Human Genomics

RESEARCH

Open Access

Enhancing prostate cancer diagnosis: a machine learning-based biomarker approach

Patricia Porras-Quesada[1,2†], Alberto Ramírez-Mena[3†], Verónica Arenas-Rodríguez[1,2], Fernando Vázquez-Alonso[4], Jesús Alcalá-Fdez[5], Beatriz Álvarez-González[2,6], Luis Javier Martínez-González[1,2*†] and María Jesús Álvarez-Cubero[1,2,7†]

Abstract

Background The lack of reliable screening biomarkers for prostate cancer (PC) diagnosis makes tissue biopsy the gold-standard strategy. However, its frequent inconclusive results often lead to repeated procedures, increasing patient burden and healthcare costs. In this context, machine learning (ML) presents a novel approach for identifying gene signatures associated with tumor presence, offering a promising avenue for improved PC detection. Therefore, the present study explored the diagnostic potential of an ML-based gene signature and its applicability in tissue and plasma samples.

Methods This study evaluated the clinical applicability of an ML-based algorithm developed using TCGA data ($n=608$) and tested in an independent dataset ($n=349$). The eleven candidate genes that contributed most to the predictive model were initially profiled by targeted RNA-Seq in prostate tissue validation cohort ($n=141$) and further validated in replication tissue ($n=75$) and plasma ($n=50$) cohorts by qPCR and dPCR, respectively.

Results Gene expression analysis in prostate tissue led to the identification of a six-gene signature (*DLX1*, *TDRD1*, *AMACR*, *HPN*, *HOXC6*, and *OR51E2)* with high diagnostic performance (AUC = 95.9%). Expression patterns supported the gene signature's potential to identify false-negative cases and correctly classify inconclusive biopsy results. In plasma, *AMACR* demonstrated added diagnostic value as a non-invasive biomarker when integrated with clinical parameters (AUC = 93.21%).

Conclusions These findings demonstrate that our ML-based gene signature can accurately distinguish PC from non-tumoral tissue and resolve ambiguous biopsy results. Its integration alongside histopathology has the potential to reduce diagnostic uncertainty, improving PC early detection and guiding clinical decision-making.

†Patricia Porras-Quesada, Alberto Ramírez-Mena, Luis Javier Martínez-González and María Jesús Álvarez-Cubero have contributed equally to this work.

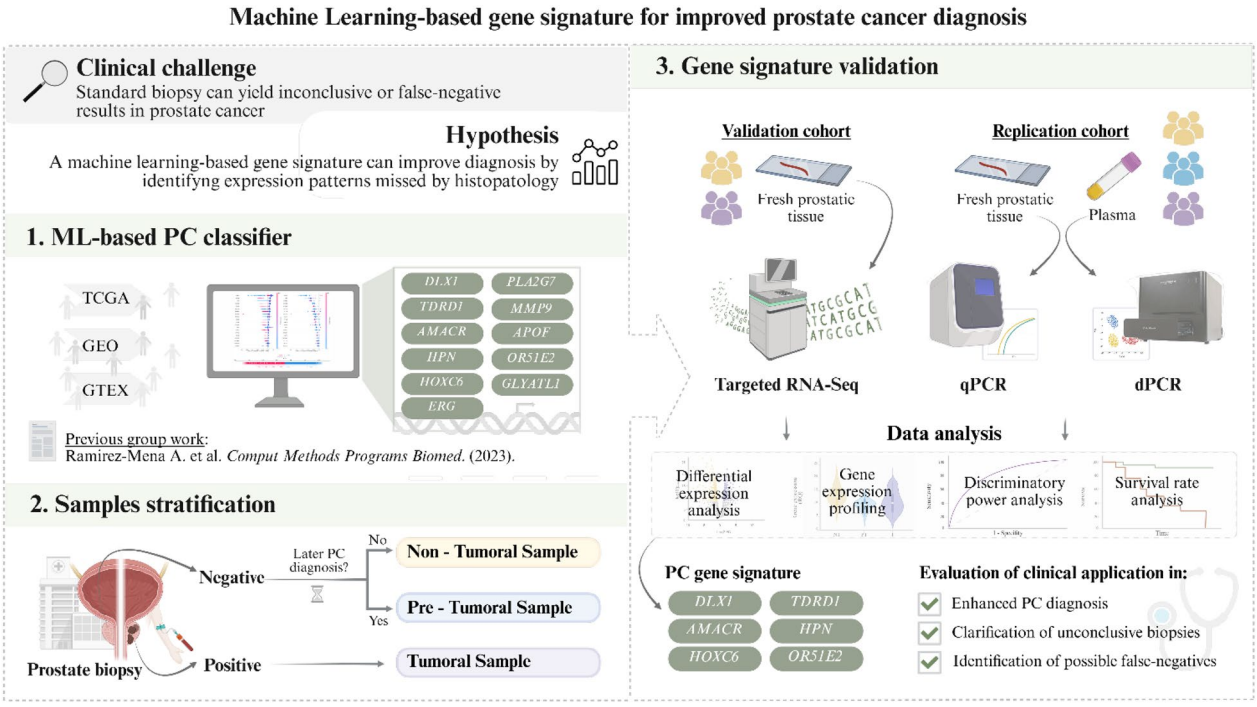
*Correspondence:
Luis Javier Martínez-González
luisjavier.martinez@genyo.es

Full list of author information is available at the end of the article

Key points

- This study validates a machine learning–derived gene expression signature that improves prostate cancer diagnosis, particularly in cases with ambiguous or false-negative biopsy results.
- The six-gene panel was experimentally validated using targeted RNA-Seq and PCR-based techniques in independent cohorts of prostate tissue and plasma, showing excellent diagnostic performance.
- By complementing standard histopathology, this molecular tool may enhance early detection, reduce unnecessary repeat biopsies, and support the integration of precision medicine into routine prostate cancer care.

Graphical Abstract



Introduction

Prostate cancer (PC) remains one of the most prevalent malignancies in men worldwide, driving extensive research efforts to enhance both treatment and diagnostic tools [1]. Despite advances, current diagnostic approaches still have considerable limitations. The most widely used screening tool, serum prostate-specific antigen (PSA) testing, is not necessarily cancer-specific, often leading to unnecessary biopsies. Moreover, prostate biopsy is subject to sampling errors and tumor heterogeneity, resulting in false negatives and frequent re-biopsies, which carry risks such as infection, bleeding, and urinary retention [2].

In recent years, the implementation of multiparametric magnetic resonance imaging (mpMRI) combined with targeted biopsy has significantly improved PC detection and reduced unnecessary procedures, increasing diagnostic sensitivity and specificity by up to 14% compared to conventional systematic biopsy approaches [2, 3]. However, even with mpMRI-guided strategies, clinically significant tumors can still be missed, particularly due to limitations in imaging resolution, operator variability, or sampling error in multifocal disease [4, 5]. In fact, false-negative results and ambiguous biopsy findings continue to pose a diagnostic challenge, often leading to patient anxiety and repeat procedures. This highlights the need for complementary molecular tools that can refine diagnostic accuracy, especially in cases where conventional methods fall short.

This diagnostic gap has driven efforts to identify molecular biomarkers that can complement current PC diagnosis methods. Among these, tissue-based gene

signatures have shown particular promise. For instance, ConfirmMDx detects DNA methylation changes in genes like *GSTP1*, *APC*, and *RASSF1*, improving the predictive accuracy beyond that of PSA by identifying epigenetic field effects in histologically negative biopsy cores [6]. SelectMDx, on the other hand, measures urinary expression of genes such as *DLX1* and *HOXC6* to predict the presence of high-grade PC upon biopsy [7]. While these signatures have demonstrated clinical utility in certain contexts, their widespread implementation has been limited by variability across cohorts, inconsistent regulatory approval, and challenges related to integration within standard diagnostic workflows.

In this context, artificial intelligence (AI) and, particularly, machine learning (ML), have emerged as a promising tool to uncover complex gene expression patterns linked to tumor biology. By integrating high-throughput molecular data with advanced computational modeling, ML-based approaches can identify precise gene signatures that outperform traditional clinicopathological variables in diagnostic and prognostic applications [8]. For instance, one such signature composed of 23 genes exhibited superior predictive performance compared to conventional clinicopathological features and previously published molecular signatures, emphasizing the synergistic role of both technologies in tackling complex diseases such as PC [9].

Building on this rationale, our group recently developed a ML–based gene expression classifier trained on publicly available transcriptomic datasets from prostate tissue, which showed excellent diagnostic performance in distinguishing tumor from non-tumor samples [10]. The model integrated gene importance rankings across independent cohorts and was optimized using ensemble algorithms to enhance robustness and reproducibility. Despite its strong in silico performance, the clinical utility of the classifier had not yet been evaluated in biological specimens. In this study, we address this gap by experimentally validating the top-ranked genes from the model using targeted RNA-Seq and PCR-based approaches in independent cohorts of prostate fresh tissue and plasma samples. Our ultimate goal is to define a clinically meaningful gene signature capable of complementing standard diagnostic methods, enhancing PC detection, and supporting informed clinical decision-making. This work represents a key step toward bridging computational discovery and real-world application in the field of molecular diagnostics.

Materials and methods

Gene selection

To guide potential translation to liquid biopsy, we prioritized genes from our previously published classifier that were both overexpressed in tumoral tissue (T), compared to non-tumoral (NT), and ranked highly for predictive relevance. Differential expression analyses were performed on three datasets used to train and validate the original model (TCGA-PRAD [11], GSE22260 [12], and GSE114740), selecting genes with higher normalized median expression in Tvs.NT samples. Predictive importance was assessed using SHapley Additive exPlanations (SHAP) scores across datasets [13]. The relevance score of each gene was calculated as the sum of the position each gene occupied when sorted in descending order of importance in each database, and the top eleven genes (rank < 100) were selected for experimental validation.

Study population

This study included data from 266 men who underwent prostate biopsy between 2012 and 2014, distributed into two independent cohorts. In both cohorts, samples were classified into non-tumoral (NT) or tumoral (T) samples, with T cases further stratified by tumor aggressiveness into T1 (ISUP < 3) and T2 (ISUP ≥ 3). The validation cohort ($n = 141$: NT = 41; T1 = 55; T2 = 45) consisted of fresh prostate tissue samples analyzed by targeted RNA-Seq. The replication cohort included tissue samples analyzed by quantitative PCR (qPCR; $n = 75$: NT = 20; T1 = 20; T2 = 20) and plasma samples analyzed by digital PCR (dPCR; $n = 50$: NT = 20; T1 = 18; T2 = 12). Updated clinical data in 2024 enabled further stratification of NT tissue samples into non-tumoral single-biopsy (NTS, $n = 13$) and non-tumoral multiple-biopsy (NTM, $n = 7$) groups, depending on whether one or more procedures were required to rule out cancer. In addition to NT and T tissue samples, a third group was included: pre-tumoral (PT, $n = 15$). Fresh prostate tissue and plasma were collected by the Urology Service at Hospital Universitario Virgen de las Nieves (Granada, Spain) and stored at − 80 °C and − 20 °C, respectively, until analysis. All participants provided written informed consent, and the study was approved by the Ethics Committee of Granada (CEI-Granada, code 1638-N-18), following the Declaration of Helsinki.

RNA isolation from tissue and plasma samples

Total RNA from tissue samples was extracted using the Trizol®/chloroform method, while RNA from 200 uL of plasma was isolated according to the manufacturer's protocol for the miRNeasy Serum/Plasma Kit (Qiagen). RNA quantity and purity were assessed by A260/A280 ratio using a NanoDrop™ 2000c spectrophotometer (ThermoFisher Scientific), and RNA integrity was evaluated with the 4200 TapeStation system (Agilent Technologies).

Gene expression analysis in fresh tissue samples by targeted RNA-Seq

Custom capture library baits were designed with SureDesign software (Agilent Technologies). Library preparation, hybridization, and capture were performed using the SureSelect XT HS2 RNA System for Illumina paired-end sequencing libraries (Agilent Technologies), following the manufacturer's instructions. Final libraries were quality-checked using the 4200 TapeStation (Agilent Technologies) and subsequently pooled for sequencing. Sequencing was carried out on an Illumina NovaSeq 6000 platform using v2 200-cycle cartridges (2 × 100 bp, Illumina).

Gene expression analysis in fresh tissue samples by qPCR

RNA reverse transcription was performed using PrimeScript™ RT reagent Kit (Takara Bio). qPCR analysis was carried out using TaqMan probes (Life Technologies Co.; Table S1). Expression analysis was performed according to the manufacturer's protocol, on a 96-well plate with QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems).

Gene expression analysis in plasma samples by dPCR

Reverse transcription was performed using SuperScript IV VILO Master Mix (ThermoFisher Scientific, Inc.). Gene expression was analyzed by dPCR on QuantStudio™ Absolute Q™ dPCR (Applied Biosystems).

Table 1 Differential expression of candidate genes in tumoral versus non-tumoral prostate tissue samples

Gene	Logarithmic fold change	*p*-value	FDR
DLX1	3.6192	8.2951E-16	4.2028E-14
TDRD1	4.2084	5.04437E-20	1.5335E-17
AMACR	4.0521	1.9054E-20	1.1585E-17
HPN	2.2004	1.4315E-13	4.5808E-12
HOXC6	2.6725	4.5910E-14	1.6420E-12
ERG	2.7777	3.0216E-15	1.4132E-13
PLA2G7	2.1090	2.2468E-13	6.8304E-12
MMP9	0.8738	0.0085	0.02064
APOF	2.5391	3.4232E-14	1.3008E-12
OR51E2	3.2831	2.4831E-16	1.6775E-14
GLYATL1	2.7639	1.7127E-16	1.3016E-14

FDR = False Discovery Rate

Bioinformatics analysis

Just the data obtained from the article [10] was the starting point for the present validation data. All bioinformatic analyses previously done were reported by Ramirez Mena, A. *et cols* [10]. Regarding RNA-Seq analysis, reads were quality-checked using FastQC (v0.12.1) and MultiQC (v1.18). Trimming and duplicate removal were performed with Agilent Trimmer (v3.1.2) and Agilent CREAK (v1.1.1), respectively. Reads were aligned using BWA (v0.7.19), and gene counts were generated with featureCounts (v2.0.6). Differential expression analysis was performed using edgeR v4.7.3, Statistical significance was set at False Discovery Rate (FDR) < 0.05.

Statistical analysis

All analyses were performed using SPSS v.22 (IBM Corporation, United States) and GraphPad Prism 10.2.3 (LCC). Data normality was assessed with the Shapiro-Wilk test and non-parametric tests (Kruskal-Wallis and U-Mann-Whitney) were applied. Statistical significance was set at $p < 0.05$. Genes showing significant differential expression were further evaluated for their diagnostic performance using Receiver Operating Characteristic (ROC) curves, from which the area under the curve (AUC) values were calculated to quantify their discriminative ability. Significant genes were then combined in a multivariable logistic regression model to assess their collective ability to predict the presence of PC.

Results

Machine learning-based gene selection for genetic signature validation

The present study is based on previous data reported [10] where we have described the top 20 gene predictors, according to their importance by SHAP obtained from the Random Forest algorithm using the TCGA-PRAD dataset. Table S2 presents a comprehensive list of genes with overexpression in tumoral (T) relative to non-tumoral (NT) tissue across all databases, sorted by their score. The eleven genes with a rank below 100 (*DLX1*, *TDRD1*, *AMACR*, *HPN*, *HOXC6*, *ERG*, *PLA2G7*, *MMP9*, *APOF*, *OR51E2*, and *GLYATL1*) were selected for their validation. More details about selected genes are shown in Table S3.

Study population description

Key characteristics of participants from validation and replication cohorts are summarized in Table S4. Compared to controls, patients with PC exhibited significantly higher PSA levels, while no differences in age were observed between groups. Among PC patients, stratification by tumor aggressiveness revealed significant differences in the proportion of metastases across all cohorts. Additionally, in the tissue cohorts, T2 patients showed higher PSA levels and increased D'Amico risk compared to T1.

Transcriptomic profiling of machine learning–selected genes

To confirm the upregulation of the 11 ML–derived genes (*DLX1*, *TDRD1*, *AMACR*, *HPN*, *HOXC6*, *ERG*, *PLA2G7*, *MMP9*, *APOF*, *OR51E2*, and *GLYATL1*), we first analyzed their expression in prostate tissue samples from the validation cohort ($n = 141$) using a targeted RNA-Seq approach. Each gene displayed significantly higher expression in tumoral (T) compared to non-tumoral (NT) samples (Table 1), supporting their potential diagnostic relevance. These differences remained consistent when patients were stratified by tumor aggressiveness (Tables S5 and S6). Based on these results, the full 11-gene panel was carried forward for experimental validation by qPCR.

Evaluation of gene expression profiling in PC diagnosis

To assess the diagnostic potential of each candidate gene, expression levels were measured by qPCR and compared between tumor (T) and non-tumor (NT) tissue samples. Six genes, *DLX1* ($p = 0.0008$), *TDRD1* ($p = 0.0041$), *AMACR* ($p = 0.0111$), *HPN* ($p = 0.0367$), *HOXC6* ($p = 0.0003$), and *OR51E2* ($p = 0.0039$), showed statistically significant differential expression (Fig. 1 Tables S7 and S8). In the stratified analysis by tumor aggressiveness (T2vs.T1), *AMACR* also showed significant differences ($p = 0.0389$), suggesting its potential as a marker of both diagnosis and aggressiveness (Fig. 1; Tables S7 and S8).

To evaluate the diagnostic performance of these six genes, ROC curve analyses were conducted. *HOXC6* and *DLX1* displayed the best individual performance, with *HOXC6* achieving an AUC = 0.8693 (95% confidence interval (CI): 0.7571–0.9815; sensitivity: 93.10%, specificity: 50%) and *DLX1* an AUC = 0.8117 (95% CI: 0.6775–0.9460; sensitivity: 95%, specificity: 57.89%) (Figure S1A–F; Table 2. A logistic regression model incorporating all six genes improved the AUC to 0.959, supporting the robustness of the combined signature (Figure S1G, Table 2).

In clinical practice, patients with negative biopsies often undergo repeated procedures to rule out cancer. In our cohort, 7 of 20 patients (35%) with initial negative biopsies required additional biopsies, all of which remained negative (Fig. 2A). We compared the gene expression profiles of NTS (single biopsy) and NTM (multiple biopsies) patients with those of confirmed tumor cases (T) to assess the model's ability to distinguish true negatives. As expected, NTS and NTM showed no significant differences in gene expression, and both were clearly distinct from T samples (Fig. 2B; Tables S9 and S10), highlighting the potential of the signature to reduce unnecessary biopsies and complement standard histopathological evaluation.

Evaluation of gene expression profiling in prostate biopsy for false-negative identification

To explore the clinical value of the proposed gene signature in detecting false-negative prostate biopsies, we analyzed a third group of individuals (Pre-Tumoral, PT) who had initial biopsy results classified as negative but were diagnosed with PC within four years (Fig. 2A). The expression profiles of the six candidate genes were assessed to determine whether these cases aligned more closely with tumor-positive or tumor-negative samples. The PT group showed gene expression patterns more similar to T than to NT samples, with *DLX1*, *TDRD1*, *HOXC6*, and *OR51E2* showing the most notable differences (Fig. 2C; Tables S11 and S12). These results suggest that incorporating the gene signature into routine biopsy analysis could help identify false-negative cases earlier, thereby reducing misclassification and supporting timely intervention.

Evaluation of gene expression profiling in PC survival

In our study population, overall survival (OS) rates were 90% at 3 years, 86.67% at 5 years, and 69.09% at 10 years. *AMACR*, *HPN*, and *OR51E2* were significant for OS, showing reasonable specificity and sensitivity values to predict disease progression (Fig. 3). In contrast, survival analyses predicated upon the expression levels of *DLX1*, *TDRD1*, and *HOXC6* exhibited no statistically significant differences (Figure S2).

Translation of gene expression profiling to liquid biopsy

Among the six candidate genes, only *AMACR* and *OR51E2* were detectable in plasma samples. *AMACR* expression was significantly higher in patients compared

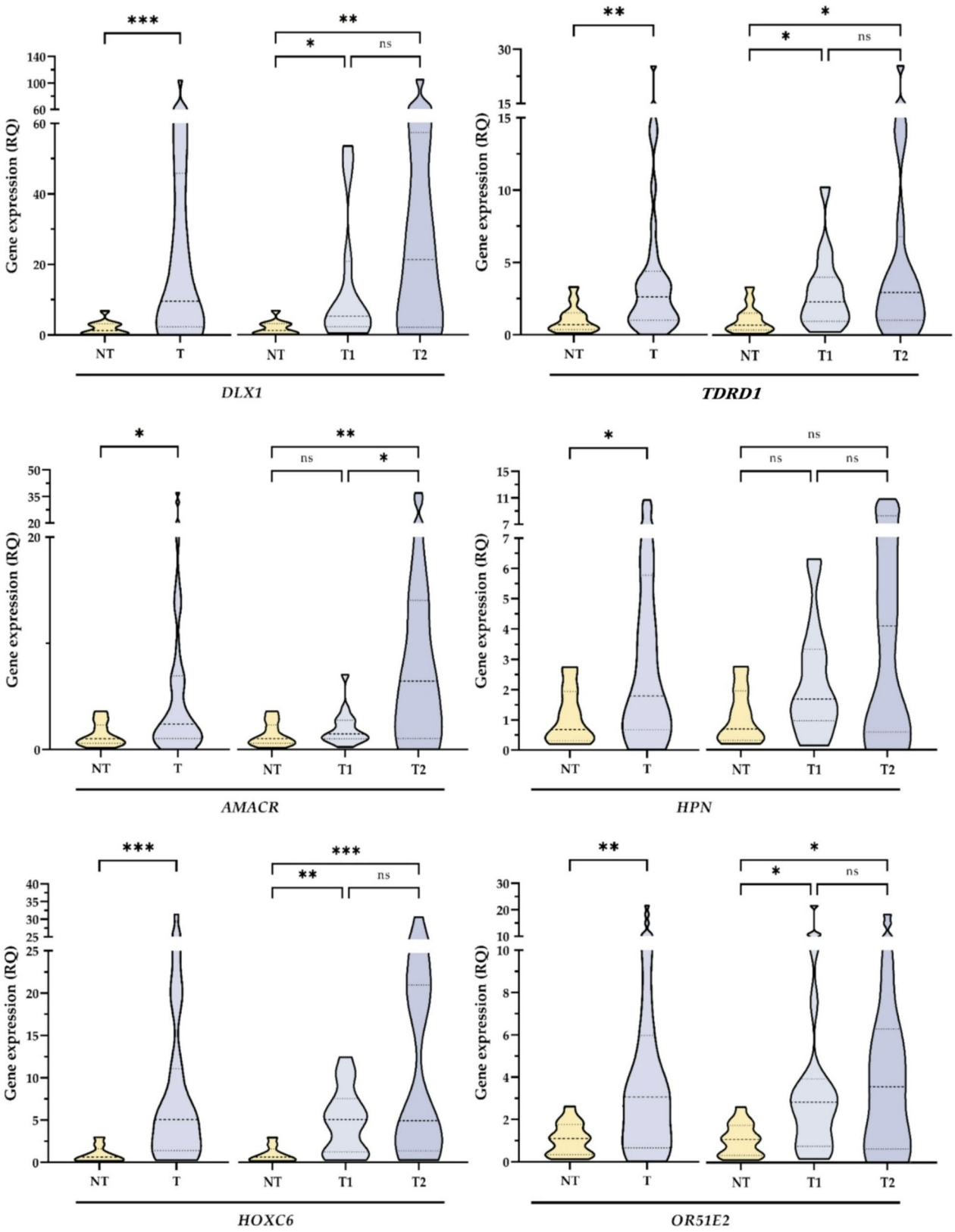


Fig. 1 Gene expression analysis of six candidate genes in prostate tissue. Comparison were made between NT and T groups, and among NT, T1, and T2 groups for *DLX1*, *TDRD1*, *AMACR*, *HPN*, *HOXC6*, and *OR51E2* expression levels. * Represents p-value < 0.05, ** represents p-value < 0.01, *** represents p-value < 0.001, and ns represents non-significant

Table 2 ROC parameters of biomarkers both individually and combined in fresh tissue samples for PC diagnosis.

Biomarker (T vs. NT)	AUC	95% CI	PPV	NPV	Sensitivity	Specificity
DLX1	0.8117	0.6775–0.9460	0.7037	0.9167	0.9500	0.5789
TDRD1	0.7806	0.6339–0.9273	0.6333	0.8333	0.9048	0.4762
AMACR	0.7141	0.5730–0.8551	0.6176	0.6667	0.7778	0.4800
HPN	0.6870	0.5422–0.8319	0.6111	0.6667	0.8148	0.4167
HOXC6	0.8693	0.7571–0.9815	0.7941	0.7778	0.9310	0.500
OR51E2	0.7426	0.6069–0.8784	0.6176	0.8750	0.9130	0.5185
All combined	0.9593	0.8950-1	0.9259	0.9000	0.9615	0.8182

AUC Area Under the Curve, *CI* Confidence Interval, *NPV* Negative Predictive Value, *PPV* Positive Predictive Value

to controls ($p = 0.0319$); however, no significant difference was observed for *OR51E2* (Figure S3; Tables S13 and S14). ROC curve analyses were performed for *AMACR*, *OR51E2*, and PSA, the current clinical biomarker for PC diagnosis, both individually and in combination through logistic regression models (Figures S4; Table S15). All combinations outperformed PSA alone, with the *AMACR* + PSA model achieving the highest diagnostic accuracy (AUC = 0.9321; 95% CI = 0.7984–1), 96.43% sensitivity, and 100% specificity (Figure S4D; Table S15).

Discussion

The development of non-invasive biomarkers remains a central goal in oncology, particularly for improving early detection in PC. While numerous studies have explored potential candidates, results remain inconclusive. Urine proteomics, for instance, has identified proteins such as SLURP1 with elevated expression in PC, although treatment-dependent variability limits their reliability [14]. Similarly, a retrospective study analyzing long non-coding RNAs, including *PCA3*, *DLX1*, *HOXC6*, and *TMPRSS2:ERG*, has shown diagnostic value in urine samples, with *PCA3* demonstrating moderate sensitivity and good specificity for distinguishing PC from non-PC cases [15].

In this study, we identified a six-gene signature (*DLX1*, *HPN*, *AMACR*, *TDRD1*, *HOXC6*, and *OR51E2)* capable of improving PC detection, offering a promising strategy for managing biopsies classified as negative or uncertain under current approaches. All six genes exhibited significant expression patterns correlated with OS, reinforcing their clinical relevance. Importantly, the classifier identified individuals potentially misclassified by conventional histopathology, highlighting its value as a complementary tool in challenging diagnostic scenarios. This molecular approach may enable earlier detection and more precise management.

Among these genes, *AMACR* showed the strongest link to PC, with previous studies reporting expression levels up to 682 times higher in tumor compared to normal tissue [16]. Other genes, such as *DLX1*, *OR51E2*, or *HOXC6*, have also been associated with PC biology. *DLX1* and *HOXC6* are part of SelectMDx, a test used for PC risk stratification [17], and *DLX1* has been proposed as a non-invasive biomarker for PC because of its involvement in ERG/AR signaling [18]. OR51E2, a surface protein related to KLK2, has been associated with treatment response and survival outcomes [19] and has also been described as a potential biomarker for PC [20]. Despite these individual associations, no previous study has evaluated the combined diagnostic potential of *DLX1*, *HPN*, *AMACR*, *TDRD1*, *HOXC6*, and *OR51E2* as an integrated gene signature in the context of prostate biopsy analysis.

AMACR and *OR51E2* also emerged as promising non-invasive biomarkers in plasma. Although the sample size limited statistical significance, the findings align with previous reports. *OR51E2* has been described as circulating mRNA with *SIM2* [21], though not yet validated in aggressive subtypes. *AMACR* has shown strong diagnostic and prognostic value across several biological matrices, including free mRNA [21], urinary exosomes [22], and plasma, where its overexpression correlates with poor outcomes [21]. These data align with our findings in both tissue and blood samples.

A major clinical challenge in PC diagnosis is managing patients with persistently negative biopsies but rising PSA levels. There is currently no standardized protocol to determine when or how re-biopsies should be conducted [23]. Clinicians increasingly demand new strategies, such as more specific PSA-based assays, imaging techniques, or molecular tissue analyses, to address this uncertainty [24]. Despite recent advances, such approaches are not yet incorporated into clinical guidelines, and consensus is still lacking [25]. We propose that incorporating molecular analysis, specifically the gene signature described here, into current diagnostic workflows may help reduce unnecessary repeat biopsies and improve risk stratification.

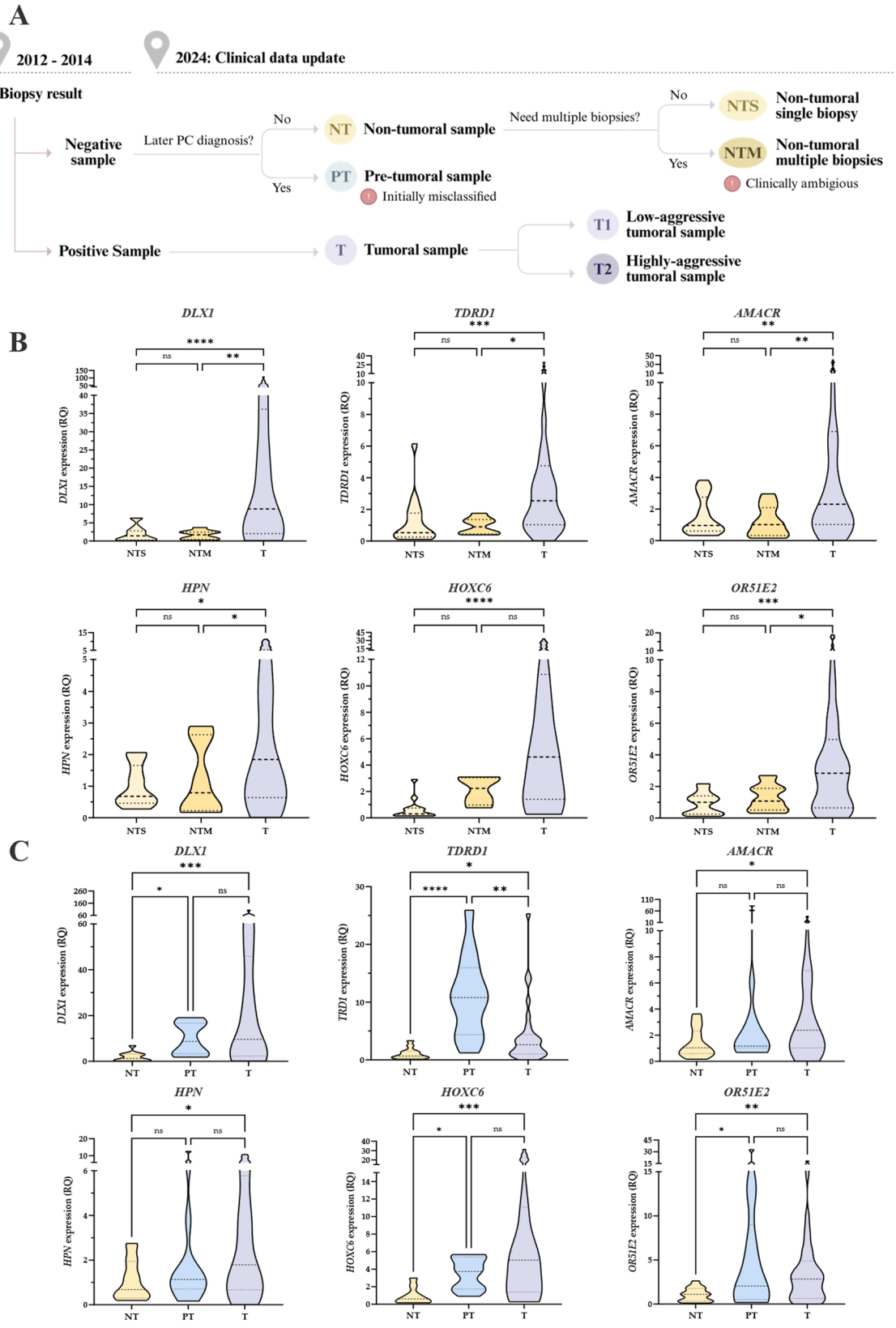


Fig. 2 Comparative expression profiling in prostate biopsies with inconclusive or false-negative results. A Schematic representation of the clinical and molecular classification of prostate tissue samples. **B** Gene expression comparison between NTS, NTM, and T groups **C** Gene expression comparison between NT, PT, and T groups. * Represents p-value < 0.05, ** represents p-value < 0.01, *** represents p-value < 0.001, **** represents p-value < 0.0001, and ns represents non-significant

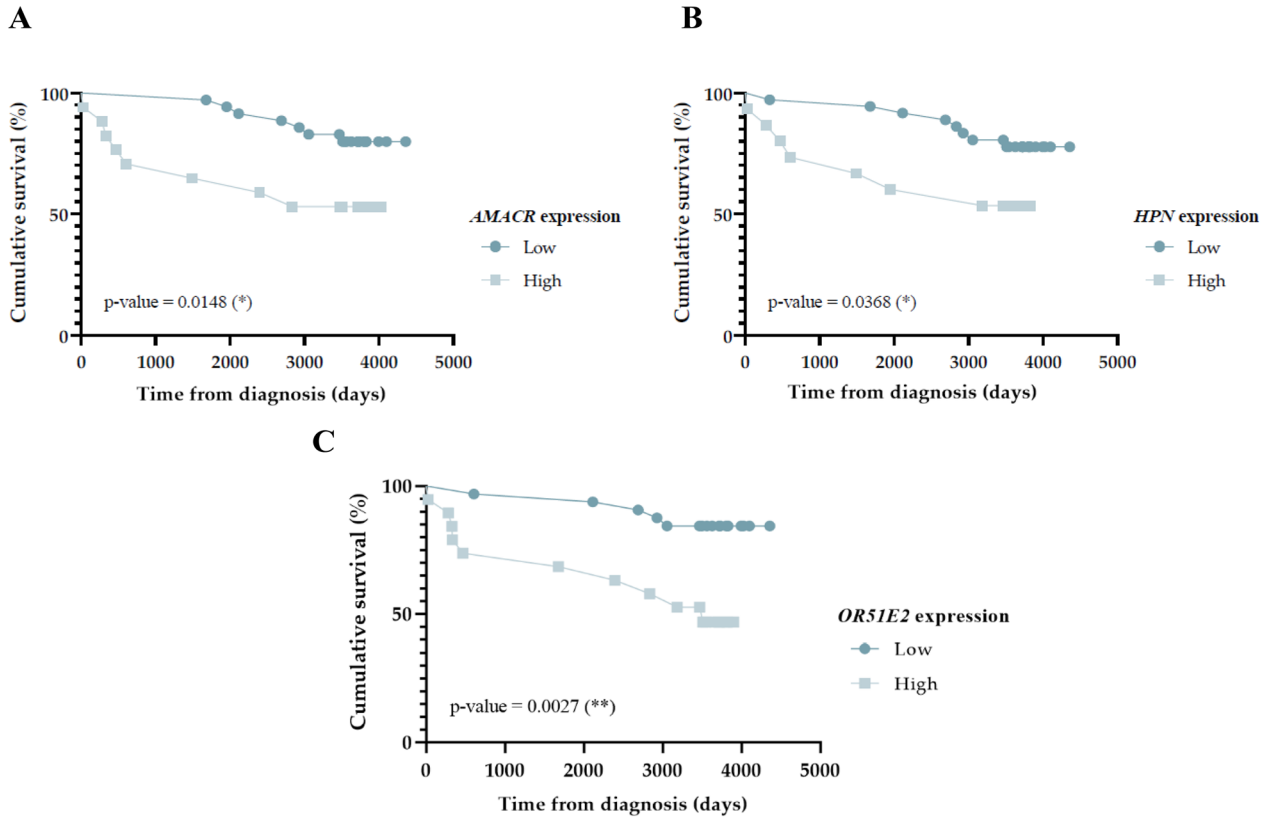


Fig. 3 Kaplan-Meier overall survival curves based on *AMACR*, *HPN*, and *OR51E2* expression levels using predefined cutoffs. Survival time is represented in days

Conclusions

This study supports the clinical value of a six-gene signature for PC diagnosis, particularly in patients with inconclusive biopsy results. These genes are associated with tumor presence and aggressiveness, enhancing early detection and reducing diagnostic uncertainty. Among them, *AMACR* stands out for its robust expression in both tissue and plasma, reinforcing its potential as a non-invasive biomarker. Integration of this signature into clinical workflows may enable more precise patient stratification and support personalized management strategies. Further validation in larger, independent cohorts will be necessary to confirm its utility and facilitate translation into routine clinical practice.

Abbreviations

AI	Artificial Intelligence
AUC	Area under the curve
CI	Confidence interval
FDR	False discovery rate
ML	Machine learning
mpMRI	Multiparametric magnetic resonance imaging
PT	Pre-Tumoral
NT	Non-Tumoral
NTM	Non-Tumoral multiple-biopsy
NTS	Non-Tumoral single-biopsy
OS	Overall survival
PC	Prostate cancer
PSA	Prostate-Specific antigen
ROC	Receiver operating characteristic
SHAP	SHapley additive exPlanations
T	Tumoral

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40246-026-00939-6.

Supplementary Material 1.

Acknowledgements

We would like to acknowledge the support of Spanish Foundation for Research in Urology (FIU, No. G80445661), University of Granada for facilitating the software for using TCGA. Moreover, to nursery of Urology Service of Virgen de las Nieves and GENYO Genomic Unit staff for making this study possible. We are also very grateful to all familiars and donors for their support and unconditional collaboration to the project. Present published results are a part of the PhD thesis of the candidate PPQ in the Biochemistry and Molecular Biology Doctoral Program of the University of Granada.

Author contributions

LJMG, MJAC, and FVA study design; PPQ and FVA samples collection; PPQ, ARM, VAR, and BAG data analysis and interpretation; ARM and JAF statistical analysis; LJMG, MJAC, ARM, and PPQ manuscript writing and critical revision; LJMG, JAF, and MJAC study supervision; LJMG and MJAC funding acquisition.

Funding

Carlos III Health Institute. TransMisiones 2023 - PMPTA23/00025 (MJAC, LJMG). Ministry of Education - FPU21/05808 (PPQ). Ministry of Education - FPU20/06611 (BAG).

Data availability
Data are available upon reasonable request to the corresponding author.

Code availability
Not applicable.

Declarations

Ethics approval and consent to participate
All study participants provided written informed consent before being enrolled, and the study was previously approved by the Research Ethics Committee of Granada Center (CEI-Granada internal code 1638-N-18) following Helsinki ethical declaration.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Biochemistry and Molecular Biology III and Immunology, Faculty of Medicine, Health Sciences Technology Park (PTS), University of Granada, Granada, Spain
[2]GENYO, Centre for Genomics and Oncological Research: Pfizer, Andalusian Regional Government; Health Sciences Technology Park (PTS), University of Granada, Granada, Spain
[3]IFMIF-DONES Spain Consortium, IFMIF-DONES, Granada, Spain
[4]Urology Department, University Hospital Virgen de las Nieves, Av. de las Fuerzas Armadas 2, Granada, Spain
[5]Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, Granada 18071, Spain
[6]Department of Legal Medicine and Toxicology, Faculty of Medicine, Health Sciences Technology Park (PTS), University of Granada, Granada, Spain
[7]Biosanitary Research Institute, ibs.GRANADA, University of Granada, Granada, Spain

Received: 24 September 2025 / Accepted: 20 February 2026
Published online: 24 February 2026

References

1. Silva KCS, Tambwe N, Mahfouz DH, Wium M, Cacciatore S, Paccez JD, et al. Transcription Factors in Prostate Cancer: Insights for Disease Development and Diagnostic and Therapeutic Approaches. Genes (Basel). 2024;15:450.
2. Friedemann M, Jandeck C, Tautz L, Gutewort K, von Rein L, Sukocheva O, et al. Blood-Based DNA Methylation Analysis by Multiplexed OBBPA-ddPCR to Verify Indications for Prostate Biopsies in Suspected Prostate Cancer Patients. Cancers (Basel). 2024;16:1324.
3. Leung DK, Chiu PK, Ng C, Teoh JY. Role of pre-biopsy multiparametric MRI in prostate cancer diagnosis: Evidence from the literature. Turk J Urol 47 (2020).
4. Alkema NG, Hoogeveen SFJS, Cauberg ECC, Witte LPW, de Kate Mv E, et al. Magnetic Resonance Imaging–targeted Prostate Biopsy Compared with Systematic Prostate Biopsy in Biopsy-naïve Patients with Suspected Prostate Cancer. Eur Urol Open Sci. 2022;44:125–30.
5. Williams C, Ahdoot M, Daneshvar MA, Hague C, Wilbur AR, Gomella PT, et al. Why Does MRI-Targeted Biopsy Miss Clinically Significant Cancer? J Urol. 2023;26:101097.
6. Van Neste L, Partin AW, Steward GD, Epstein JI, Harrison DJ, Van Criekinge W. Risk score predicts high-grade prostate cancer in DNA-methylation positive, histopathologically negative biopsies. Prostate. 2016;76:1078–87.
7. Haese A, Trooskens G, Steyaert S, Hessels D, Brawer M, Vlaeminck-Guillem V, et al. J Urol. 2019;202:256–63.
8. Ghamande SS, Cline JK, Sayyid RK, Klaassen Z. Advancing Precision Oncology With Artificial Intelligence: Ushering in the ArteraAI Prostate Test. Urology. 2024;188:20–3.
9. He X, Hu S, Wang C, Yang Y, Li Z, Zeng M, et al. Predicting prostate cancer recurrence: Introducing PCRPS, an advanced online web server. Heliyon. 2024;10:e28878.
10. Ramírez-Mena A, Andrés-León E, Alvarez-Cubero MJ, Anguita-Ruiz A, Martinez-Gonzalez LJ, Alcala-Fdez J. Explainable artificial intelligence to predict and identify prostate cancer tissue by gene expression. Comput Methods Programs Biomed. 2023;240:107719.
11. Li J, Xu C, Lee HJ, Ren S, Zi X, Zhang Z, et al. A genomic and epigenomic atlas of prostate cancer in Asian populations. Nature. 2020;580:93–9.
12. Kannan K, Wang L, Wang J, Ittmann MM, Li W, Yen L. Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. Proc Natl Acad Sci U S A. 2011;108:9172–7.
13. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. arXiv [cs AI] 4768–77, (2017).
14. Yang T, Liu T, Lei T, Li T, Liu N, Zhang M. Unveiling the potential of SLURP1 protein as a biomarker for prostate cancer screening. Front Oncol. 2024;14:1365615.
15. Mytsyk Y, Nakonechnyi Y, Dosenko V, Kowal P, Pietrus M, Gazdikova K, et al. The performance and limitations of PCA3, TMPRSS2:ERG, HOXC6 and DLX1 urinary markers combined in the improvement of prostate cancer diagnostics. Clin Biochem. 2023;116:120–7.
16. Kong G, Lee H, Tran Q, Kim C, Park J, Kwon SH, et al. Corrigendum: Current Knowledge on the Function of α-Methyl Acyl-CoA Racemase in Human Diseases. Front Mol Biosci. 2021;8:639164.
17. Katzendorn O, von Klot CAJ, Mahjoub S, Faraj Tabrizi P, Harke NN, Tezval H, et al. Combination of PI-RADS score and mRNA urine test-A novel scoring system for improved detection of prostate cancer. PLoS ONE. 2022;17:e0271981.
18. Goel S, Bhatia V, Kundu S, Biswas T, Carskadon S, Gupta N, et al. Transcriptional network involving ERG and AR orchestrates Distal-less homeobox-1 mediated prostate cancer progression. Nat Commun. 2021;12:5325.
19. Paniagua-Herranz L, Moreno I, Nieto-Jiménez C, Garcia-Lorenzo E, Díaz-Tejeiro C, Sanvicente A, et al. Genomic and Immunologic Correlates in Prostate Cancer with High Expression of KLK2. Int J Mol Sci. 2024;25:2222.
20. Qian Y, Feng D, Wang J, Wei W, Wei Q, Han P, et al. Establishment of cancer-associated fibroblasts-related subtypes and prognostic index for prostate cancer through single-cell and bulk RNA transcriptome. Sci Rep. 2023;13:9016.
21. Souza MF, Kuasne H, Barros-Filho MC, Cilião HL, Marchi FA, Fuganti PE, et al. Circulating mRNA signature as a marker for high-risk prostate cancer. Carcinogenesis. 2020;41:139–45.
22. Jin X, Ji J, Niu D, Yang Y, Tao S, Wan L, et al. Urine Exosomal AMACR Is a Novel Biomarker for Prostate Cancer Detection at Initial Biopsy. Front Oncol. 2022;12:904315.
23. Dani H, Loeb S. The role of prostate cancer biomarkers in undiagnosed men. Curr Opin Urol. 2017;27:210–6.
24. Blute ML Jr, Abel EJ, Downs TM, Kelcz F, Jarrard DF. Addressing the need for repeat prostate biopsy: new technology and approaches. Nat Rev Urol. 2015;12:435–44.
25. Tosoian JJ, Sessine MS, Trock BJ, Ross AE, Xie C, Zheng Y, et al. MyProstateScore in men considering repeat biopsy: validation of a simple testing approach. Prostate Cancer Prostatic Dis. 2023;26:563–7.

Publisher's Note