Discover Oncology

Research

Analysing DNA methylation and transcriptomic signatures to predict prostate cancer recurrence risk

Fahad M. Aldakheel[1] · Hadeel Alnajran[1] · Shatha A. Alduraywish[2] · Ayesha Mateen[3] · Mohammed S. Alqahtani[3] · Rabbani Syed[3]

Received: 10 August 2024 / Accepted: 17 January 2025
Published online: 01 February 2025

Abstract

Prostate cancer (PCa) remains a significant global health challenge, with approximately 1.6 million new cases and 366,000 deaths annually. Despite high survival rates for localized prostate cancer, recurrence poses a substantial risk due to inherent biological factors and residual disease. Early detection and intervention are essential for enhancing patient outcomes and reducing mortality. However, traditional diagnostics such as PSA tests, digital rectal examinations, and biopsies often lack specificity resulting in overdiagnosis. There is a pressing need for novel biomarkers to enhance precision medicine approaches for PCa. This study employs a machine learning approach to identify DNA methylation and RNA expression biomarkers predictive of PCa recurrence using datasets from The Cancer Genome Atlas (TCGA). We analyzed 49,133 genes, identifying 684 differentially methylated genes (DMGs) and 691 differentially expressed genes (DEGs) between recurrence and non-recurrence groups. Ten genes (TNNI2, SPIN2, COL5A3, RNF169, CCND1, FGFR1, SLC17A2, FAMM71F2, RREB1, AOX1) were found to have significant correlations between methylation and expression, forming the basis for our predictive model. A support vector machine (SVM) model was developed using these ten genes, achieving an area under the curve (AUC) of 0.773, demonstrating robust predictive capability. Multivariate regression analysis confirmed the SVM score as an independent predictor of recurrence (HR = 0.45; 95% CI 0.28–0.69, $P < 0.001$). The analysis of recurrence-free survival suggested that patients with low-risk scores experienced significantly better outcomes compared to those with high-risk scores. Functional enrichment analyses of DMGs revealed significant involvement in biological processes such as transcription regulation, signal transduction, and immune response, highlighting the potential mechanistic pathways of these biomarkers. Validation using real-time PCR confirmed differential expression and methylation patterns of the identified genes in prostate cancer (PC3) and non-cancerous cell lines (PNT2). In conclusion, our study hihglights the DNA methylation biomarkers linked to PCa recurrence and introduces a promising SVM model for early prediction, potentially improving treatment outcomes. Further research is needed to explore the biological roles of these genes in PCa aiming to refine therapeutic approaches.

Keywords Methylation · Expression · Recurrence · Prostate cancer · Cell lines · Machine learning

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-01833-8.

✉ Rabbani Syed, rsyed@ksu.edu.sa; rsyedrabbani@rediffmail.com; Fahad M. Aldakheel, faldakheel@ksu.edu.sa; Hadeel Alnajran, halnajrean@ksu.edu.sa; Shatha A. Alduraywish, saduraywesh@ksu.edu.sa; Ayesha Mateen, amateen.c@ksu.edu.sa; Mohammed S. Alqahtani, msalqahtani@ksu.edu.sa | [1]Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Saud University, P.O. Box 10219, 11433 Riyadh, Saudi Arabia. [2]Department of Family and Community Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia. [3]Department of Pharmaceutics, College of Pharmacy, King Saud University, PO Box 2457, Riyadh-11451, Saudi Arabia.

1 Introduction

Prostate cancer (PCa) has been considered as the largest global health problem among men with about 1.6 million newly diagnosed cases and 366,000 annual deaths globally. It is found second most frequently diagnosed cancer globally while placing fifth among the leading causes of cancer-related death for men. Interestingly, nearly all men detected with localized prostate cancer have a five-year survival rate. [1]. Contrastingly up to approximately 40% of presentations indicate increased probability for recurrence as a result of the inherent biological factors, distant micrometastases from the prostate and localized residual disease despite presentation at what is known as the 'curative' stage, provided by radical prostatectomy [2, 3]. Prompt therapy at recurrence may significantly lower the risk of distant metastasis, extend life and some cases adequately cure. Thus early detection of PCa is important very towards improvement of the patient prognosis and decrease in the mortality rate [4, 5].

Traditional prostate cancer (PCa) diagnosis methods, such as PSA blood tests, DRE, and biopsies, have lacked specificity, leading to overdiagnosis and overtreatment. This highlights the urgent clinical need for novel biomarkers that can predict therapeutic responses and prognosis, potentially transforming PCa treatment through precision medicine and overcoming current diagnostic limitations [6]. Currently, major factors that are considered in deciding adjuvant treatment in cases of high-risk prostate cancer include staging of the tumor, Gleason score, surgical margins, PSA levels, lymph node involvement and patient's preferences [7–10]. These approaches, however, may give relevant information that can be very useful for the clinical management but are not accurate in prediction of cancer recurrence when compared to the use of genetic, epigenetic and molecular biomarkers as explained. It is therefore of fundamental importance that there be development of more improved biomarker sets that will aid more accurately predict recurrence risk and better tailor new treatment approaches in most solid malignancies [11, 12].

DNA methylation, being a key epigenetic modification for the gene transcriptional regulation without altering DNA sequence, is linked with early tumorigenesis. The hypermethylation of tumor suppression genes and the hypomethylation of oncogenes are one of the key early events during early tumorigenesis. [13–17]. Drugs that target these mechanistic targets have been proven to be fruitful in the treatment of prostate cancer. The DNA methylation shows relative stability and has experienced extensive variations across various types of tumor hence influencing the path Physiology, progression, and prognosis of cancer [18, 19]. Other studies investigated the relationship between DNA methylation and prostate cancer, its incidence, gene functions, as well as methylation-based classifications of tumors, and predictors for radiotherapy outcomes [20–24]. Few studies, however, specifically target the development of recurrence prediction model based on DNA methylation in prostate cancer, which is clinically informative to accurately identify the at-risk patients and further devise proactive strategies in clinical treatment This study employs a machine learning approach to analyze DNA methylation and RNA expression data from the Cancer Genome Atlas (TCGA) to identify epigenetic markers predicting prostate cancer recurrence risk, with further validation using prostate cancer specific in vitro- cell line model.

2 Results

2.1 Clinical characteristics of patients

After screening and filtering, 490 patients diagnosed with prostate cancer each of whom comprehensive methylation and clinical records were obtained in TCGA portal. The median age at the time of diagnosis was 60 years (40–80). There was a reoccurrence rate to 15% within five years among the entire patients. The T stage of the patients with prostate cancer was classified, hence ranging between II to IV stages. More precisely, 184 patients (37.55%) were stage I, 286 patients (58.37%) were stage III and 10 patients (2.04%) were stage IV. The racial background of the study participants was as follows, Whites 396 (80.82%), Asians 14 (2.86%), Blacks 55 (11.22%), and undetermined racial backgrounds 20 (4.08%). Based on Gleason scores, patients were stratified into three groups: < 7 (43 patients, 8.78%), $= 7$ (241 patients, 49.18%), and > 7 (196 patients, 40.00%) (Table 1). All the patients were grouped into training and refrence cohorts keeping a proportion of 70:30 to ease in analysis. Figure 1 provides an overview of the study's overall design and workflow.

Table 1 Clinical characteristics of prostate cancer patients

Characteristics	Total Patients (n = 490)		Training set (n = 325)		Testing set (n = 165)	
	n	%	n	%	n	%
Age						
≤ 55	110	22.45	80	24.62	35	21.21
> 55	380	77.55	245	75.38	130	78.79
Race						
White	396	80.82	270	83.08	125	75.76
Asian	14	2.86	11	3.38	5	3.03
Black	55	11.22	29	8.92	26	15.76
Unknown	20	4.08	15	4.62	9	5.45
Gleason score						
< 7	43	8.78	20	6.15	25	15.15
= 7	241	49.18	170	52.31	79	47.88
> 7	196	40.00	135	41.54	61	36.97
T stage						
T2	184	37.55	125	38.46	66	40.00
T3	286	58.37	190	58.46	91	55.15
T4	10	2.04	10	3.08	8	4.85

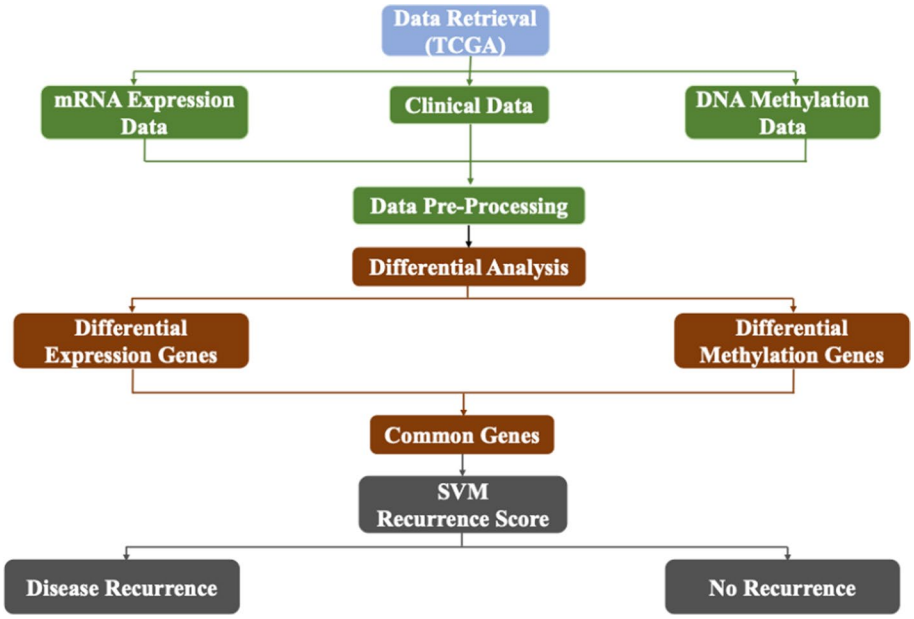

Fig. 1 Study outline for machine learned based identification of gene expression and methylation biomarkers for prostate cancer recurrence risk

2.2 Identification of DMGs and DEGs

To identify significantly differentially methylated genes (DMGs) and differential expression gene (DEG) sets, we explored 22,602 DNA methylated genes and 20,531 RNA expression data sets in the TCGA database between recurrence versus no-recurrence groups (Fig. 2a). Thus, volcano plots represented the division of DNA methylation or gene expression fold changes by each group (Fig. 2a, b). Overall, we defined 684 DMGs ($p < 0.01$; logFC > 0.025) and 691 DEGs ($p < 0.01$; logFC > 0.4). Of these, six genes were found hypermethylated while four were hypomethylated, and they overlapped with the DEGs. It, therefore, implied that a negative correlation ($p < 0.05$) was discovered between DNA methylation and mRNA expression levels in ten of these genes (TNNI2, SPIN2, COL5A3, RNF169, CCND1, FGFR1, SLC17A2, FAMM71F2, RREB1, AOX1). The recurrent prediction model was then developed based on the ten DNA-methylated genes.

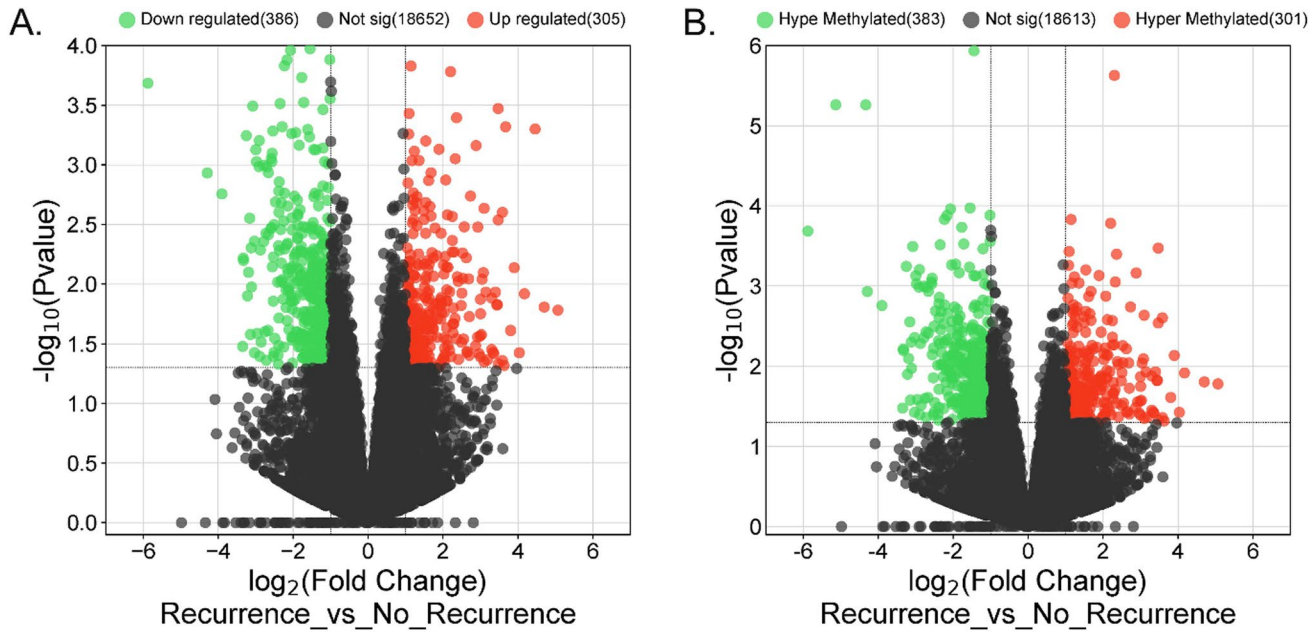

Fig. 2 Selection of DMGs in recurrent and nonrecurrent prostate cancer samples. **a** Volcano plot displaying DMGs. **b** Volcano plot showing DEGs

Here we identified ten genes with significant epigenetic alterations, specifically six hypermethylated and four hypomethylated, all of which overlapped with the differentially expressed genes (DEGs). This overlap highlights a crucial relationship between DNA methylation and gene expression levels. Statistical analysis revealed a significant negative correlation ($p < 0.05$) between DNA methylation and mRNA expression for these genes, indicating that hypermethylation was associated with decreased expression, while hypomethylation correlated with increased expression.

2.3 The SVM-based predictions

We next trained the method of SVM model for recurrence outcome based on ten overlapped genes from both methylation differential expression data. In this, Fig. 3A**,** it is the receiver operating characteristic (ROC) curve with an area under the curve (AUC) of 0.773. Using the below formula hypotheses the recurrence rate (β values) of top DNA-methylated genes using SVM based recurrence construction.

$$\begin{aligned} RecurrenceScore = {} & 0.46 \times \beta_{TNNNI2} - 0.59 \times \beta_{SPIN2} + 0.15 \times \beta_{COL5A3} + 0.35 \times \beta_{RNF169} + 0.10 \times \beta_{CCND1} \\ & + 0.23 \times \beta_{FGFR1} + 0.37 \times \beta_{SLC17A2} + 0.61 \times \beta_{FAMM71F2} + 0.19 \times \beta_{RREB1} + 0.72 \times \beta_{AOX1} \end{aligned}$$

2.4 Regression analyses

On univariate analysis of these patients' SVM scores and clinical features. Univariate analysis of the risk factors, including SVM score, Gleason score, M stage, N stage as well as the pathologic grade ($p < 0.05$). Furthermore, the risk factors were used for multivariate regression analysis, which found that the SVM score (hazard ratio [HR] = 0.45; 95% confidence interval [CI] 0.28–0.69, $P < 0.001$) and N stage (HR = 2.96; 95% CI 1.21–7.31, $P < 0.05$) could independently predict clinical recurrence.

2.5 Recurrence-free survival by methylation-derived recurrence scores

We were testing for the effect of methylation risk score people on prostate cancer subject's RFS. It turned out that there is significant difference in the low-risk score and the high-risk score people in terms of RFS. Figure 3B shows that the

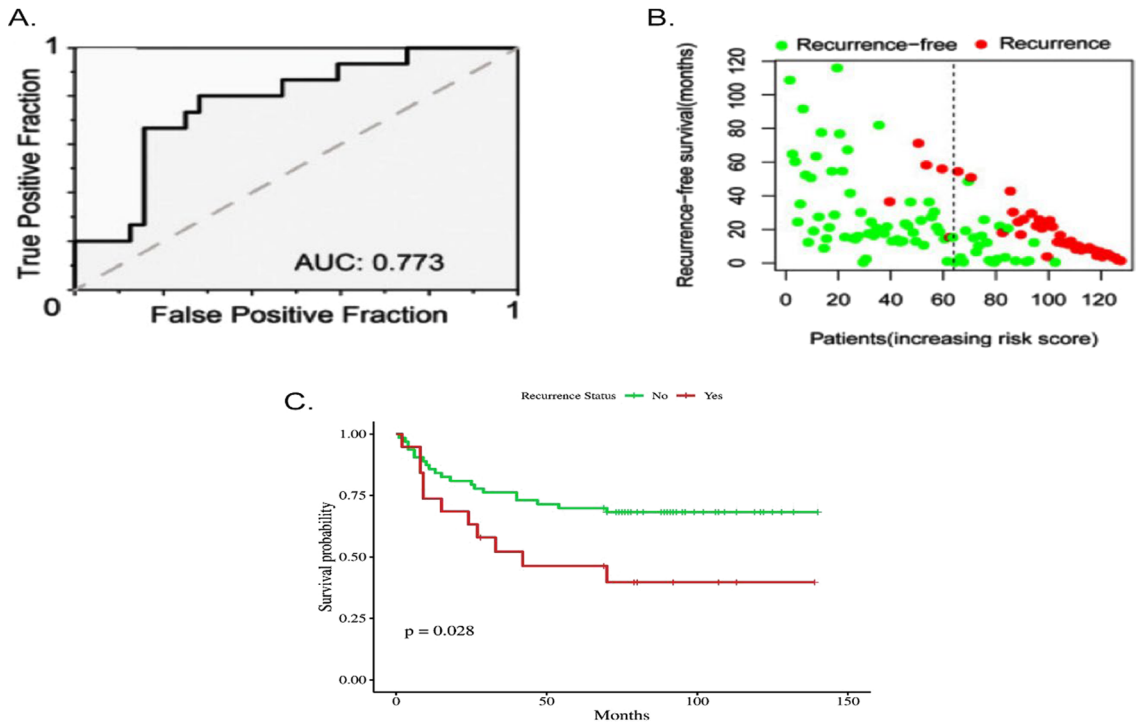

Fig. 3 **A** ROC curve of SVM-based recurrence prediction with the AUC **B** Visualization and correlation DNA methylation-based recurrence score with RFS. Red points represent patients with recurrence, while blue points represent patients without recurrence. **C** Survival analysis of the prostate cancer patients using the prediction model recurrence score

patients defined as low-risk had consistently significantly better RFS than their high-risk counterparts, without consideration of the other differences. This is further clarified by a Kaplan–Meier curve (Fig. 3C) that the model based low-risk group showed an apparently extended recurrence free survival than the high risk.

2.6 GO and KEGG analysis

To examine the molecular functional landscape of DMGs, gene ontology (GO) analysis for biological processes (BPs), molecular functions (MFs), and cellular components (CCs) was conducted using DAVID. MFs enriched protein binding, GTPase activator activity poly(A) RNA binding, and protein homodimerization activity. BPs were enriched in regulation of transcription, signal transduction and positive regulation of activity immune response. CCs were enriched in nucleoplasm membrane plasma membrane, and cytosol. Furthermore, KEGG analysis results indicated enrichment of DMGs in pathways that related to human T-lymphotropic virus 1 infection, cytokine–cytokine receptor interaction, T-cell receptor signaling (TCR) pathway and natural killer cell-mediated cytotoxicity (Figure S1).

2.7 Findings of promoter methylation and realtime PCR

In this study, we confirmed the differential expression of ten genes (TNNI2, SPIN2, COL5A3, RNF169, CCND1, FGFR1, SLC17A2, FAMM71F2, RREB1, and AOX1) through quantitative real-time PCR (qRT-PCR). The gene expression levels in PC3 cells were compared to non-cancerous PNT2 cells, and the LogFC values were calculated to reflect fold changes in gene expression between the two cell types. The analysis revealed that TNNI2 (LogFC = 1.68), SPIN2 (LogFC = 1.54), COL5A3 (LogFC = 2.42), SLC17A2 (LogFC = 1.92), and AOX1 (LogFC = 1.25) were significantly downregulated in PC3 cells compared to PNT2 cells. In contrast, RNF169 (LogFC = -0.66), CCND1 (LogFC = – 1.47), FGFR1 (LogFC = -1.74), FAMM71F2 (LogFC = – 1.52), and RREB1 (LogFC = -1.43) were upregulated in PC3 cells. These results highlight notable differences in gene expression between prostate cancer (PC3) and non-cancerous prostate (PNT2)

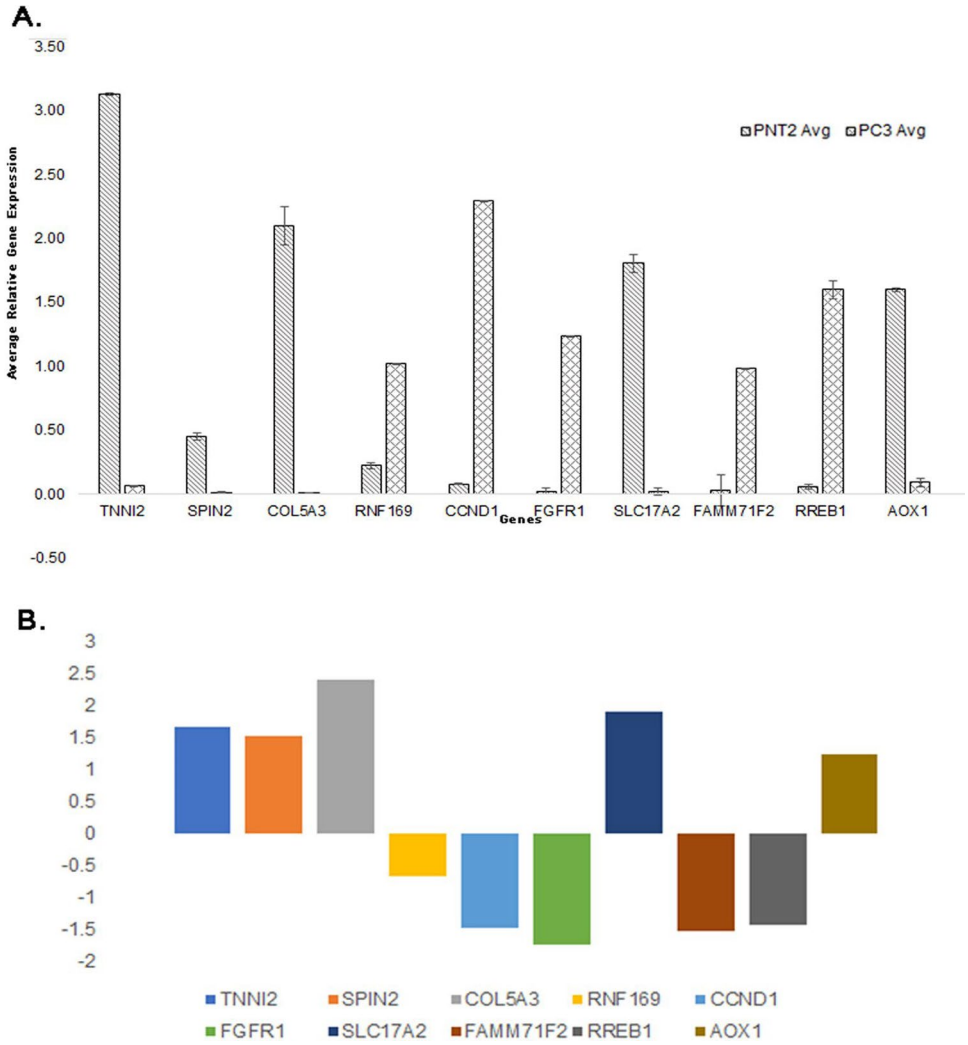

Fig. 4 qRT-PCR Analysis of Gene Expression in PC3 and PNT2 Cells. **A** Relative gene expression levels of TNNI2, SPIN2, COL5A3, RNF169, CCND1, FGFR1, SLC17A2, FAMM71F2, RREB1, and AOX1 in PC3 and PNT2 cells. Error bars show standard deviation. **B** Log fold changes (LogFC) of gene expression in PC3 vs. PNT2 cells. Positive values indicate upregulation, and negative values indicate downregulation in PC3 cells

cells, suggesting that these genes may play significant roles in the progression or suppression of prostate cancer. The log fold change (LogFC) values reflect the magnitude of expression differences, with higher positive values indicating greater downregulation in PC3 cells and negative values indicating upregulation (Fig. 4AB).

2.8 Validation of differentially expressed and methylated genes in prostate cancer cells using qRT-PCR

In this study, ten genes (TNNI2, SPIN2, COL5A3, RNF169, CCND1, FGFR1, SLC17A2, FAM71F2, RREB1, and AOX1), initially identified through TCGA analysis, were further validated for their differential expression at the transcript level and methylation patterns using quantitative real-time PCR (qRT-PCR). The gene expression trends closely mirrored the LogFC values observed from the microarray data. Specifically, TNNI2, SPIN2, COL5A3, AOX1, and SLC17A2 showed significant upregulation in PC3 cells compared to PNT2 cells, as shown in Fig. 5. In contrast, RNF169, CCND1, FGFR1, FAM71F2, and RREB1 were downregulated. These findings underscore notable gene expression differences between PC3 and PNT2 cells (Fig. 5A), highlighting the differential expression profiles of these genes.

Moreover, methylation profiles, as indicated by CT values, revealed clear hypermethylation of key genes like COL5A3, FGFR1, and SLC17A2 in PC3 cells (Fig. 2). This pattern aligns well with the microarray LogFC values, further corroborating our earlier findings. In summary, the data emphasize the significant role these genes play in distinguishing between Prostate Cancer (PC3) and Non-Cancerous (PNT2) cell lines.

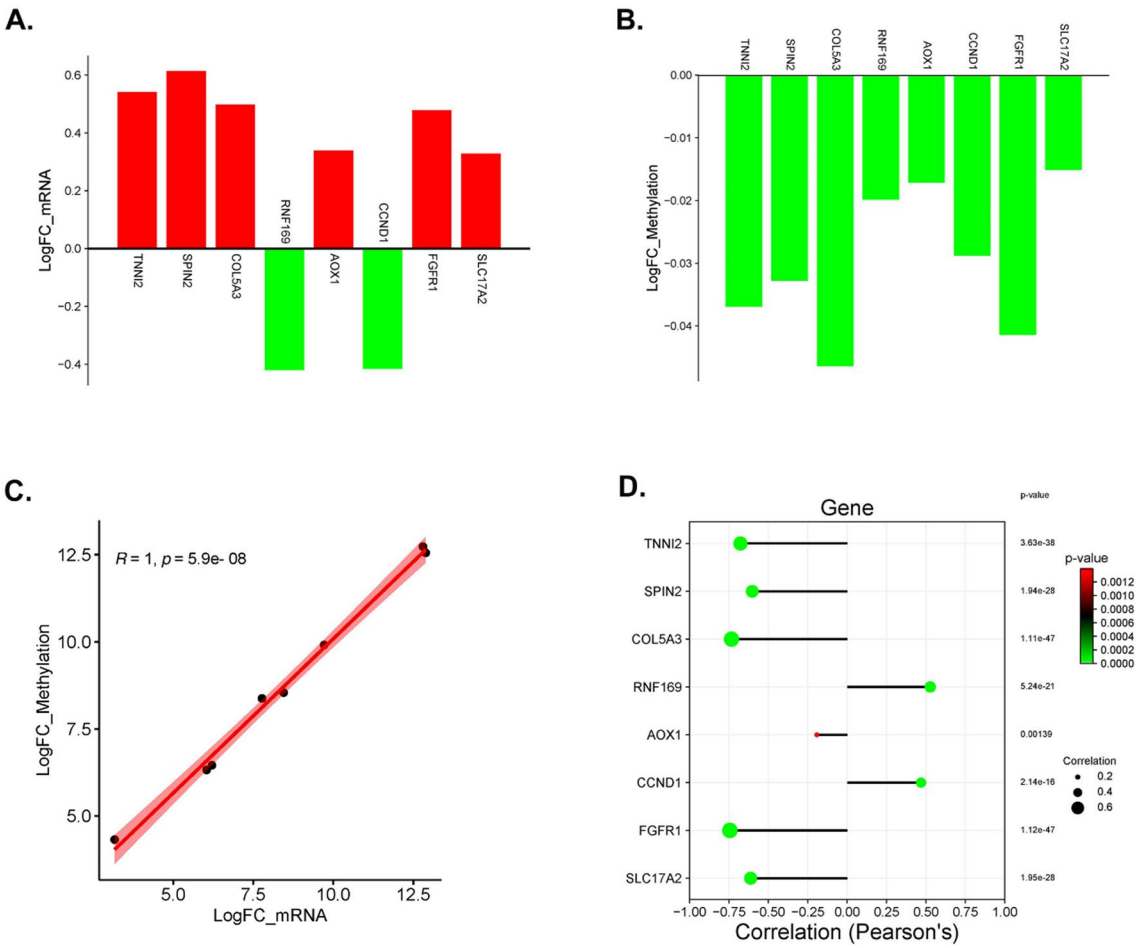

Fig. 5 Gene Expression Analysis of PC3 vs. PNT2 cell lines. **A** LogFC values of gene expression in PC3 vs. PNT2, **B** LogFC values of methylation in PC3 vs. PNT2 Methylation in PC3 vs. PNT2, **C** R squared values showing the strength of relation between methylation versus gene expression data **D**. Pearson's values showing the linear relationship between methylation and expression levels

Methylation analysis similarly showed distinct LogFC values between PC3 and PNT2 cells (Fig. 5B), with specific regions exhibiting increased or decreased methylation in PC3 cells. Correlation analysis revealed varying strengths of association between methylation and gene expression, as reflected in the R-squared values. Some gene-methylation pairs showed higher R-squared values, indicating a strong relationship between methylation changes and gene expression. Pearson's correlation coefficients further illustrated the linear relationship between methylation and gene expression levels, with both positive and negative correlations highlighting how alterations in methylation correspond to shifts in gene expression (Figs. 5C, D).

3 Materials and methods

3.1 Data collection and processing

For this study, clinical datasets for DNA methylation (494 samples), RNA expression (samples 493), and from The Cancer Genome Atlas (TCGA) clinical information were downloaded using the FireBrowse Data Portal. Samples without survival data were essentially excluded but were included based on disease-free survival and recurrence status.

Table 2 Correlation analysis between DNA methylation and mRNA expression levels in ten of these genes

Gene	LogFC_Methylation	LogFC_mRNA	r2 (R-sqaured)	p-value	Correlation (Pearson's)
TNNI2	– 0.03692	0.541041	0.4590612	3.63E– 38	– 0.6775406
SPIN2	– 0.0328	0.613551	0.3627484	1.94E– 28	– 0.602286
COL5A3	– 0.04641	0.498091	0.5392871	1.11E– 47	– 0.7343617
RNF169	– 0.01987	– 0.42003	0.2779979	5.24E– 21	0.527255
AOX1	– 0.01717	0.338443	0.0369399	0.00139	– 0.1921976
CCND1	– 0.02881	– 0.41618	0.2200395	2.14E– 16	0.4690837
FGFR1	– 0.04142	0.478092	0.5592871	1.12E– 47	– 0.7443617
SLC17A2	– 0.01512	0.328353	0.5627484	1.95E– 28	– 0.612286
FAMM71F2	– 0.02519	0.438450	0.0399399	0.00119	– 0.1821976
RREB1	– 0.05652	0.459097	0.2100395	2.24E– 16	0.5690837

3.2 Analysis of differential methylation and expression

TCGA DNA methylation and RNA expression datasets were studied with R (version 3.6.0). We identified differentially methylated genes (DMGs) using a p-value < 0.01 and an absolute log fold change >|0.025| to identify significant changes in methylation levels (β values). Similarly, differentially expressed genes (DEGs) observed where RNA expression levels of DEGs showed significant absolute log fold change more than 0.4 with a p-value less than 0.01. We intended to identify in the top 1% percentile. Correlation analysis performed on DMGs and DEGS in R (limma function). The SVM prediction model was built by selecting the significant DEMGs and DEGs (based on negative correlation) (Table 2).

3.3 SVM model

We applied the SVM, is a machine-learning procedure implemented in binary classification. This set of algorithms handles data in a high-dimensional space to come up with the optimal hyperplane for class separation. The SVM score was estimated via their coefficients, values of the ten significant DMGs β, and a constant bias. The two classes of scores, where the SVM score values were > 0 or ≤ 0, classified patients as high risk and low risk respectively. The svm function from the e1071 package was used for classification. A confusion matrix was created to validate predictions of the SVM classifier model. Various performance measures were for the evaluation of the model such as a receiver-operating characteristic (ROC) curve and the area under the curve (AUC). We employed the t-tests and the Fisher's exact tests to determine those clinical features that significantly influenced recurrence, with only those conditions substantial at $p < 0.05$ being maintained for multivariate regression assessment.

3.4 Functional enrichment analysis

The Gene Ontology (GO) enrichment analysis, namely cellular component, biological process, and molecular function as well as a Kyoto Encyclopedia of Genes and Genomes pathway (KEGG) enrichment analysis for DMGs, using the Database for Annotation, Visualization, and Integrated Discovery (DAVID). Enrichment was considered significant at $p < 0.05$.

3.5 Comparative analysis of gene expression and methylation for prostate cancer PC3 and non-cancerous PNT2 cell lines using real-time quantitative PCR

The PC3 prostate cancer cell lines were cultured in RPMI-1640 medium, supplemented with 10% FBS and 1% penicillin/streptomycin, in a humidified incubator at 37 °C with 5% CO2. The benign prostate cell line PNT2, used as a control, was cultured in Keratinocyte-SFM supplemented with human EGF and bovine pituitary extract. After overnight adherence of both PNT2 and PC3 cells, the cell density of PC3 was adjusted to 2 × 10^5 cells/well, and experimental treatments were conducted for an additional 24 h. The PNT2 cells were maintained in their untreated state as controls. DNA was extracted from the treated samples using the Qiagen "QIAamp DNA Mini Kit" protocol to ensure preservation for subsequent

methylation analysis. Total RNA was isolated using the RNeasy Mini Kit, and its quantity and purity were assessed using a NanoDrop spectrophotometer.

Reverse transcription of total RNA to cDNA was carried out using the SuperScript III First-Strand Synthesis System for gene expression analysis. Gene expression was analyzed using real-time PCR on a StepOnePlus Real-Time PCR System, utilizing TaqMan Gene Expression Assays for selected genes (TNNI2, SPIN2, COL5A3, RNF169, CCND1, FGFR1, SLC17A2, FAM71F2, RREB1, and AOX1) with GAPDH as the reference gene. Delta CT values were calculated by subtracting the mean GAPDH CT value from the target gene CT values. The 2^-ΔCT method was used to determine gene expression levels, while logFC represented the difference in expression between treated and control cells. Statistical analysis was performed using GraphPad Software.

The methylation-sensitive real-time PCR method further refined the detection of DNA methylation in the targeted gene promoters. In this approach, bisulfite conversion successfully converted unmethylated cytosines into uracils within the pre-isolated DNA, followed by real-time PCR using the StepOnePlus System. Custom primers were designed for the promoters of the selected genes (TNNI2, SPIN2, COL5A3, RNF169, CCND1, FGFR1, SLC17A2, FAM71F2, RREB1, and AOX1). Methylation levels were quantified by comparing the melting temperatures (Tms) of PCR products to known standards, demonstrating a direct correlation between Tm and methylation percentage. Exact methylation levels were determined through Ct value analysis, with assay accuracy confirmed using controls and reference genes. The correlation between gene expression and methylation was calculated, and statistical significance ($p < 0.05$) was assessed using Pearson analysis in GraphPad Software.

Using Pearson analysis, the correlation between gene expression and methylation was determined. The correlation coefficient (r) indicated the strength of the relationship, illustrating how methylation influences gene expression. Statistical analysis was performed using GraphPad Software, with significance defined as $p < 0.05$.

4 Discussion

One of the essential epigenetic mechanisms, DNA methylation, has been emphasized playing a role in multiple biological processes of the human genome, such as gene expression regulation, cell differentiation, and inflammation [13–16]. DNA methylation takes place at the CpG sites that in normal cases are un-methylated or weakly methylated [25, 26]. The hypermethylation of these CpG sites, however, may lead to the inactivation of tumor suppressor genes and subsequently predispose an individual to cancer. For instance, a key clinical goal is therefore to identify abnormal methylation at certain genes to assist with disease diagnosis, prognosis, and prediction of therapeutic response to personalized therapy against cancer.

Notably, various other cancer types had distinct DNA methylation signatures that had shown great prediction potentials. For instance, 13 DNA methylation signature showed precision and accuracy in the prediction of recurrence-free survival in stage I lung cancer [15, 27]. Similarly, the relevance to predict recurrence-free survival in papillary thyroid cancer using a 6-DNA methylation signature has also been revealed in a recent study [16]. Although a nomogram model incorporating 11 DNA methylation sites and clinicopathological features has been developed [14, 28], there remains no established quantitative method for early recurrence prediction in PCa based on methylation signatures.

In this study, we have completely performed data analysis of the prostate cancer patients within the TCGA dataset. As a result, we compared those patients to recurrent-free ones. Furthermore, using the SVM machine-learning algorithm, and with support from the SVM scores generated from it, we developed a predictive model. On the other hand, recurrence majorly depended on the independent risk factor, determined by multiple Cox regression analysis—the SVM score (Table 3). Traditional tumor staging parameters, including Gleason score, T, and M stages, lacked independent predictive power for recurrence. The SVM recurrence score, however, proved more closely aligned with tumor recurrence than conventional staging and reliably predicted recurrence across various pathologic types and N stages. Through ROC curve, it

Table 3 Stratification analysis of recurrence score in patients

	Univariate analysis			Cox regression analysis	
Characteristic	No recurrence (N = 295)	Recurrence (N = 198)	*P-value*	HR (95% CI)	*P-value*
Recurrence Score (Mean ± SD)	1.42 ± 1.14	−0.23 ± 1.23	2.52E − 11	0.458 (0.240–0.670)	3.02E − 04

was evident that the accuracy of the model was extremely high. Independent datasets from GEO applied our validated model on survival analysis. In a clinical setting, eight DNA methylation biomarkers by clinicians would be able to identify high-risk recurrence patients using this model.

The ten DEGs (Differentially Expressed Genes)—TNNI2, SPIN2, COL5A3, RNF169, CCND1, FGFR1, SLC17A2, FAM71F2, RREB1, and AOX1—were found to exhibit distinct methylation profiles among patients with and without disease recurrence. Some of these genes have previously been reported to be associated with cancer and other diseases. For instance, FGFR1 has been linked to metastasis in colorectal carcinoma and malignancy in renal cell carcinoma. High expression of TNNI2 has been associated with peritoneal metastasis in gastric cancer. SPIN2 is recognized as a promoter of the tumor suppressor gene that inhibits proteases involved in cancer progression. CCND1 is a critical regulator of the cell cycle, playing an essential role in the tumorigenesis process through uncontrolled proliferation [29, 30]. These findings are consistent with established evidence that DNA methylation typically silences gene expression by impeding transcription factor binding or recruiting repressive complexes [31, 32]. The ten identified genes—TNNI2, SPIN2, COL5A3, RNF169, CCND1, FGFR1, SLC17A2, FAM71F2, RREB1, and AOX1—were notable for their roles in cellular processes. For instance, CCND1 (Cyclin D1) is known to regulate the cell cycle and has been implicated in various cancers, while FGFR1 (Fibroblast Growth Factor Receptor 1) plays a role in signaling pathways critical for cell proliferation and differentiation, with alterations linked to tumorigenesis [33]. This supports the hypothesis that epigenetic changes in these genes may contribute to disease pathology.

Building upon this discovery, the researchers developed a recurrent prediction model using these ten DNA-methylated genes. This model integrates the methylation-expression dynamics to improve predictions of gene expression patterns and disease outcomes. Such approaches align with broader efforts to leverage epigenetic data for diagnostic and therapeutic advancements in precision medicine [34].

While in normal cells the expression of CCND1 is strictly regulated, in cancer the heightened activity of CCND1 occurs through myriad mechanisms. Although AOX1-related metabolites were shown to predict advanced bladder cancer [15], this gene was implicated in the pathogenesis of colorectal cancer by participating in the transcriptional activation of CD133 through phosphoinositide 3-kinase (PI3K)/Akt pathway [35]. Furthermore, AOX1 gene expression levels have shown a correlation with prostate cancer recurrence [36]. Although roles of most of these genes in prostate cancer are uncharted territories, there remains an urgent eligibility for further feed to their biological functions and mechanisms in the context of prostate cancer. The current study contained certain limitations. For example, in differentiating recurrence of prostate cancer from those that did not recur, patients without a recurrence and whose follow-up duration was very short had to be eliminated. Two adverse results would be this had reduced sample size and patients who did not have recurrence in the first 2 years after treatment were categorized as "no recurrence" group, which could introduce some degree of bias to be associated due to limited follow-up duration. These genes may be potential biomarkers and therapeutic targets for prostate cancer. More mechanistic investigations should be done to reveal what functions and mechanisms mediated by these genes in prostate cancer.

In conclusion, our study identified ten DNA methylation biomarkers significantly associated with prostate cancer recurrence. We developed a machine learning model for early and efficient recurrence prediction using a limited number of biomarkers, which could aid in proactive treatment planning for high-risk patients in clinical settings.

Acknowledgements The authors extend their appreciation to the Deputyship for Research & Innovation, "Ministry of Education" in Saudi Arabia for funding this research work through the project number (IFKSUDR_H206).

Author contributions FMA, HA and RS conceived, designed, and supervised the research. HA, SAA, AM and MSA collected the related literature. FMA and RS wrote the manuscript. The final draft of the work has been read and approved by all writers.

Data availability The data that support the findings of this study are available from the authors but restrictions apply to the availability of these data, which were used under license from the Natural History Museum (London) for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

1. Wang L, Lu B, He M, Wang Y, Wang Z, Du L. Prostate cancer incidence and mortality: global status and temporal trends in 89 countries from 2000 to 2019. Front Public Health. 2022;10: 811044. https://doi.org/10.3389/fpubh.2022.811044.
2. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, Fossati N, Gross T, Henry AM, Joniau S, et al. EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol. 2017;71:618–29. https://doi.org/10.1016/j.eururo.2016.08.003.
3. Sanda MG, Cadeddu JA, Kirkby E, Chen RC, Crispino T, Fontanarosa J, Freedland SJ, Greene K, Klotz LH, Makarov DV, et al. Clinically localized prostate cancer: AUA/ASTRO/SUO Guideline. Part I: Risk Stratification, Shared Decision Making, and Care Options. J Urol. 2018;199:683–90. https://doi.org/10.1016/j.juro.2017.11.095.
4. Brockman JA, Alanee S, Vickers AJ, Scardino PT, Wood DP, Kibel AS, Lin DW, Bianco FJ Jr, Rabah DM, Klein EA, et al. Nomogram predicting prostate cancer-specific mortality for men with biochemical recurrence after radical prostatectomy. Eur Urol. 2015;67:1160–7. https://doi.org/10.1016/j.eururo.2014.09.019.
5. Awan Z, Kutbi HI, Ahmad A, Syed R, Alsulaimany FAS, Shaik NA. Molecular design, synthesis and biological characterization of novel Resveratrol derivative as potential anticancer agent targeting NF-κB. J Appl Biomed. 2020;18:8–17. https://doi.org/10.32725/jab.2020.001.
6. Ilic D, Djulbegovic M, Jung JH, Hwang EC, Zhou Q, Cleves A, Agoritsas T, Dahm P. Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. BMJ. 2018;362: k3519. https://doi.org/10.1136/bmj.k3519.
7. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, Tomaszewski JE, Renshaw AA, Kaplan I, Beard CJ, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA. 1998;280:969–74. https://doi.org/10.1001/jama.280.11.969.
8. Cooperberg MR, Broering JM, Carroll PR. Time trends and local variation in primary treatment of localized prostate cancer. J Clin Oncol. 2010;28:1117–23. https://doi.org/10.1200/jco.2009.26.0133.
9. Spratt DE, Zhang J, Santiago-Jiménez M, Dess RT, Davis JW, Den RB, Dicker AP, Kane CJ, Pollack A, Stoyanova R, et al. Development and validation of a novel integrated clinical-genomic risk group classification for localized prostate cancer. J Clin Oncol. 2018;36:581–90. https://doi.org/10.1200/jco.2017.74.2940.
10. Chang AJ, Autio KA, Roach M. High-risk prostate cancer-classification and therapy. Nat Rev Clin Oncol. 2014;11:308–23. https://doi.org/10.1038/nrclinonc.2014.68.
11. Porzycki P, Ciszkowicz E. Modern biomarkers in prostate cancer diagnosis. Central Eur J Urol. 2020;73:300–6. https://doi.org/10.5173/ceju.2020.0067R.
12. Sahly NN, Banaganapalli B, Sahly AN, Aligiraigri AH, Nasser KK, Shinawi T, Mohammed A, Alamri AS, Bondagji N, Elango R, et al. Molecular differential analysis of uterine leiomyomas and leiomyosarcomas through weighted gene network and pathway tracing approaches. Systems Biol Reprod Med. 2021;67:209–20. https://doi.org/10.1080/19396368.2021.1876179.
13. Guerra J, Pereira BMS, Cruz J, Scherer NM, Furtado C, Montalvão de Azevedo R, Oliveira PSL, Faria P, Boroni M, de Camargo B, et al. Genes controlled by DNA methylation are involved in wilms tumor progression. Cells. 2019. https://doi.org/10.3390/cells8080921.
14. Ma X, Cheng J, Zhao P, Li L, Tao K, Chen H. DNA methylation profiling to predict recurrence risk in stage I lung adenocarcinoma: development and validation of a nomogram to clinical management. J Cell Mol Med. 2020;24:7576–89. https://doi.org/10.1111/jcmm.15393.
15. Vantaku V, Putluri V, Bader DA, Maity S, Ma J, Arnold JM, Rajapakshe K, Donepudi SR, von Rundstedt FC, Devarakonda V, et al. Epigenetic loss of AOX1 expression via EZH2 leads to metabolic deregulations and promotes bladder cancer progression. Oncogene. 2020;39:6265–85. https://doi.org/10.1038/s41388-019-0902-7.
16. Chen H, Ma X, Yang M, Wang M, Li L, Huang T. A methylomics-associated nomogram predicts recurrence-free survival of thyroid papillary carcinoma. Cancer Med. 2020;9:7183–93. https://doi.org/10.1002/cam4.3388.
17. Vaidya S, Shaik NA, Latha M, Chava S, Mohiuddin K, Yalla A, Rao KP, Kodati VL, Hasan Q. No evidence for the role of somatic mutations and promoter hypermethylation of FH gene in the tumorigenesis of nonsyndromic uterine leiomyomas. Tumour Biol. 2012;33:1411–8. https://doi.org/10.1007/s13277-012-0391-6.
18. Kulis M, Merkel A, Heath S, Queirós AC, Schuyler RP, Castellano G, Beekman R, Raineri E, Esteve A, Clot G, et al. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. Nat Genet. 2015;47:746–56. https://doi.org/10.1038/ng.3291.
19. Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, Flagg K, Hou J, Zhang H, Yi S, et al. DNA methylation markers for diagnosis and prognosis of common cancers. Proc Natl Acad Sci USA. 2017;114:7414–9. https://doi.org/10.1073/pnas.1703577114.
20. Jones PA, Laird PW. Cancer epigenetics comes of age. Nat Genet. 1999;21:163–7. https://doi.org/10.1038/5947.
21. Ai T, Zhang J, Wang X, Zheng X, Qin X, Zhang Q, Li W, Hu W, Lin J, Chen F. DNA methylation profile is associated with the osteogenic potential of three distinct human odontogenic stem cells. Signal Transduct Target Ther. 2018;3:1. https://doi.org/10.1038/s41392-017-0001-6.
22. Pu Y, Li C, Yuan H, Wang X. Identification of prostate cancer specific methylation biomarkers from a multi-cancer analysis. BMC Bioinf. 2021;22:492. https://doi.org/10.1186/s12859-021-04416-w.

23. Zubeda S, Kaipa PR, Shaik NA, Mohiuddin MK, Vaidya S, Pavani B, Srinivasulu M, Latha MM, Hasan Q. Her-2/neu status: a neglected marker of prognostication and management of breast cancer patients in India. Asian Pacific J Cancer Prevent. 2013;14:2231–5. https://doi.org/10.7314/apjcp.2013.14.4.2231.
24. Morgan SC, Holmes OE, Craig J, Grimes S, Malone S. Long-term outcomes of prostate radiotherapy for newly-diagnosed metastatic prostate cancer. Prostate Cancer Prostatic Dis. 2021;24:1041–7. https://doi.org/10.1038/s41391-021-00339-y.
25. Hou X, Zhang Y, Han S, Hou B. A novel DNA methylation 10-CpG prognostic signature of disease-free survival reveal that MYBL2 is associated with high risk in prostate cancer. Expert Rev Anticancer Ther. 2020;20:1107–19. https://doi.org/10.1080/14737140.2020.1838280.
26. Masoodi TA, Shaik NA, Burhan S, Shafi G, Talluri VR. RNA-Seq reveals skipping of exon 3 in a breast cancer patient carrying G118D PIK3CA mutation. Gene Reports. 2020;20: 100704. https://doi.org/10.1016/j.genrep.2020.100704.
27. Masoodi TA, Banaganapalli B, Vaidyanathan V, Talluri VR, Shaik NA. Computational analysis of breast cancer GWAS loci identifies the putative deleterious effect of STXBP4 and ZNF404 gene variants. J Cell Biochem. 2017;118:4296–307. https://doi.org/10.1002/jcb.26080.
28. Bima AI, Elsamanoudy AZ, Alamri AS, Felimban R, Felemban M, Alghamdi KS, Kaipa PR, Elango R, Shaik NA, Banaganapalli B. Integrative global co-expression analysis identifies key microRNA-target gene networks as key blood biomarkers for obesity. Minerva Med. 2022;113:532–41. https://doi.org/10.23736/S0026-4806.21.07478-4.
29. Ajabnoor G, Alsubhi F, Shinawi T, Habhab W, Albaqami WF, Alqahtani HS, Nasief H, Bondagji N, Elango R, Shaik NA, et al. Computational approaches for discovering significant microRNAs, microRNA-mRNA regulatory pathways, and therapeutic protein targets in endometrial cancer. Front Genet. 2022;13:1105173. https://doi.org/10.3389/fgene.2022.1105173.
30. Shinawi T, Nasser KK, Moradi FA, Mujalli A, Albaqami WF, Almukadi HS, Elango R, Shaik NA, Banaganapalli B. A comparative mRNA- and miRNA transcriptomics reveals novel molecular signatures associated with metastatic prostate cancers. Front Genet. 2022;13:1066118. https://doi.org/10.3389/fgene.2022.1066118.
31 Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007;128(4):683–92.
32. Robertson KD. DNA methylation and human disease. Nat Rev Genet. 2005;6(8):597–610.
33. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet. 2007;8(4):286–98.
34 Feinberg AP, Tycko B. The history of cancer epigenetics. Nat Rev Cancer. 2004;4(2):143–53.
35. Stefani C, Miricescu D, Stanescu-Spinu I-I, Nica RI, Greabu M, Totan AR, Jinga M. Growth factors, PI3K/AKT/mTOR and MAPK signaling pathways in colorectal cancer pathogenesis: where are we now? Int J Mol Sci. 2021;22:10260.
36. Li W, Middha M, Bicak M, Sjoberg DD, Vertosick E, Dahlin A, Häggström C, Hallmans G, Rönn AC, Stattin P, et al. Genome-wide scan identifies role for AOX1 in prostate cancer survival. Eur Urol. 2018;74:710–9. https://doi.org/10.1016/j.eururo.2018.06.021.