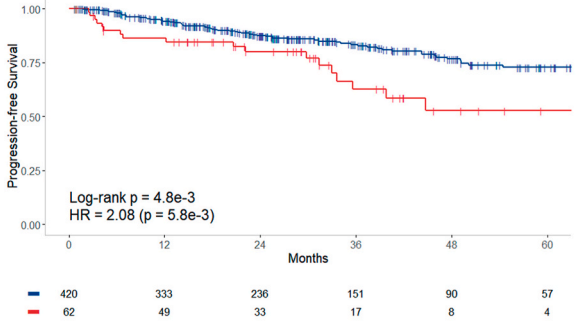
D



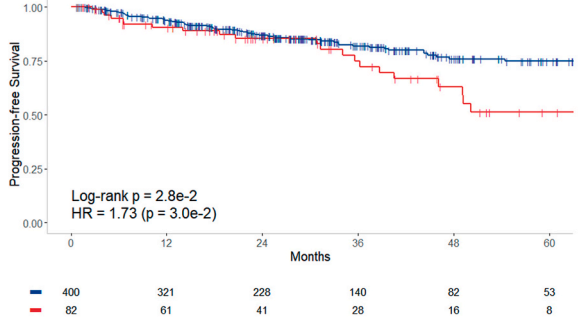
E



F



G



H

Characteristic	HR*	95% CI*	p-value
PCA3	0.48	0.31 - 0.76	1.4e-3
MIR924HG	0.23	0.12 - 0.44	1.3e-5
PGM5-AS1	0.27	0.15 - 0.48	1.1e-5
RRS1-AS1	0.27	0.11 - 0.68	5.3e-3
LINC00457	0.28	0.16 - 0.49	9.1e-6
LINC01087	2.17	1.27 - 3.70	4.5e-3
LINC01095	1.90	1.14 - 3.15	1.3e-2

*HR = Hazard Ratio, CI = Confidence interval

(caption on next page)

Fig. 3. Progression-free survival (PFS) analysis for patients with low- or high-expression of the 7 selected lncRNAs.
The figure illustrates the PFS analyses for the 7 lncRNAs selected used to construct the final risk score (results for the other 17 lncRNAs are presented in supplemental Figure A1 and Table A2). Panels A to G depict the PFS Kaplan-Meier curves for patients with low (blue) or high (red) expression of each individual lncRNA: *PCA3* (A), *MIR924HG* (B), *PGM5-AS1* (C), *RRS1-AS1* (D), *LINC00457* (E), *LINC01087* (F) and *LINC01095* (G). The Log-rank p, univariate Hazard Ratios (HR) and Cox p-values are provided within each plot, the numbers below each graph show the number of patients at risk at each time point, and censored patients are denoted with the vertical lines. The low-expression group was used as the reference in all comparisons. The table in panel H shows the Hazard Ratios (HR) estimated using a multivariate Cox proportional hazard model for the 7 lncRNAs selected (refer to supplemental Table A2 for additional information).

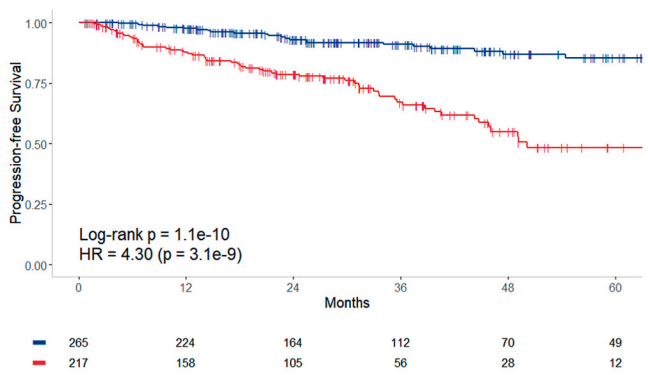
x Exp *LINC00457*) + (0.77 x Exp *LINC01087*) + (0.64 x Exp *LINC01095*).

Then, we calculated the multi-lncRNA score for each patient and categorized them into a low- or high-score using the median as the cutoff value (median multi-lncRNA score = 1.27) and performed a PFS analysis comparing these two groups. Patients with a high multi-lncRNA score were 4 times more likely to progress than patients with a low score (HR = 4.30, 95 %CI = 2.66–6.97, p = 3.1e-9; Fig. 4A).

3.4. The multi-lncRNA score is an independent risk factor for prostate cancer progression and enhances the predictive value of the ISUP group

Current clinico-pathological features used in standard clinical practice to predict patient outcomes include PSA at diagnosis, pT stage and Gleason score/ISUP group. Consequently, we performed a multivariate Cox hazard model including clinico-pathological and transcriptomic data as covariates. We found that the ISUP group and the multi-lncRNA score were the only variables that remained significant after adjusting for covariables (HR = 1.53, 95 %CI = 1.24–1.89, p = 7.4e-5 and HR =

A

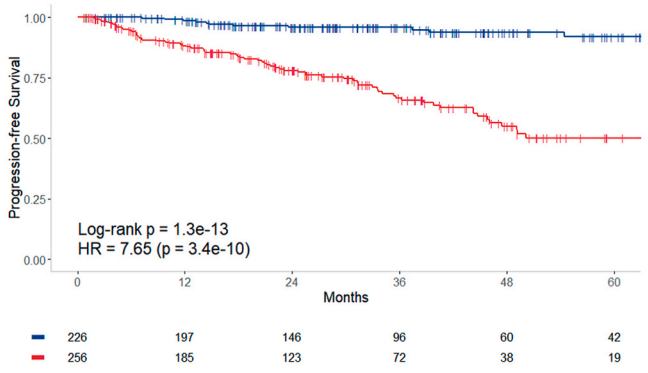


B

Characteristic	HR*	95% CI*	p-value
Preoperative PSA	1.00	0.99 – 1.02	0.81
Pathologic T			
T2	–	–	
T3	1.67	0.87 – 3.20	0.13
T4	0.89	0.18 – 4.32	0.89
ISUP grade group	1.53	1.24 – 1.89	7.4e-5
Multi lncRNA score	2.78	1.68 – 4.59	6.4e-5

*HR = Hazard Ratio, CI = Confidence interval

C



D

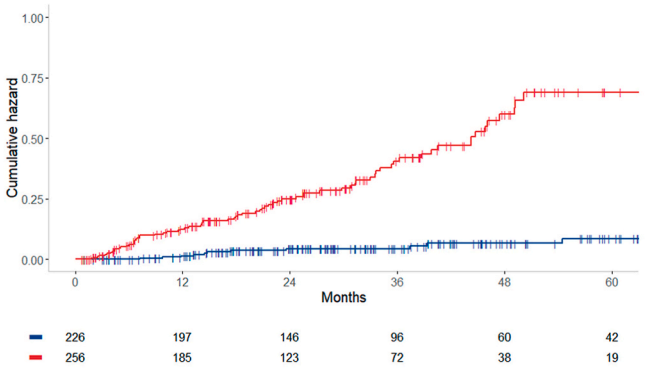


Fig. 4. Progression-free survival (PFS) analysis using the multi-lncRNA and combined risk scores.
A) PFS Kaplan-Meier curves for patients with low (blue) or high (red) multi-lncRNA score. Patients with a high multi-lncRNA score exhibited significantly shorter PFS and a 4-fold risk of progression when compared to patients with a low score. B) The table shows the Hazard Ratio estimated by a multivariate Cox proportional hazard model including PSA, pT, ISUP group and multi-lncRNA score as covariates. After the adjustment for covariates, only the ISUP group and the multi-lncRNA score remained as independent risk factors for disease progression. C) PFS Kaplan-Meier curves for patients with low (blue) or high (red) combined risk score. Patients with a high combined risk score had significantly poorer PFS and a nearly 8-fold risk of progression compared to patients with a low score. D) Cumulative risk curve for patients with low (blue) or high (red) combined risk score. Patients with a high combined risk score had greater cumulative risk at all-time points compared to patients with a low combined score.
The numbers below each Kaplan-Meier graph show the number of patients at risk at each time point, and censored patients are denoted with the vertical lines.

2.78, 95 %CI = 1.68–4.59, p = 6.4e-5, respectively; Fig. 4B).

Next, we combined the gene expression and clinico-pathological variables into a new risk score (Equation (2)). We coded the expression of the multi-lncRNAs as 0 (reference) or 1 (poor-prognostic score) as determined by the multivariate Cox proportional hazard model for PFS; for the ISUP covariable we used the patient's ISUP group. The combined risk score for each patient was calculated as: (1.06 x multi-lncRNA score) + (0.50 x ISUP). Patients were then assigned to a low-

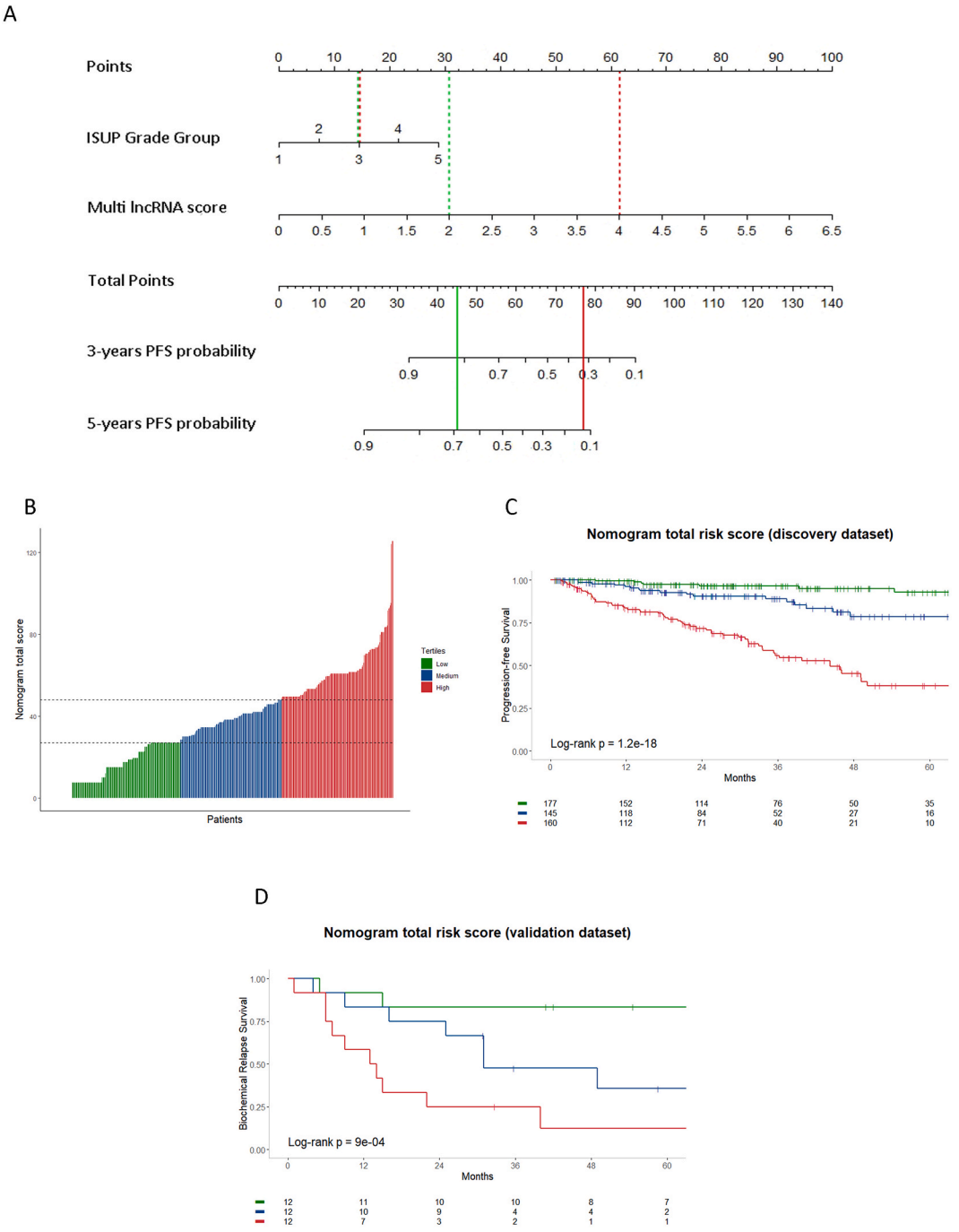


Fig. 5. Construction and validation of a nomogram for the prediction of prostate cancer progression.
A) Nomogram for predicting 3- and 5-year PFS in men with prostate cancer. The probability of being progression-free at 3 and 5 years for each patient could be easily estimated as follows: i) draw a vertical line upward connecting the value of each variable with the Points on the upper scale (dashed lines), ii) add all points and, using the Total Points scale, draw a vertical line downward to estimate patient PFS probability (solid lines). The figure shows two examples: 1) ISUP = 3 and multi-lncRNA score = 2 (green lines) and 2) ISUP = 3 and multi-lncRNA score = 4 (red lines). B) Waterfall plot showing the distribution of patients according to their total nomogram score. Patients were categorized into low-, medium- or high-risk using the nomogram total score tertiles: (0–27], (27–48] and (48–125]; respectively. The dotted lines denote the cutoff values. C) PFS analysis to validate the nomogram. As expected, the worse survival was observed for the high-risk group (red line), followed by the medium-risk group (blue line) and finally the low-risk group (green line). The numbers below the Kaplan-Meier curve show the number of patients at risk at each time point, and censored patients are denoted with the vertical lines. D) Biochemical relapse-free survival analysis for the validation dataset GSE46602. We observed that the higher the risk score (green: low-, blue: medium-, red: high-risk scores), the worse the patient's survival. The numbers below the Kaplan-Meier curve show the number of patients at risk at each time point, and censored patients are denoted with the vertical lines.

or high-score group using the median combined risk score (median = 2.07) as a cutoff value; and we performed a PFS analysis using this score. We found that patients with a high score had nearly 8-fold higher risk of progression (HR = 7.65, 95 %CI = 4.05–14.44, p = 3.4e-10; Fig. 4C). The cumulative risk of disease progression is shown in Fig. 4D.

In order to evaluate the performance of the multi-lncRNAs score, ISUP groups and the combined risk score in predicting 3- and 5-year PFS, we implemented a ROC curve analysis and compared the AUCs. We did not observe significant differences between the AUCs for ISUP-only and multi-lncRNAs-only scores. However, there was a significant enhancement in the performance of the classification method when the combined risk score was used: AUC 3-years 0.76 vs 0.72 (p = 0.046) and AUC 5-years 0.76 vs 0.72 (p = 0.036) for the combined risk score and ISUP-only groups, respectively (supplemental Figure A2).

3.5. Construction and validation of a nomogram

We constructed a nomogram to predict the 3- and 5-year PFS based on the ISUP group and multi-lncRNA score (Fig. 5A). The nomogram shows that the multi-lncRNA contributed the most to the prognosis of men with prostate cancer. As an example, a patient with an ISUP = 3 and multi-lncRNA score = 2 would have ~80 % probability of surviving without progression at 3 years and ~70 % at 5 years (Fig. 5A, green lines); while the probability of surviving free of progression would be ~30 % at 3 years and ~10 % at 5 years for a patient with an ISUP = 3 and multi-lncRNA score = 4 (Fig. 5A, red lines).

Finally, in order to validate the prognostic capacity of the nomogram, we stratified the patients from TCGA-PRAD into low-, medium- or high-risk groups based on the nomogram total score (Fig. 5B). As expected, patients in the high-risk scores had significantly shorter PFS with a median time to progression of 44 months, while the median was not reached for the medium- and low-risk score groups (HR = 12.97, 95 %CI = 5.91–28.44, p = 1.6e-10 for high- vs low-risk; and HR = 3.89, 95 %CI = 2.27–6.68, p = 7.8e-7 for high- vs medium-risk; Fig. 5C). Moreover, the group with a medium score had significantly worse PFS compared to patients with a low score (HR = 3.35, 95 %CI = 1.39–8.08, p = 0.007; Fig. 5C). The proportion of patients free of progression for the three groups were: 96 % (low-risk), 89 % (medium-risk) and 56 % (high-risk) at 3 years; and 93 % (low-risk), 79 % (medium-risk) and 38 % (high-risk) at 5 years.

3.6. Validation on an independent dataset

In order to validate the multi-lncRNA score and nomogram, we sought independent public datasets that analyzed transcriptome data and assessed prostate cancer progression. Unfortunately, no additional datasets using RNA-seq as a platform for transcriptomic analysis were found. Consequently, we expanded our search to include platforms based on gene expression microarrays.

While several datasets were identified, we encountered few limitations for the validation of the multi-lncRNA score and nomogram. First, the microarrays used in these studies did not include probes for all 7 lncRNAs integrated into our multi-lncRNA score. This may be attributed to the recent emergence of the field of lncRNAs in disease, with the majority of this molecules being recently discovered, sequenced and validated. Second, these studies primarily focused on single progression outcomes (biochemical relapse, metastasis or death); whereas our analysis comprised a broader range of progression outcomes, including progression of the disease, locoregional recurrence, distant metastasis, new primary tumor, and death with tumor. Third, tumor grading was commonly reported as Gleason score without discriminating between 7 (3 + 4) and 7(4 + 3), which correspond to two different ISUP grades (ISUP 2 and ISUP 3, respectively).

Despite these limitations, we used the GSE46602 dataset [31] to validate our nomogram. This dataset included probes for 6 out of the 7 lncRNAs (*LINC00457* was not analyzed due to the lack of probes in the microarray) and studied only 36 patients. Additionally, because the authors used a gene expression microarray as a transcriptome platform, we were not able to use the same regression coefficients calculated from the RNA-seq data. Therefore, we calculated new regression coefficients for the microarray data using the same pipeline described here. Finally, since the authors reported Gleason scores of 7, we opted to combine ISUP 2 (Gleason score 7(3 + 4)) and ISUP 3 (Gleason score 7(4 + 3)) into a single group for constructing the validation nomogram.

The analysis of this validation dataset is presented in Fig. 5D. After stratifying the patients from this dataset into tertiles of nomogram total risk score including the 6-lncRNA score and ISUP grades (ISUP 1, ISUP 2/3, ISUP 4 and ISUP 5), we observed that the higher the total risk score, the worse the biochemical relapse-free survival, with an overall Log-rank p = 9e-4 (Fig. 5D).

In summary, despite all the limitations encountered in this validation, the results support the potential of the multi-lncRNA score and nomogram as a valuable tool in assessing prostate cancer progression.

4. Discussion

Prostate cancer remains a significant global health concern, needing innovative approaches to enhance its diagnosis and prognosis. In this context, the present study introduces a pioneering investigation into the potential of lncRNAs as crucial markers for guiding clinical decisions in prostate cancer management.

Upon prostate cancer diagnosis, uro-oncologists assess the risk of aggressive/life-threatening disease by evaluating various clinico-pathological features that include the Gleason score, ISUP grading, TNM stage and serum PSA levels, among others [2]. Even though these features are able to identify the aggressive potential of prostate tumors and disease progression, and despite the advancements in traditional diagnostic and prognostic tools, the inherent heterogeneity of prostate cancer poses challenges in accurately predicting disease progression [3]. Addressing this critical gap, our study aimed to unravel lncRNAs molecular changes in prostate cancer, shedding light on their differential expression patterns across distinct stages of the disease. By integrating comprehensive lncRNA profiles with clinico-pathological features, we have constructed a predictive nomogram that empowers healthcare practitioners to tailor treatment strategies and patient follow-up, ultimately optimizing the precision of prostate cancer care. This work thus not only pushes the boundaries of our understanding of prostate cancer at the molecular level but also promises tangible clinical benefits by ushering in a new era of personalized and effective management strategies.

lncRNAs have a myriad of cellular functions, acting as molecular hubs due to their ability to interact and modulate several proteins, mRNAs and miRNAs. Over recent years, is has been an exponential increase of evidence indicating that dysregulation of lncRNA contributes to the development and progression of disease, including prostate cancer [18,19]. In this study, we analyzed RNA-seq data from public repositories and evaluated differential lncRNAs expression across different stages of prostate cancer, aiming to identify dysregulated lncRNAs with potential prognostic significance. After performing differential expression using two R-base packages, univariate, multivariate models and PFS analyses, we identified 7 lncRNAs that displayed the following characteristics: i) they responded to testosterone levels, as they were significantly downregulated during ADT; ii) they were dysregulated during the transition to mCRPC, as their expression was higher in mCRPC tissues compared to post-ADT primary tumors, and even higher to those observed in pre-ADT primary tumors; and iii) their expression was associated with PFS in univariate and multivariate analysis, considering clinico-pathological features as covariates. Of note, *PCA3* was among them and has been previously implicated in prostate cancer development and progression [43]; thus, validating the robustness of the pipeline used. Moreover, there are FDA-approved diagnostic tests (Progensa™ *PCA3* [44] and APTIMA® *PCA3* [45]) that were designed to

quantify *PCA3* levels in urine. Interestingly, *PCA3* is overexpressed in prostate cancer compared with normal prostate tissue [46]; however, it falls short in detecting high-grade prostate tumors, potentially leading to under-diagnosis of aggressive cancer [47]. Consistent with this, the findings presented here and those reported by Alshalalfa et al. [48] showed that low levels of tumoral *PCA3* were associated with worse prognosis.

The analysis of individual gene expression has been extensively used in oncology; however, it has yet to yield sufficient sensitivity and specificity for disease diagnosis and prognosis. Herein, we pooled the expression of 7 lncRNAs associated with prostate cancer progression and survival to create a multi-lncRNA risk score. Notably, we found that this score outperformed the risk prediction of the individual genes. Some coding gene expression profiles have previously shown predictive capability in discriminating aggressive prostate cancer [7,8]. Nevertheless, these scores do not combine clinico-pathological data within their predictive models; instead, they use a sequential approach [5]. Our study overcame this limitation by combining the multi-lncRNA risk score with pathologic features to establish a comprehensive multivariable risk score. After multivariate PFS analysis, only the multi-lncRNA score and ISUP group remained significant and, therefore, they were incorporated in the final molecular and pathological risk model. Our combined score outperformed the ISUP-only and multi-lncRNA-only risk score, highlighting the importance of integrating all aspects of tumor physiopathology as a prognostic tool.

Existing diagnostic/predictive kits for prostate cancer include *Oncotype Dx Genomic Prostate Score*® and *Decipher*®. The initial investigation of the *Oncotype Dx Genomic Prostate Score*® demonstrated better predictive accuracy when adding this genomic score to the CAPRA score (Cancer of the Prostate Risk Assessment score, which integrates information on age at diagnosis, PSA, Gleason score of the biopsy, clinical T stage and percentage of biopsy cores with cancer); resulting in a higher AUC (0.67 vs 0.63 for the combined score and CAPRA-only score, respectively) [8]. In a more recent study, the *Oncotype Dx Genomic Prostate Score*® (derived from biopsies) was evaluated in relation to adverse prostate tumor features after radical prostatectomy. The authors reported a significant improvement in AUCs when incorporating the Genomic Prostatic Score into predictive models, with values of 0.78 vs. 0.74 ($p = 0.004$) for seminal vesicle invasion and 0.70 vs. 0.68 ($p < 0.001$) for extra-prostatic extension [11]. The *Decipher*® gene signature was developed to predict early metastasis following radical prostatectomy. In the discovery and validation study, the authors reported that the combined genomic-clinical classifier had greater ability to identify patients at risk of distant metastasis (AUC of 0.74 for the genomic-clinical score compared to an AUC of 0.69 for the clinical-alone variables in the validation cohort) [6]. Karnes et al. later reported comparable results using the same gene signature for predicting metastasis in at-risk patients (AUC = 0.82 vs 0.79 for the genomic-clinical and clinical-alone classifiers, respectively) [7]. More recently, a large multicenter cohort study (total n = 6928) concluded that the *Decipher*® gene signature combined with clinico-pathological risk groups is a better predictor of 5-year risk of distant metastasis (AUC = 0.66 for the CAPRA-only score, AUC = 0.65 for the NCCN (National Comprehensive Cancer Network) 4-risk-group only, and AUC = 0.75 for the genomic-clinical classifier) [9]. While the evaluated outcomes and technologies (gene expression microarray or TaqMan RT-qPCR) varied among these genomic tests, their common goal is to identify patients at risk of aggressive/progressive disease. Notably, all AUCs were similar to or lower than those reported here, which used RNA-seq data. Overall, our results provide valuable insights into the potential of the multi-lncRNA score as an innovative predictive tool for prostate cancer progression.

Finally, in view that genetic scores might be challenging in terms of interpretation when predicting patient outcome, we constructed a nomogram that included the multi-lncRNA risk score and ISUP group. The proposed nomogram exhibited significantly superior performance compared to the predictive models based on clinico-pathological features alone.

The findings of our study highlight the importance of integrating molecular/genomic and clinical data into healthcare practice and suggest that not all men with identical clinico-pathological tumor characteristics may benefit from standardized treatments. The incorporation of a multi-lncRNA score could facilitate an empirical assessment of the risk of prostate cancer progression for each patient, enabling a tailored decision-making and treatment. However, it is important to acknowledge some limitations of this study. First, the datasets used lack detailed data on clinico-pathological features and information on treatment that could influence survival outcomes. Second, although the discovery of differentially expressed lncRNAs was performed on datasets from biopsies and FFPE tissues, the PFS analysis and construction of the nomogram was performed using expression data from radical prostatectomies; therefore, the prediction capacity of the nomogram should be further validated in biopsy samples. Third, although we were able to validate the multi-lncRNA and nomogram in an independent dataset with a limited number of samples and incomplete data on the 7 lncRNAs and ISUP groups, these results should be further validated in prospective multicenter cohorts; hence, large prospective clinical studies are needed to confirm the predictive capacity of the nomogram. Additionally, considering the potential utility of the multi-lncRNA score as a biomarker for prostate cancer progression and the possibility of detecting lncRNAs in urine and blood samples (e.g. *PCA3* [5,44,45]), these results warrant further investigation into the usage of the multi-lncRNA score as a liquid biopsy for the diagnosis and prognosis of prostate cancer.

5. Conclusions

In conclusion, this study bridges the gap between genomic/transcriptomic information and clinical practice, providing a practical framework for predicting prostate cancer progression and optimizing patient care. By contributing to the growing field of precision medicine, our findings pave the way for a more personalized and effective approach to prostate cancer management, ultimately improving patient outcomes and quality of life. Further research and validation are warranted to fully realize the transformative potential of lncRNA-based predictive models in the realm of prostate cancer diagnosis and prognosis.

CRediT authorship contribution statement

Sabrina Ledesma-Bazan: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing, Visualization. **Florencia Cascardo:** Formal analysis, Investigation, Methodology, Writing – review & editing. **Juan Bizzotto:** Formal analysis, Methodology, Data curation, Writing – review & editing. **Santiago Olszevicki:** Methodology, Data curation, Writing – review & editing. **Elba Vazquez:** Investigation, Writing – review & editing, Formal analysis, Funding acquisition. **Geraldine Gueron:** Formal analysis, Funding acquisition, Investigation, Writing – review & editing, Conceptualization. **Javier Cotignola:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing, Resources.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgement

The results shown here are part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

RNA-seq sequences were aligned to the reference genome on the high-performance computational clusers CeCAR (Centro de Computación de Alto Rendimiento, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina) and CCAD (Centro de Computación de Alto Desempeño, Universidad Nacional de Córdoba, Argentina).

This study was funded by: Universidad de Buenos Aires (20020170100585BA), Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación (PICT-2019-03215, PICT-2021-III-A-00080).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ncrna.2024.01.014.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (2021) 209–249, https://doi.org/10.3322/caac.21660.

[2] N. Mottet, R.C.N. van den Bergh, E. Briers, T. Van den Broeck, M.G. Cumberbatch, M. De Santis, S. Fanti, N. Fossati, G. Gandaglia, S. Gillessen, N. Grivas, J. Grummet, A.M. Henry, T.H. van der Kwast, T.B. Lam, M. Lardas, M. Liew, M.D. Mason, L. Moris, D.E. Oprea-Lager, H.G. van der Poel, O. Rouviere, I.G. Schoots, D. Tilki, T. Wiegel, P.M. Willemse, P. Cornford, EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. Part 1: screening, diagnosis, and local treatment with curative intent, Eur. Urol. 79 (2021) 243–262, https://doi.org/10.1016/j.eururo.2020.09.042.

[3] M.C. Haffner, W. Zwart, M.P. Roudier, L.D. True, W.G. Nelson, J.I. Epstein, A.M. De Marzo, P.S. Nelson, S. Yegnasubramanian, Genomic and phenotypic heterogeneity in prostate cancer, Nat. Rev. Urol. 18 (2021) 79–92, https://doi.org/10.1038/s41585-020-00400-w.

[4] S. Loeb, M.A. Bjurlin, J. Nicholson, T.L. Tammela, D.F. Penson, H.B. Carter, P. Carroll, R. Etzioni, Overdiagnosis and overtreatment of prostate cancer, Eur. Urol. 65 (2014) 1046–1055, https://doi.org/10.1016/j.eururo.2013.12.062.

[5] B.E. Boehm, M.E. York, G. Petrovics, I. Kohaar, G.T. Chesnut, Biomarkers of aggressive prostate cancer at diagnosis, Int. J. Mol. Sci. 24 (2023), https://doi.org/10.3390/ijms24032185.

[6] N. Erho, A. Crisan, I.A. Vergara, A.P. Mitra, M. Ghadessi, C. Buerki, E.J. Bergstralh, T. Kollmeyer, S. Fink, Z. Haddad, B. Zimmermann, T. Sierocinski, K.V. Ballman, T. J. Triche, P.C. Black, R.J. Karnes, G. Klee, E. Davicioni, R.B. Jenkins, Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy, PLoS One 8 (2013) e66855, https://doi.org/10.1371/journal.pone.0066855.

[7] R.J. Karnes, E.J. Bergstralh, E. Davicioni, M. Ghadessi, C. Buerki, A.P. Mitra, A. Crisan, N. Erho, I.A. Vergara, L.L. Lam, R. Carlson, D.J. Thompson, Z. Haddad, B. Zimmermann, T. Sierocinski, T.J. Triche, T. Kollmeyer, K.V. Ballman, P.C. Black, G.G. Klee, R.B. Jenkins, Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at risk patient population, J. Urol. 190 (2013) 2047–2053, https://doi.org/10.1016/j.juro.2013.06.017.

[8] E.A. Klein, M.R. Cooperberg, C. Magi-Galluzzi, J.P. Simko, S.M. Falzarano, T. Maddala, J.M. Chan, J. Li, J.E. Cowan, A.C. Tsiatis, D.B. Cherbavaz, R.J. Pelham, I. Tenggara-Hunter, F.L. Baehner, D. Knezevic, P.G. Febbo, S. Shak, M.W. Kattan, M. Lee, P.R. Carroll, A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling, Eur. Urol. 66 (2014) 550–560, https://doi.org/10.1016/j.eururo.2014.05.004.

[9] D.E. Spratt, J. Zhang, M. Santiago-Jimenez, R.T. Dess, J.W. Davis, R.B. Den, A. P. Dicker, C.J. Kane, A. Pollack, R. Stoyanova, F. Abdollah, A.E. Ross, A. Cole, E. Uchio, J.M. Randall, H. Nguyen, S.G. Zhao, R. Mehra, A.G. Glass, L.L.C. Lam, J. Chelliserry, M. du Plessis, V. Choeurng, M. Aranes, T. Kolisnik, J. Margrave, J. Alter, J. Jordan, C. Buerki, K. Yousefi, Z. Haddad, E. Davicioni, E.J. Trabulsi, S. Loeb, A. Tewari, P.R. Carroll, S. Weinmann, E.M. Schaeffer, E.A. Klein, R. J. Karnes, F.Y. Feng, P.L. Nguyen, Development and validation of a novel integrated clinical-genomic risk group classification for localized prostate cancer, J. Clin. Oncol. 36 (2018) 581–590, https://doi.org/10.1200/JCO.2017.74.2940.

[10] X. Wu, D. Lv, M. Eftekhar, A. Khan, C. Cai, Z. Zhao, D. Gu, Y. Liu, A new risk stratification system of prostate cancer to identify high-risk biochemical recurrence patients, Transl. Androl. Urol. 9 (2020) 2572–2586, https://doi.org/10.21037/tau-20-1019.

[11] M. Covas Moschovas, C. Chew, S. Bhat, M. Sandri, T. Rogers, P. Dell'Oglio, S. Roof, S. Reddy, M.C. Sighinolfi, B. Rocco, V. Patel, Association between Oncotype DX genomic prostate score and adverse tumor Pathology after radical prostatectomy, Eur Urol Focus 8 (2022) 418–424, https://doi.org/10.1016/j.euf.2021.03.015.

[12] Z. Mou, J. Spencer, B. Knight, J. John, P. McCullagh, J.S. McGrath, L.W. Harries, Gene expression analysis reveals a 5-gene signature for progression-free survival in prostate cancer, Front. Oncol. 12 (2022) 914078, https://doi.org/10.3389/fonc.2022.914078.

[13] Y. Zhao, Z. Tao, L. Li, J. Zheng, X. Chen, Predicting biochemical-recurrence-free survival using a three-metabolic-gene risk score model in prostate cancer patients, BMC Cancer 22 (2022) 239, https://doi.org/10.1186/s12885-022-09331-8.

[14] Y. Mitobe, K.i. Takayama, K. Horie-Inoue, S. Inoue, Prostate cancer-associated lncRNAs, Cancer Lett. 418 (2018) 159–166, https://doi.org/10.1016/j.canlet.2018.01.012.

[15] L. Statello, C.J. Guo, L.L. Chen, M. Huarte, Gene regulation by long non-coding RNAs and its biological functions, Nat. Rev. Mol. Cell Biol. 22 (2021) 96–118, https://doi.org/10.1038/s41580-020-00315-9.

[16] M.M. Balas, A.M. Johnson, Exploring the mechanisms behind long noncoding RNAs and cancer, Noncoding RNA Res 3 (2018) 108–117, https://doi.org/10.1016/j.ncrna.2018.03.001.

[17] J. Carlevaro-Fita, A. Lanzos, L. Feuerbach, C. Hong, D. Mas-Ponte, J.S. Pedersen, P. Drivers, G. Functional Interpretation, R. Johnson, P. Consortium, Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis, Commun. Biol. 3 (2020) 56, https://doi.org/10.1038/s42003-019-0741-7.

[18] Y. Qian, L. Shi, Z. Luo, Long non-coding RNAs in cancer: implications for diagnosis, prognosis, and therapy, Front. Med. 7 (2020) 612393, https://doi.org/10.3389/fmed.2020.612393.

[19] C. An, I. Wang, X. Li, R. Xia, F. Deng, Long non-coding RNA in prostate cancer, Am J Clin Exp Urol 10 (2022) 170–179.

[20] M. Tellez-Gabriel, D. Heymann, Exosomal lncRNAs: the newest promising liquid biopsy, Cancer Drug Resist 2 (2019) 1002–1017, https://doi.org/10.20517/cdr.2019.69.

[21] O. Beylerli, I. Gareev, A. Sufianov, T. Ilyasova, Y. Guang, Long noncoding RNAs as promising biomarkers in cancer, Noncoding RNA Res 7 (2022) 66–70, https://doi.org/10.1016/j.ncrna.2022.02.004.

[22] T. Guo, X.H. Tang, X.Y. Gao, Y. Zhou, B. Jin, Z.Q. Deng, Y. Hu, X.F. Xing, Z.Y. Li, J. F. Ji, A liquid biopsy signature of circulating exosome-derived mRNAs, miRNAs and lncRNAs predict therapeutic efficacy to neoadjuvant chemotherapy in patients with advanced gastric cancer, Mol. Cancer 21 (2022) 216, https://doi.org/10.1186/s12943-022-01684-9.

[23] J. Hu, C. Lai, Z. Shen, H. Yu, J. Lin, W. Xie, H. Su, J. Kong, J. Han, A prognostic model of bladder cancer based on metabolism-related long non-coding RNAs, Front. Oncol. 12 (2022) 833763, https://doi.org/10.3389/fonc.2022.833763.

[24] G. Huang, H. Cao, G. Liu, J. Chen, Role of androgen receptor signaling pathway-related lncRNAs in the prognosis and immune infiltration of breast cancer, Sci. Rep. 12 (2022) 20631, https://doi.org/10.1038/s41598-022-25231-0.

[25] P. Rajan, J. Stockley, I.M. Sudbery, J.T. Fleming, A. Hedley, G. Kalna, D. Sims, C. P. Ponting, A. Heger, C.N. Robson, R.M. McMenemin, I.D. Pedley, H.Y. Leung, Identification of a candidate prognostic gene signature by transcriptome analysis of matched pre- and post-treatment prostatic biopsies from patients with advanced prostate cancer, BMC Cancer 14 (2014) 977, https://doi.org/10.1186/1471-2407-14-977.

[26] P. Rajan, I.M. Sudbery, M.E. Villasevil, E. Mui, J. Fleming, M. Davis, I. Ahmad, J. Edwards, O.J. Sansom, D. Sims, C.P. Ponting, A. Heger, R.M. McMenemin, I. D. Pedley, H.Y. Leung, Next-generation sequencing of advanced prostate cancer treated with androgen-deprivation therapy, Eur. Urol. 66 (2014) 32–39, https://doi.org/10.1016/j.eururo.2013.08.011.

[27] Q. Long, J. Xu, A.O. Osunkoya, S. Sannigrahi, B.A. Johnson, W. Zhou, T. Gillespie, J.Y. Park, R.K. Nam, L. Sugar, A. Stanimirovic, A.K. Seth, J.A. Petros, C.S. Moreno, Global transcriptome analysis of formalin-fixed prostate cancer specimens identifies biomarkers of disease recurrence, Cancer Res. 74 (2014) 3228–3237, https://doi.org/10.1158/0008-5472.CAN-13-2699.

[28] M.H. Traka, A. Melchini, J. Coode-Bate, O. Al Kadhi, S. Saha, M. Defernez, P. Troncoso-Rey, H. Kibblewhite, C.M. O'Neill, F. Bernuzzi, L. Mythen, J. Hughes, P.W. Needs, J.R. Dainty, G.M. Savva, R.D. Mills, R.Y. Ball, C.S. Cooper, R. F. Mithen, Transcriptional changes in prostate of men on active surveillance after a 12-mo glucoraphanin-rich broccoli intervention-results from the Effect of Sulforaphane on prostate CAncer PrEvention (ESCAPE) randomized controlled trial, Am. J. Clin. Nutr. 109 (2019) 1133–1144, https://doi.org/10.1093/ajcn/nqz012.

[29] A.G. Sowalsky, Z. Xia, L. Wang, H. Zhao, S. Chen, G.J. Bubley, S.P. Balk, W. Li, Whole transcriptome sequencing reveals extensive unspliced mRNA in metastatic castration-resistant prostate cancer, Mol. Cancer Res. : MCR 13 (2015) 98–106, https://doi.org/10.1158/1541-7786.MCR-14-0273.

[30] M.J. Goldman, B. Craft, M. Hastie, K. Repecka, F. McDade, A. Kamath, A. Banerjee, Y. Luo, D. Rogers, A.N. Brooks, J. Zhu, D. Haussler, Visualizing and interpreting cancer genomics data via the Xena platform, Nat. Biotechnol. 38 (2020) 675–678, https://doi.org/10.1038/s41587-020-0546-8.

[31] M.M. Mortensen, S. Hoyer, A.S. Lynnerup, T.F. Orntoft, K.D. Sorensen, M. Borre, L. Dyrskjot, Expression profiling of prostate cancer tissue delineates genes associated with recurrence after prostatectomy, Sci. Rep. 5 (2015) 16018, https://doi.org/10.1038/srep16018.

[32] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120, https://doi.org/10.1093/bioinformatics/btu170.

[33] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21, https://doi.org/10.1093/bioinformatics/bts635.

[34] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, https://doi.org/10.1186/s13059-014-0550-8.

[35] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140, https://doi.org/10.1093/bioinformatics/btp616.
[36] J. Budczies, F. Klauschen, B.V. Sinn, B. Gyorffy, W.D. Schmitt, S. Darb-Esfahani, C. Denkert, Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization, PLoS One 7 (2012) e51862, https://doi.org/10.1371/journal.pone.0051862.
[37] T.M. Therneau, P.M. Grambsch, Modeling Survival Data: Extending the Cox Model, 1 ed., Springer New York, NY, 2000.
[38] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, M. Muller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinf. 12 (2011) 77, https://doi.org/10.1186/1471-2105-12-77.
[39] H. Wickham, ggplot2. Elegant Graphics for Data Analysis, Springer, New York, NY, 2009.
[40] G. Petrovics, W. Zhang, M. Makarem, J.P. Street, R. Connelly, L. Sun, I. A. Sesterhenn, V. Srikantan, J.W. Moul, S. Srivastava, Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients, Oncogene 23 (2004) 605–611, https://doi.org/10.1038/sj.onc.1207069.
[41] G.O. Ifere, G.A. Ananaba, Prostate cancer gene expression marker 1 (PCGEM1): a patented prostate- specific non-coding gene and regulator of prostate cancer progression, Recent Pat. DNA Gene Sequences 3 (2009) 151–163, https://doi.org/10.2174/187221509789318360.
[42] R.P. Pal, N.U. Maitra, J.K. Mellon, M.A. Khan, Defining prostate cancer risk before prostate biopsy, Urol. Oncol. 31 (2013) 1408–1418, https://doi.org/10.1016/j.urolonc.2012.05.012.
[43] S. Ghafouri-Fard, T. Khoshbakht, B.M. Hussen, A. Baniahmad, M. Taheri, F. Rashnoo, A review on the role of PCA3 lncRNA in carcinogenesis with an especial focus on prostate cancer, Pathol. Res. Pract. 231 (2022) 153800, https://doi.org/10.1016/j.prp.2022.153800.
[44] X. Durand, S. Moutereau, E. Xylinas, A. de la Taille, Progensa PCA3 test for prostate cancer, Expert Rev. Mol. Diagn 11 (2011) 137–144, https://doi.org/10.1586/erm.10.122.
[45] J. Groskopf, S.M. Aubin, I.L. Deras, A. Blase, S. Bodrug, C. Clark, S. Brentano, J. Mathis, J. Pham, T. Meyer, M. Cass, P. Hodge, M.L. Macairan, L.S. Marks, H. Rittenhouse, APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer, Clin. Chem. 52 (2006) 1089–1095, https://doi.org/10.1373/clinchem.2005.063289.
[46] M. Salagierski, J.A. Schalken, PCA3 and TMPRSS2-ERG: promising biomarkers in prostate cancer diagnosis, Cancers 2 (2010) 1432–1440, https://doi.org/10.3390/cancers2031432.
[47] J.T. Wei, Z. Feng, A.W. Partin, E. Brown, I. Thompson, L. Sokoll, D.W. Chan, Y. Lotan, A.S. Kibel, J.E. Busby, M. Bidair, D.W. Lin, S.S. Taneja, R. Viterbo, A. Y. Joon, J. Dahlgren, J. Kagan, S. Srivastava, M.G. Sanda, Can urinary PCA3 supplement PSA in the early detection of prostate cancer? J. Clin. Oncol. 32 (2014) 4066–4072, https://doi.org/10.1200/JCO.2013.52.8505.
[48] M. Alshalalfa, G.W. Verhaegh, E.A. Gibb, M. Santiago-Jimenez, N. Erho, J. Jordan, K. Yousefi, L.L.C. Lam, T. Kolisnik, J. Chelissery, R. Seiler, A.E. Ross, R.J. Karnes, E. M. Schaeffer, T.T. Lotan, R.B. Den, S.J. Freedland, E. Davicioni, E.A. Klein, J. A. Schalken, Low PCA3 expression is a marker of poor differentiation in localized prostate tumors: exploratory analysis from 12,076 patients, Oncotarget 8 (2017) 50804–50813, https://doi.org/10.18632/oncotarget.15133.