Salud, Ciencia y Tecnología. 2025; 5:1644
doi: 10.56294/saludcyt20251644

ORIGINAL

Development of a risk model for BCR-free prognosis in prostate cancer patients linked to nucleotide metabolism

Desarrollo de un modelo de riesgo para el pronóstico sin BCR en pacientes con cáncer de próstata relacionado con el metabolismo de nucleótidos

Wei Liu[1,2], Hong Zhi[3], Lai Fu Han[4], Shiirevnyamba Avirmed[1] ✉, Bilegtsaikhan Tsolmon[1], Ganbayar Batmunkh[1], Jianlong Weng[2], Ling Qin[2], Tao Feng[2], Naranjargal Davaadorj[1], Munkhtuya Tumurkhuu[1,5], Pingping Wang[6]

[1]Graduate School, Mongolian National University of Medical Sciences. Ulaanbaatar, Mongolian, 11000.
[2]Kezuo Houqi People's Hospital. TongLiao, Inner Mongolia, China, 028100.
[3]Affiliated Hospital of Inner Mongolia University for Nationalities. TongLiao, Inner Monglia, China, 028000.
[4]Inner Mongolia Baotou City Cancer Hospital. Inner Mongolia, China, 014000.
[5]East Carolina University. NS, USA, 27858.
[6]The Second people's Hospital of Horqin District. TongLiao, Inner Mongolia, China, 028000.

Citar como: Liu W, Tumurkhuu M, Davaadorj N, Feng T, Qin L, Weng J, et al. Development of a risk model for bcr-free prognosis in prostate cancer patients linked to nucleotide metabolism. Salud, Ciencia y Tecnología. 2025; 5:1644. https://doi.org/10.56294/saludcyt20251644

Enviado: 26-09-2024 **Revisado:** 13-12-2024 **Aceptado:** 26-02-2025 **Publicado:** 27-02-2025

Editor: Prof. Dr. William Castillo-González

Autor para la correspondencia: Shiirevnyamba Avirmed ✉

ABSTRACT

Introduction: prostate cancer is a prevalent malignancy among elderly men, with bioinformatics playing a crucial role in advancing diagnosis and treatment paradigms. Recent studies have highlighted the significance of nucleotide metabolism (NM) in Prostate cancer development and progression, linking it to aggressive cancer phenotypes characterized by uncontrolled proliferation and metastasis. Understanding NM-related genes (NMRGs) could provide insights into Prostate cancer pathogenesis and therapeutic targets.
Method: this paper analyzed TCGA-PRAD and GSE70769 datasets to identify critical modules associated with NMRGs using weighted gene co-expression network analysis (WGCNA). Differentially expressed genes (DEGs) between Prostate cancer and control samples were extracted from the TCGA-PRAD dataset, with overlaps identified as NM-related DEGs (DE-NMRGs). A biochemical recurrence (BCR)-free risk model was constructed from 396 Prostate cancer samples, and patients were classified into high- and low-risk groups based on median risk scores. A nomogram model integrating key prognostic factors was developed to predict BCR rates.
Results: this paper identified 5 prognostic genes: RGS11, KAT2A, MXD3, TARBP1, and WFIKKN. The low-risk group exhibited significantly higher BCR-free survival rates, ESTIMATE scores, and immunophenoscore (IPS) scores compared to the high-risk group. Additionally, potential therapeutic agents, including KU-55933 and Wee1 inhibitors, were proposed.
Conclusions: the identified prognostic genes present promising targets for Prostate cancer diagnosis and treatment, emphasizing their importance in predicting biochemical recurrence and tailoring personalized therapeutic strategies for patients.

Keywords: Prostate Cancer; Risk Model; Nucleotide Metabolism; Prognostic Genes; Prostate Cancer; Risk Prediction; Nucleotides; Genes.

RESUMEN

Introducción: el cáncer de próstata es una neoplasia maligna prevalente en hombres de edad avanzada, y la bioinformática juega un papel crucial en el avance de los paradigmas de diagnóstico y tratamiento. Estudios

recientes han destacado la importancia del metabolismo de nucleótidos (NM) en el desarrollo y la progresión del CaP, vinculado a fenotipos de cáncer agresivo caracterizado por la proliferación incontrolada y metástasis. La comprensión de los genes relacionados con el metabolismo de nucleótidos (NMRGs) podría proporcionar información sobre la patogénesis y las dianas terapéuticas.
Método: este documento analizó los conjuntos de datos TCGA-PRAD y GSE70769 para identificar módulos críticos asociados con NMRGs utilizando análisis de redes de coexpresión de genes ponderados (WGCNA). Los genes diferencialmente expresados (DEGs) entre CaP y muestras de control fueron extraídos del conjunto de datos TCGA-PRAD, con superposiciones identificadas como DEGs relacionados con NM (DE-NMRGs). Se construyó un modelo de riesgo libre de recidiva bioquímica (BCR) a partir de 396 muestras de CaP y los pacientes se clasificaron en grupos de riesgo alto y bajo con base en los puntajes de riesgo medio. Se desarrolló un modelo de nomograma que integra factores pronósticos clave para predecir las tasas de BCR.
Resultados: este documento identificó 5 genes pronósticos: RGS11, KAT2A, MXD3, TARBP1 y WFIKKN. El grupo de riesgo bajo presentó tasas de supervivencia libre de BCR, puntajes estimados y puntajes de inmunofenotipo (IPS) significativamente más altos en comparación con el grupo de riesgo alto. Adicionalmente, se han propuesto posibles agentes terapéuticos, incluyendo los inhibidores de KU-55933 y Wee1.
Conclusiones: los genes pronósticos identificados presentan dianas prometedoras para el diagnóstico y tratamiento del CaP, destacando su importancia en la predicción de la recidiva bioquímica y la adaptación de estrategias terapéuticas personalizadas para los pacientes.

Palabras clave: Cáncer de Próstata; Modelo de Riesgo; Metabolismo de Nucleótidos; Genes Pronósticos; Predicción de Riesgos.

INTRODUCTION

Prostate cancer is a common malignant tumor affecting elderly men. In recent years, the detection rate of Prostate cancer has risen significantly, largely due to the widespread implementation of screening method.[1,2] Within the realm of Prostate cancer pathology, chromosomal aberrations play a pivotal role in early stages, leading to abnormal activation of the androgen receptor (AR) signaling pathway.[3,4] This aberrant activation induces alterations in metabolic pathways, thereby accelerating the proliferation of Prostate cancer cells.[5] The primary method for screening Prostate cancer involves a blood test for prostate-specific antigen, often complemented by a digital rectal examination (DRE). However, determining the appropriate threshold level for PSA testing remains a contentious issue due to the risk of overdiagnosis.[6,7] Diagnostic evaluation of suspected Prostate cancer typically involves transrectal ultrasound-guided biopsy of prostate tissue specimens, with subsequent assessment with the Gleason grading system.[8] The Gleason score serves as a crucial prognostic indicator for prostate cancer, with higher scores indicative of more aggressive tumors and poorer prognoses.[9,10]

These findings underscore the imperative of comprehending the pathogenic mechanisms underlying Prostate cancer and highlight the urgent need for the development of more effictive treatment modalities.

Research has revealed that heightened expression of specific genes can profoundly influence nucleotide metabolism NM), a pivotal susceptibility trait implicated in cancer initiation and progression.[11,12] These genes are intricately involved in modulating various facets of nucleotide biology, including synthesis, repair, and degradation, thereby exerting a significant impact on cancer-associated metabolic reprogramming. For instance, certain genes may facilitate the synthetic pathway, boosting nucleotide production, while others may impede nucleotide breakdown, thus reducing their consumption.[13,14] Bioinformatics research stands at the forefront of transforming cancer diagnosis and treatment paradigms.[15,16] Through the integration of high-throughput sequencing data and advanced computational tools, this field offers novel insights into the molecular understanding of cancer.[17,18] By meticulously analyzing vast datasets, bioinformatics identifies emerging biomarkers and prognostic indicators, thereby paving the way for more precise and personalized approaches to cancer care.[19,20] The application of bioinformatics in Prostate cancer research has unveiled a multifaceted understanding of disease onset and progression, elucidating the pivotal roles played by gene mutations, expression profiles, NM, and immune cell infiltration.[21,22] Particularly, NM emerges as a critical driver in Prostate cancer development and evolution, with aberrant nucleotide metabolism, influenced by overexpressed genes, identified as a fundamental susceptibility trait in cancer.[23,24] Cancer cells exhibit aggressive phenotypes, including uncontrolled proliferation, chemotherapy resistance, immune evasion, and metastasis, and are intrinsically associated with increased levels of nuclear material (NM). This metabolic reprogramming not only fuels the relentless growth of tumors but also underpins their resistance to treatment and ability to spread. Notably, nucleotide synthesis inhibitors, among the earliest anti-cancer agents discovered, continue to serve as cornerstone therapeutics across various cancer types. The pervasive upregulation of NM in cancer cells underscores its indispensability in sustaining tumor growth and malignant behaviors.

https://doi.org/10.56294/saludcyt20251644

Utilizing transcriptome data, this study identified genes associated with NM and employed machine learning techniques to develop a risk model for predicting recurrence in Prostate cancer patients post-radical prostatectomy. The constructed model effectively predicts the likelihood of biochemical recurrence in Prostate cancer patients, potentially revealing novel biomarkers for enhanced diagnosis and treatment strategies. By integrating advanced computational methods with transcriptomic insights, this research offers valuable insights into the management of Prostate cancer patients, aiding clinicians in making informed decisions regarding patient care.

METHOD

Experiment information

Place and date of its realization: Mongolia.

Type of study: Bioinformatics analysis.

Ethical parameters: Respect for autonomy; privacy protection; the principle of non-injury; the principle of justice; ethical review and supervision; international cooperation and cultural sensitivity; responsible innovation; academic integrity.

Data acquisition

The TCGA-Prostate Adenocarcinoma (PRAD) including 481 prostate gland of Prostate cancer (Prostate cancer) and 51 normal samples was sourced from TCGA as a training dataset. Thereinto, 396 Prostate cancer samples that contained recurrence information and time were selected to construct a risk model, including 346 biochemical recurrence-free survival (BCRFS) and 50 biochemical recurrences (BCR). The primary tumors tissue of 92 Prostate cancer patients including 47 BCRFS and 45 BCR were screened from the GSE70769 of GEO database (https://www.ncbi.nlm.nih.gov/geo/), and the GSE70769 was considered as a testing dataset based on Illumina HumanHT-12 V4.0 expression beadchip to verify the risk model. Afterward, we used 'Nucleotide metabolism' as a keyword to select nucleotide metabolism-related genes (NMRGs) from the Molecular Signatures Database v7.1 (MSigDB, https://www.gsea-msigdb.org/gsea/msigdb), and 97 NMRGs were obtained. Moreover, 90 NMRGs were obtained from the published literature. After removing duplicates, 104 Nuclear Material-Related Genes (NMRGs) were used in the study.

Identification of DEGs

In TCGA-PRAD, the DEGs were identified between Prostate cancer (n=481) and control (n=51) samples by 'DESeq2 (v. 1.36.0)' R package, and the thresholds were | log2FC | > 1 and Padj. <0,05. The DEGs were displayed by volcano maps and heat maps using the 'ggplot2 (v. 3.3.6)' and 'pheatmap (v. 1.0.12)' R packages.

WGCNA

The 481 Prostate cancer samples were used to perform the WGCNA by the 'WGCNA (v. 1.72.1)' R package. The R2 was closest to 0,8 when the soft-threshold ß = 7. Then, several gene modules were generated by dynamic tree cutting (genes number200). To further explore the modules that highly correlated with NMRGs, the NMRGs scores in TCGA-PRAD were calculated via the ssGSEA algorithm of 'GSVA (v. 1.44.5). The highly correlated two modules were screened based on Spearman's correlation. Finally, we overlapped the DEGs and module genes by 'ggvenn (v. 0.1.9)' R package as nucleotide metabolism-related DEGs (DE-NMRGs) in Prostate cancer patients.

Functional enrichment analysis and construction of PPI networks

The GO and KEGG analysis were used for understanding the DE-NMRGs related biological functions and pathways by the 'clusterProfiler (v. 4.7.1.001)' R package. The GO items and KEGG pathways were filtered out when P > 0,05, and the GO items and KEGG pathways were generated chordal graphs via the 'DOSE (v. 3.26.2)' and 'GOplot (v. 1.0.2)' R packages, respectively. Additionally, the PPI network of DE-NMRGs was constructed using data from the STRING database (http://string.embl.de/) (medium confidence = 0,4).

The prognostic risk models were established and validated

Firstly, univariate COX analysis was performed on data from 396 Prostate cancer samples to identify candidate genes for NMRG. Selection criteria included a Hazard Ratio (HR) ≠ 1 and a significance threshold of P < 0,05. Subsequently, feature genes were further refined based on the fulfillment of the proportional hazards (PH) assumption at P < 0,05. Then, we constructed a LASSO regression analysis to determine the characteristic genes again. Next, prognostic genes were identified through multivariate Cox analysis, forming the basis for constructing the prognostic risk model. To further assess the efficacy of the risk model, ROC curves were generated using data from the TCGA-PRAD and GSE70769 datasets. According to the median risk score, the 396 Prostate cancer patients in the TCGA-PRAD dataset were subsequently divided into low-risk (n=198) and high-risk (n=198) groups. Similarly, the 92 Prostate cancer patients in the GSE70769 dataset were divided into

low-risk (n=46) and high-risk (n=46) groups. Finally, using the 'survminer (v. 0.4.9)' R package (https://cran.r-project.org/package=survminer), Kaplan-Meier (KM) curves of Subsequently, based on the median risk score, the 396 Prostate cancer patients were divided into low- (n=198) and high- (n=198) risk groups in TCGA-PRAD, and the 92 Prostate cancer patients were divided into low- (n=46) and high- (n=46) risk groups in GSE70769. Subsequently, based on the median risk score, the 396 Prostate cancer patients were divided into low- (n=198) and high- (n=198) risk groups in TCGA-PRAD, and the 92 Prostate cancer patients were divided into low- (n=46) and high- (n=46) risk groups in GSE70769. Lastly, using the 'survminer (v. 0.4.9)' R package (https://cran.r-project.org/package=survminer), the BCR probabilities between high- and low-risk groups were compared by Kaplan-Meier (KM) curves in TCGA-PRAD and GSE70769, respectively. Additionally, we computed the survival time of Prostate cancer patients and analyzed the expression levels of prognostic genes in the two risk groups within both the TCGA-PRAD and GSE70769 datasets.

GSEA

The correlation between prognostic genes and all of other genes were counted by Spearman's correction. The correlation coefficients of genes were ranked to perform the GSEA. Based on the 'clusterProfiler (v. 4.7.1.001)' R packages and '3.h.all.v2023.2.Hs.symbols.gmt' from the MSigDB database, the related-pathways were enriched at |NES| > 1 and Padj. < 0,05.

Correlation between clinical indicators and risk scores

The correlation between risk scores and clinical indicators (Age, Gleason, T categories, N categories, and prostate-specific antigen (PSA) level) were compared in two risk groups of TCGA-PRAD. We conducted univariate and multivariate Cox for risk scores and those clinical indicators, and effective factors was screened as independent prognostic. The screening criteria for univariate Cox was HR≠1 and P < 0,05, and multivariate Cox was HR≠1 and P < 0,02. By the 'rms (v. 6.3-0)' R package, the nomogram model was constructed based on the independent prognostic factors. The points corresponding to each independent prognostic predictor were aggregated to calculate the total point, which was then utilized to predict the 1-, 3-, and 5-year BCR. Subsequently, the model was evaluated using calibration curves and ROC curves.

Tumor microenvironment

The ssGSEA algorithm of the 'GSVA (v. 1.44.5)' R package was applied to compute the 28 immune cell scores in 396 Prostate cancer samples. The 28 infiltrating immune cells in the two risk groups of TCGA-PRAD were displayed by 'pheatmap (v. 1.0.12)' R packages. The difference of infiltrating was compared by the Wilcoxon test (P < 0,05). The Spearman's correlation between prognostic genes and 28 immune cell types was analyzed using the 'ggcor (version 0.9.8.1)' R package. Based on the Wilcoxon test, we compared the difference of Immune Scores, Stromal Scores, and Estimate scores between high- and low-risk groups using the 'estimate (v. 1.0.13) R packages. Besides, the 4 immunophenoscore (IPS) scores of Prostate cancer samples were scoured from the cancer immunome atlas.

Drug sensitivity analysis

We downloaded 198 drugs from the GDSC database. The IC50 values of 198 drugs in Prostate cancer patients were calculated via the 'oncoPredict (v. 0.2)' R package and compared between two groups via the Wilcoxon test (P < 0,05) in TCGA-PRAD. The Spearman's correlation between IC50 and risk score was analyzed, and the drugs with |cor| > 0,3 and P < 0,05 were considered potential drugs in Prostate cancer.

ceRNA network

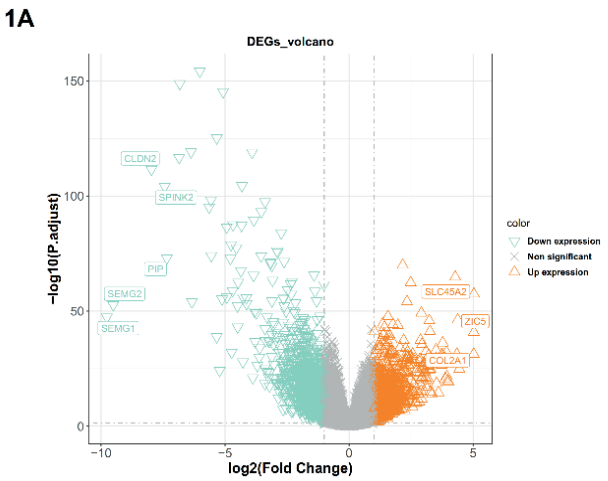
The miRNAs linked with prognostic genes were obtained on the DIANA microT-CDS and miRDB database, respectively. The intersection miRNAs of the two databases were used to generate the miRNA-mRNA pairs. Then, based on the intersection miRNAs, the LncRNAs were predicted on the starBase database as the clipExpNum>70 criteria. Last, the ceRNA network was constructed based on the above relation pairs.

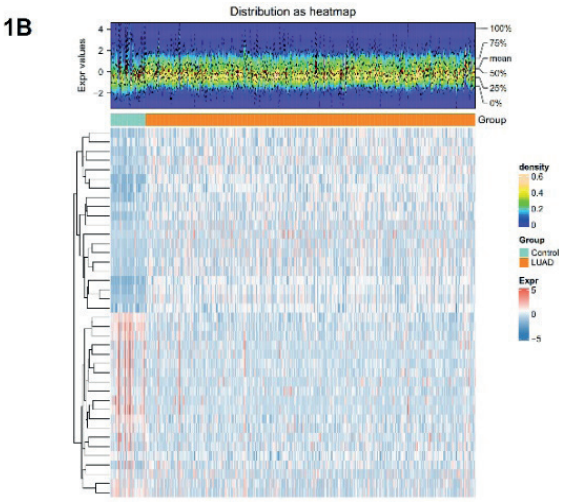
RESULTS

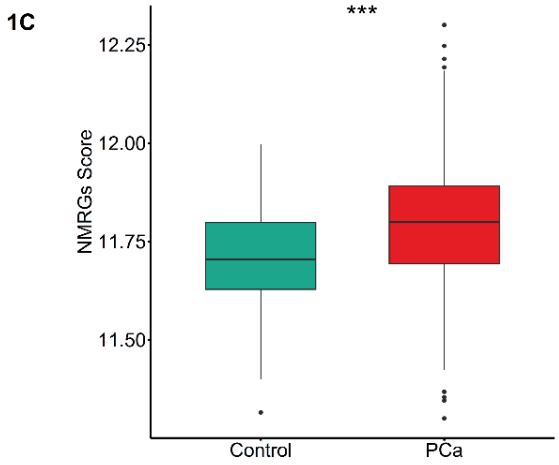
A total of 105 DE-NMRGs were identified in Prostate cancer patients

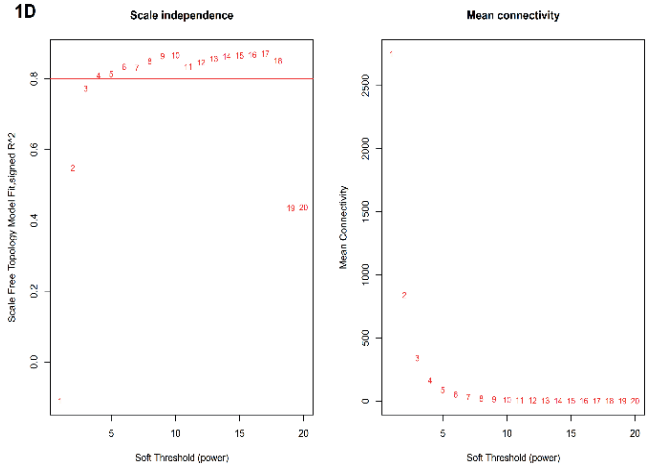
Between Prostate cancer and healthy controls, there were 867 upregulated DEGs and 1,398 downregulated DEGs in TCGA-PRAD (figure 1A, B). The NMRG score was significantly higher in the Prostate cancer samples compared to the healthy controls (P < 0,001) (figure 1C). The WGCNA methodology was employed to identify gene modules associated with NMRGs. Through the clustering analysis of 481 Prostate cancer samples, there was no abnormal sample (figure 1). After determining β = 7, all of the genes were divided into 7 modules (figure 1D, E). Then, the MEpink (528 genes) and MEblack (585 genes) modules that highly correlated with the NMRGs score were selected as key modules, which contained 1,113 module genes (figure 1F). Ultimately, the 2,265 DEGs and

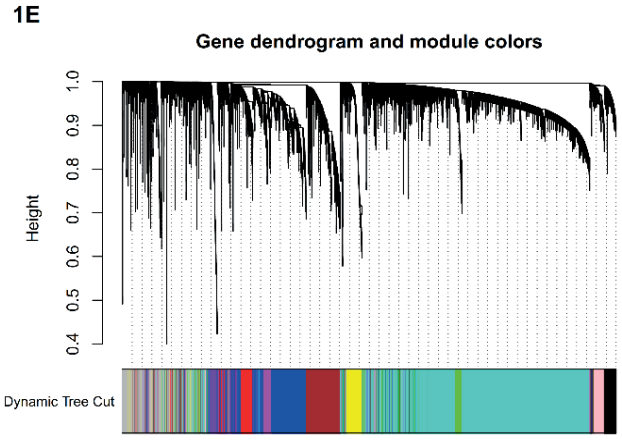
1,113 module genes were overlapped to obtain 105 DE-NMRGs (figure 1G). Additionally, GO results suggested that the DE-NMRGs were enriched into ‘homologous recombination’, ‘reciprocal meiotic recombination’, and ‘reciprocal homologous recombination’, etc (figure 1H, I). KEGG results showed that the DE-NMRGs were linked with the ‘fanconi anemia pathway’, ‘mineral absorption’, ‘steroid hormone biosynthesis’, etc (figure 1J). The PPI network showed that ‘KAT2A-HDAC10’, ‘HGFAC-GSD1’, etc. had a strong interaction (figure 1K).











https://doi.org/10.56294/saludcyt20251644

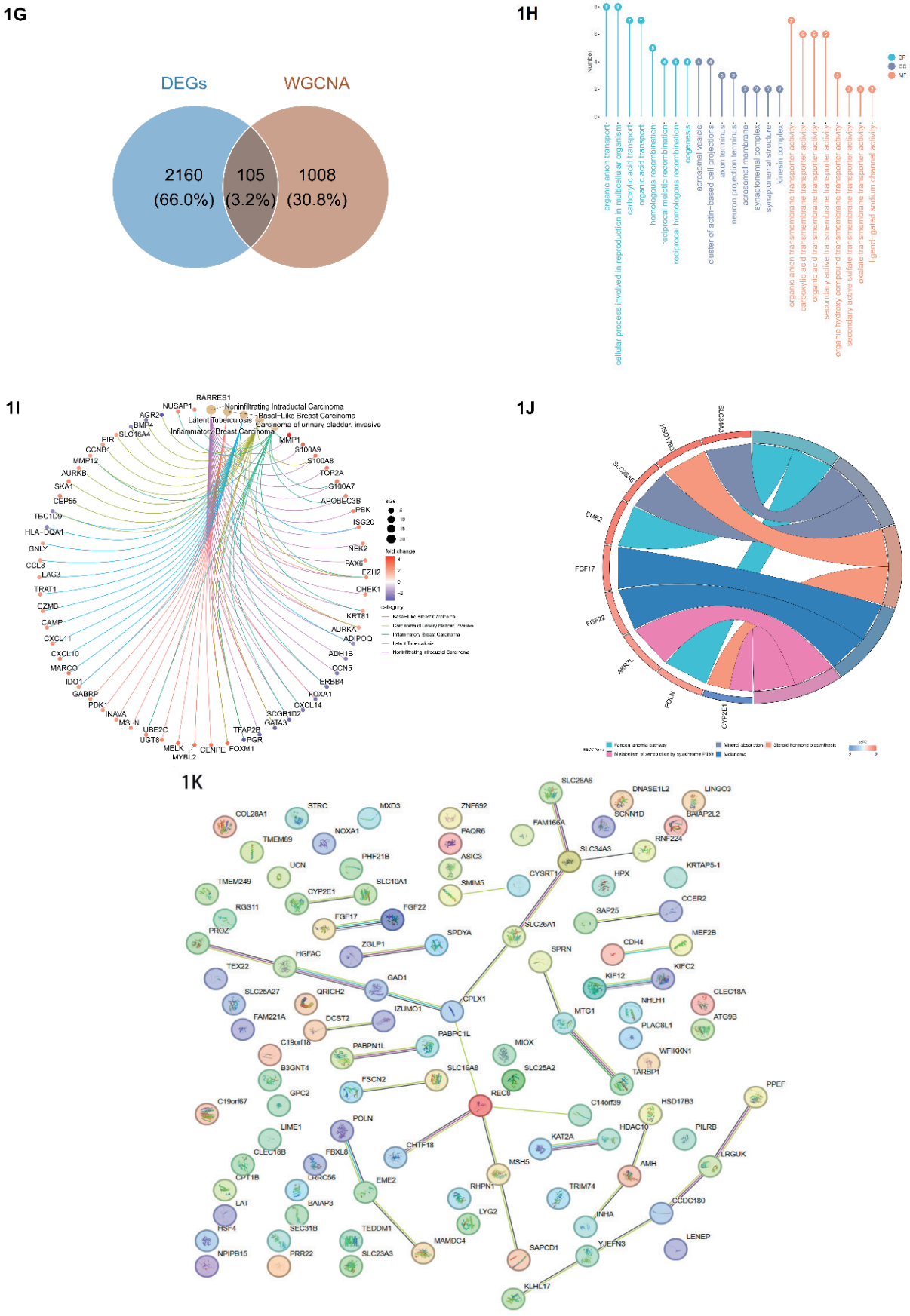


Figure 1. A total of 105 DE-NMRGs were identified in Prostate cancer patients, A. Differential gene map, B. Differential gene expression calorimetric map, C. Differences in NMRGs scores between Prostate cancer and Control groups, D. Left: scale-free fit index; right: average of all gene contiguity functions, E. Construct a co-expression network, F. Key module screening, G. Venn diagram of DE-PMRGs, H. Functional enrichment analysis of DE-NMRGs, I-J. Functional enrichment analysis of DE-NMRGs, K. DE-NMRGs PPI Network

https://doi.org/10.56294/saludcyt20251644

Figure 1A: Differential gene map, the orange equilateral triangle signifies up-regulated differentially expressed genes, the green inverted triangle represents down-regulated differentially expressed genes, and the gray X-shaped symbols denote genes that lack statistical significance.

Figure 1B: Differential gene expression calorimetric map, the NMRG score was significantly higher in the Prostate cancer samples compared to the healthy controls (P < 0,001).

Figure 1C: Differences in NMRGs scores between Prostate cancer and Control groups, significant variations were observed in NMRGs scores between these groups, indicating a distinct role for NMRGs in Prostate cancer.

Figure 1D: In the left figure, the vertical axis shows the scale-free fit index (signed R2), where a higher value indicates a stronger adherence to a scale-free network distribution. In the right figure, the vertical axis represents the average of all gene contiguity functions in the corresponding gene module.

Figure 1E: Construct a co-expression network. Tens of thousands of genes in the expression matrix were clustered into modules, forming a systematic hierarchical tree. Each gene module contained at least 200 genes, resulting in a total of 11 modules.

Figure 1F: Key module screening: Heat map presenting the correlation between modules and scores. Among the 11 modules obtained from WGCNA analysis, MEpink (528 genes) and MEblack (585 genes), the two modules with the highest absolute NMRGs score, were selected as key modules, with a total of 1113 genes.

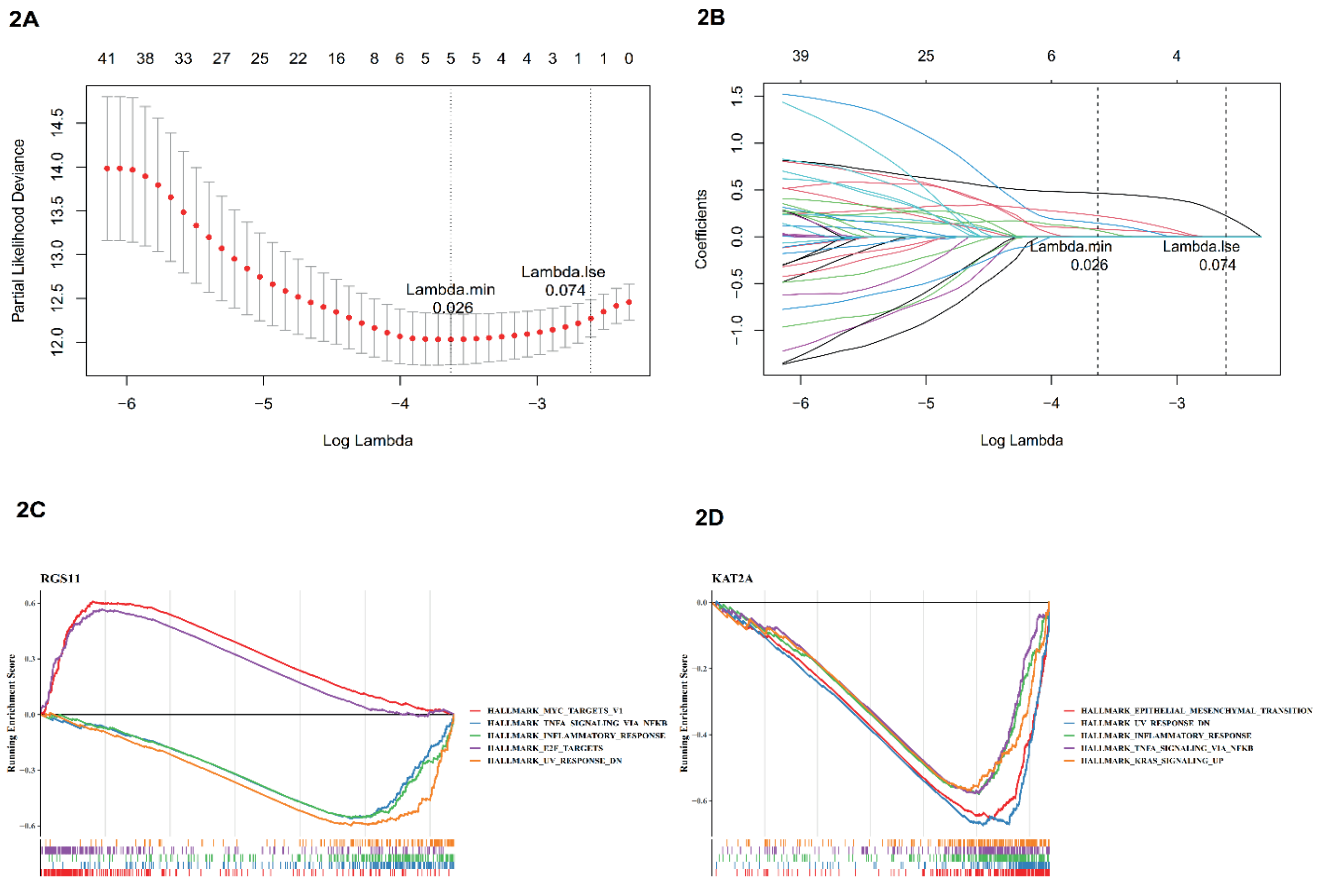
Figure 1G: Venn diagram of DE-PMRGs: This process yielded a total of 105 intersection genes, which were identified as DE-NMRGs.

Figure 1H: Functional enrichment analysis of DE-NMRGs. The top 8 entries from each of the three sections of GO enrichment are visualized in ascending order based on their P-values

Figure 1I-1J: Functional enrichment analysis of DE-NMRGs. Figure 1I: The top 8 entries from each of the three sections of GO enrichment are visualized in ascending order based on their P-values. Figure 1J: The left half of the circle depicts genes, with colors indicating the logFC (logarithm of fold change) value: red denotes up-regulated expression, while blue signifies down-regulated expression. The intensity of the color reflects the magnitude of the fold change, with darker shades indicating larger differences.

Figure 1K: DE-NMRGs PPI Network. An interaction network of 101 genes was obtained, containing 101 nodes and 33 interacting pairs.

Five prognostic genes were selected by the LASSO-Cox regression analyses



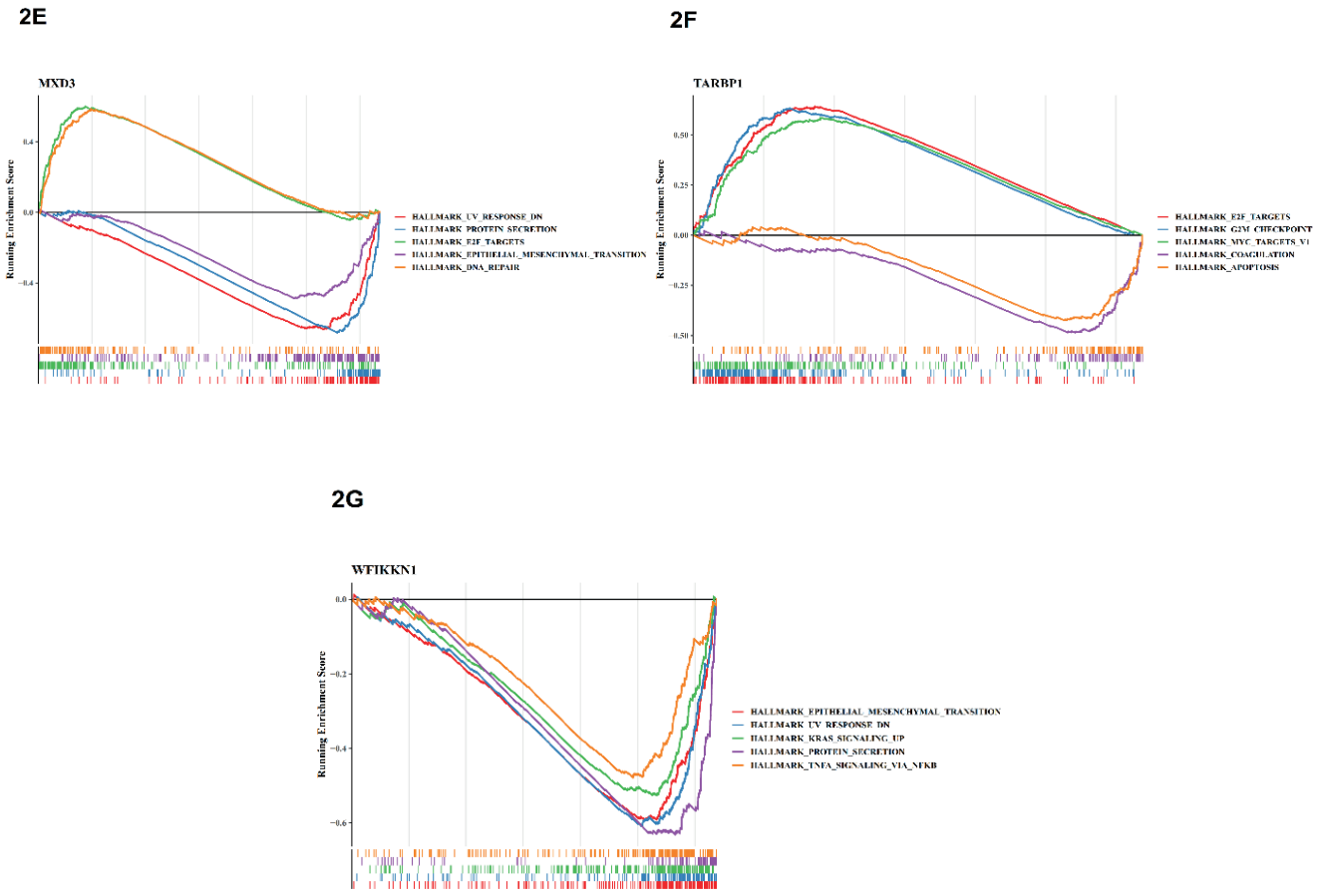


Figure 2. A-B. Prognostic gene screening, C-G. GSEA Enrichment Analysis

Based on the 391 Prostate cancer samples with recurrence information and time, 60 feature genes were screened by univariate Cox analysis. After the PH test, 55 feature genes were retained (P < 0,05). Subsequently, feature genes were narrowed down to 5 by the LASSO analysis (figure 2A, B). The 5 remaining feature genes were determined again as prognostic genes via multivariate Cox analysis, including RGS11, KAT2A, MXD3, TARBP1, and WFIKKN1. Therefore, the 5 prognostic genes of the risk model could be independent prognostic predictors. The GSEA results indicated that all of the prognostic genes except TARBP1 were enriched 'UV response DN', etc. (figure 2C-G). RGS11 and TARBP1 were linked with 'MYC target V1' and 'E2F targets' (figure 2C, F). RGS11, KAT2A, and WFIKKN1 were related to 'TNFA signaling via NFKB' (figure 2C, D, G).

Figure 2A-2B: Prognostic gene screening, figure 2A Lasso Regression Analysis: AXIS X offers a magazine (Lambda), but the axis shows a cross -check error there.

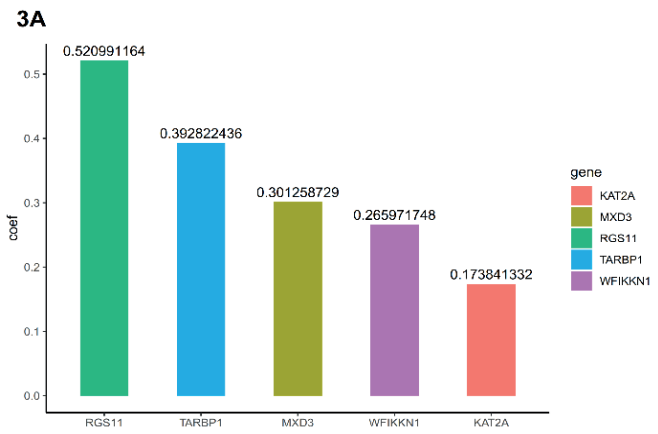
Figure 2B: The horizontal axis displays deviance, indicating the proportion of the residual explained by the model, with the number of genes plotted against the proportion of the residual explained. The y-axis represents the coefficient of each gene.

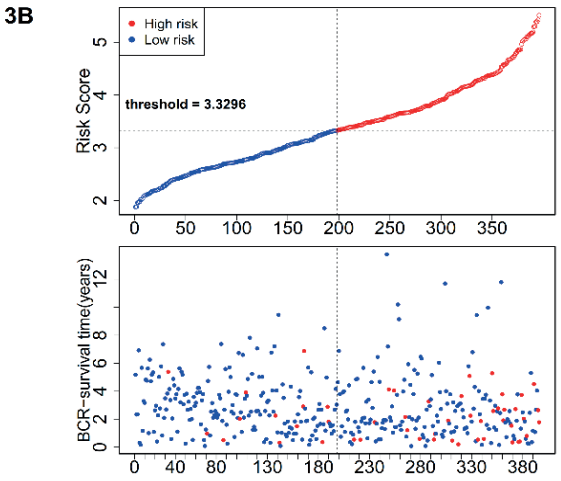
Figure 2C-G: GSEA Enrichment Analysis: The top 5 Hallmarks pathways are displayed based on their P. adjust values, arranged from small to large.

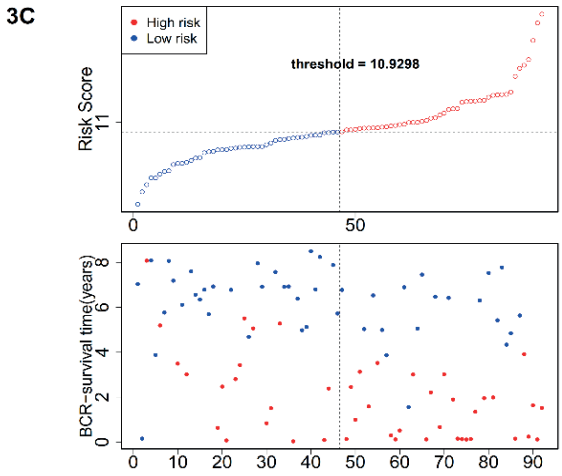
The BCR-free rate in the high-risk group was lower than that in the low-risk group

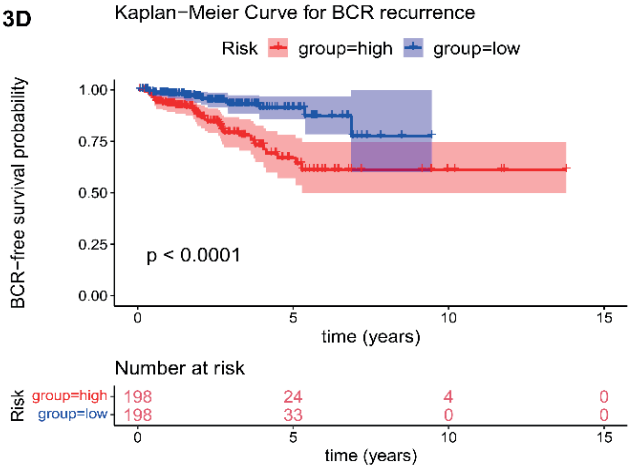
The risk model was established based on the five prognostic genes, with the risk coefficients shown in figure 3A. The Prostate cancer patients in TCGA-PRAD and GSE70769 were categorized as high- and low-risk, respectively. In TCGA-PRAD and GSE70769, the BCR-free rate was significantly lower in the high-risk group than in the low-risk group (figure 3B, C). Moreover, the BCR-free survival probability of Prostate cancer patients in the high-risk group was lower (figure 3D, E). The area under the curve (AUC) for 1-, 3-, and 5-year periods was above 0,6 in both TCGA-PRAD and GSE70769 datasets, indicating that the prognostic risk model can effectively predict the BCR-free probability of Prostate cancer patients (figure 3F, figure 3G). The expression levels of 5 prognostic genes in low-risk groups were lower than those of high-risk groups (figure 3H, I).

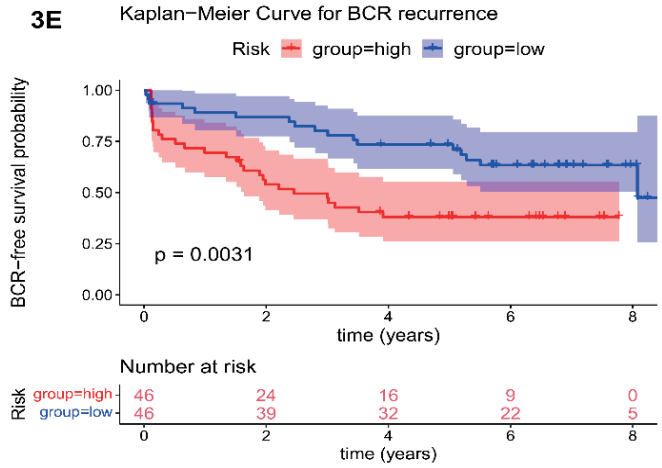
https://doi.org/10.56294/saludcyt20251644

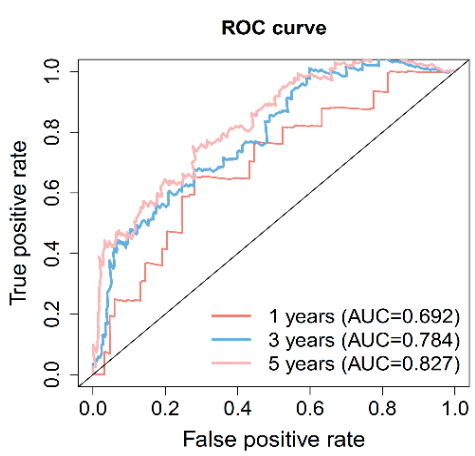


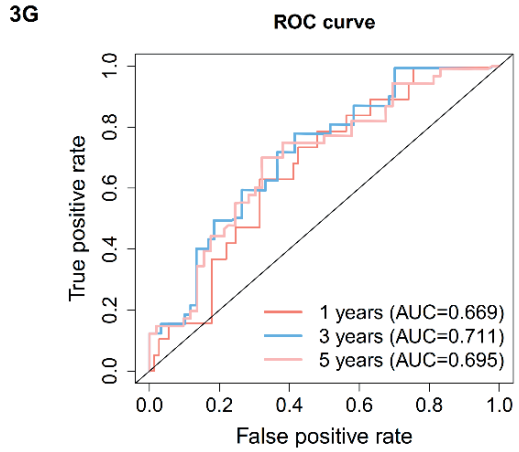


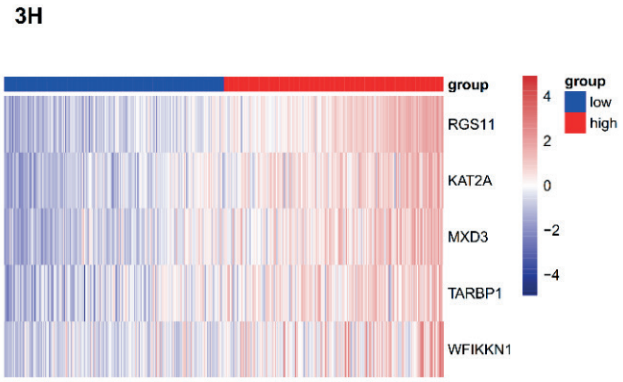












https://doi.org/10.56294/saludcyt20251644

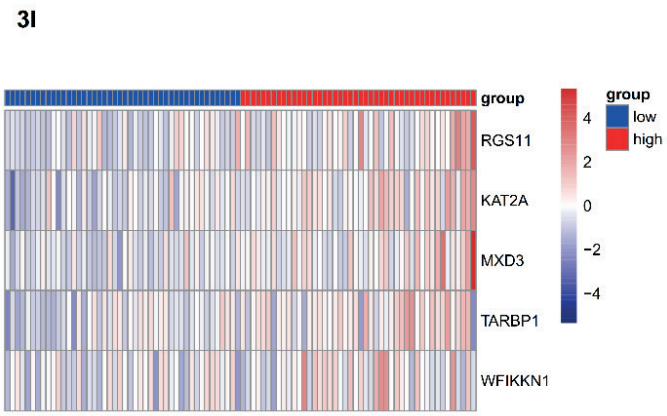


Figure 3. A. Risk Model Coefficients, B. Risk Curve (Training Set), C. The high-risk group had a higher number of recurrences than the low-risk group, D-E. Recurrence rates of Prostate cancer patients, F. ROC Curve (Training Set), G. ROC Curve (GSE70769 Validation Set), H. Gene expression heat maps for high-risk and low-risk group models (training set), I. Risk model

Figure 3A: Risk Model Coefficients: This model demonstrates effective prediction of recurrence status in Prostate cancer patients, thereby demonstrating improved prognostic performance.

Figure 3B Risk Curve (Training Set). There were more deceased cases observed in the high-risk group compared to the low-risk group. The BCR recurrence rate of Prostate cancer patients was lower in the low-risk group than in the high-risk group.

Figure 3C: The high-risk group had a higher number of recurrences than the low-risk group. The recurrence rate of BCR samples in the high-risk group surpassed that in the low-risk group, aligning with the findings from the training set. This consistency underscores the stability and applicability of prognostic risk models.

Figure 3D-H: Establishment and Verification of the Risk Model:

Figure 3D: It demonstrated a significant difference ($p < 0,05$) in the recurrence rates of Prostate cancer patients between the two groups. Specifically, patients in the high-risk group exhibited a significantly higher recurrence rate than those in the low-risk group.

Figure 3E: The recurrence rate of the high-risk group was higher than that of the low-risk group ($p < 0,05$), consistent with the findings from the training set.

Figure 3F: ROC Curve (Training Set): It illustrated a significant difference ($p < 0,05$) in the recurrence rates of Prostate cancer patients between the two groups. Specifically, the high-risk group had a significantly higher recurrence rate than the low-risk group.

Figure 3G: ROC Curve (GSE70769 Validation Set): The recurrence rate in the high-risk group was significantly higher than that in the low-risk group ($p < 0,05$), consistent with the results observed in the training set.

Figure 3H: Gene expression heat maps for high-risk and low-risk group models (training set). Expression heat maps showing the expression patterns of prognostic genes for high-risk and low-risk groups in the training set were also provided.

Fig 3I: The five target genes was tested in different risk groups to validate the risk model.

Nomogram model of Prostate cancer patients was constructed

Here, the risk scores varied significantly across different subgroups of ages, Gleason scores, T categories, and N categories. Additionally, there were significant discrepancies in Gleason scores, T categories, and N categories between the two risk groups (figure 4A). In addition, clinical indicators and risk scores were analyzed to further develop the risk prediction model. Subsequently, the risk score, Gleason, T categories, and PSA levels were screened as independent prognostic predictors by univariate and multivariate Cox, PH test (figure 4B, figure 4C). Later, those independent prognostic predictors were used to develop a nomogram model aimed at better predicting the BCR rate of Prostate cancer patients (figure 4D). Furthermore, the AUCs for all the three time periods (1, 3, and 5 years) were above 0,7, indicating that the nomogram exhibited excellent predictive ability (figure 4E). The calibration curves of the three periods (1, 3, and 5 years) suggested that the nomogram was in good agreement with the actual BCR of Prostate cancer patients (figure 4F).

Figure 4A: Among the five target genes, significant differences were observed across different subtypes of Gleason, N, and T among the two groups.

Figure 4B-4C: Multivariate Cox regression analysis:

Figure 4B Univariate Cox analysis was performed on Prostate cancer samples in the TCGA-Prostate cancer dataset, as well as HR≠1 criteria and P values. The risk score, Gleason score, T, N, and PSA were obtained.

https://doi.org/10.56294/saludcyt20251644

Figure 4C Risk score, Gleason, T, N, and PSA were included for multivariate Cox prognostic analysis. HR≠1 and P value<0,2 were used as criteria for screening. Risk score, Gleason, T, and PSA were identified as independent prognostic factors.

Figure 4D: Nomogram of independent prognostic factors: The R language rms package is used to score based on riskScore, Gleason, T, and PSA. Each factor corresponds to a score, and the sum of the scores of each factor corresponds to the TotalPoint. The recurrence rate at 1, 3, and 5 years is predicted according to the total score.

Figure 4E: ROC curve of Nomogram model: The AUC values of the Nomogram model for the first, third, and fifth years are all greater than 0,6, demonstrating that the Nomogram model has good predictive ability.

Figure 4F: Nomogram Calibration Curve: The horizontal coordinate represents the predicted event incidence, while the vertical coordinate denotes the actual event incidence. Both coordinates range from 0 to 1, indicating the percentage of event occurrence.

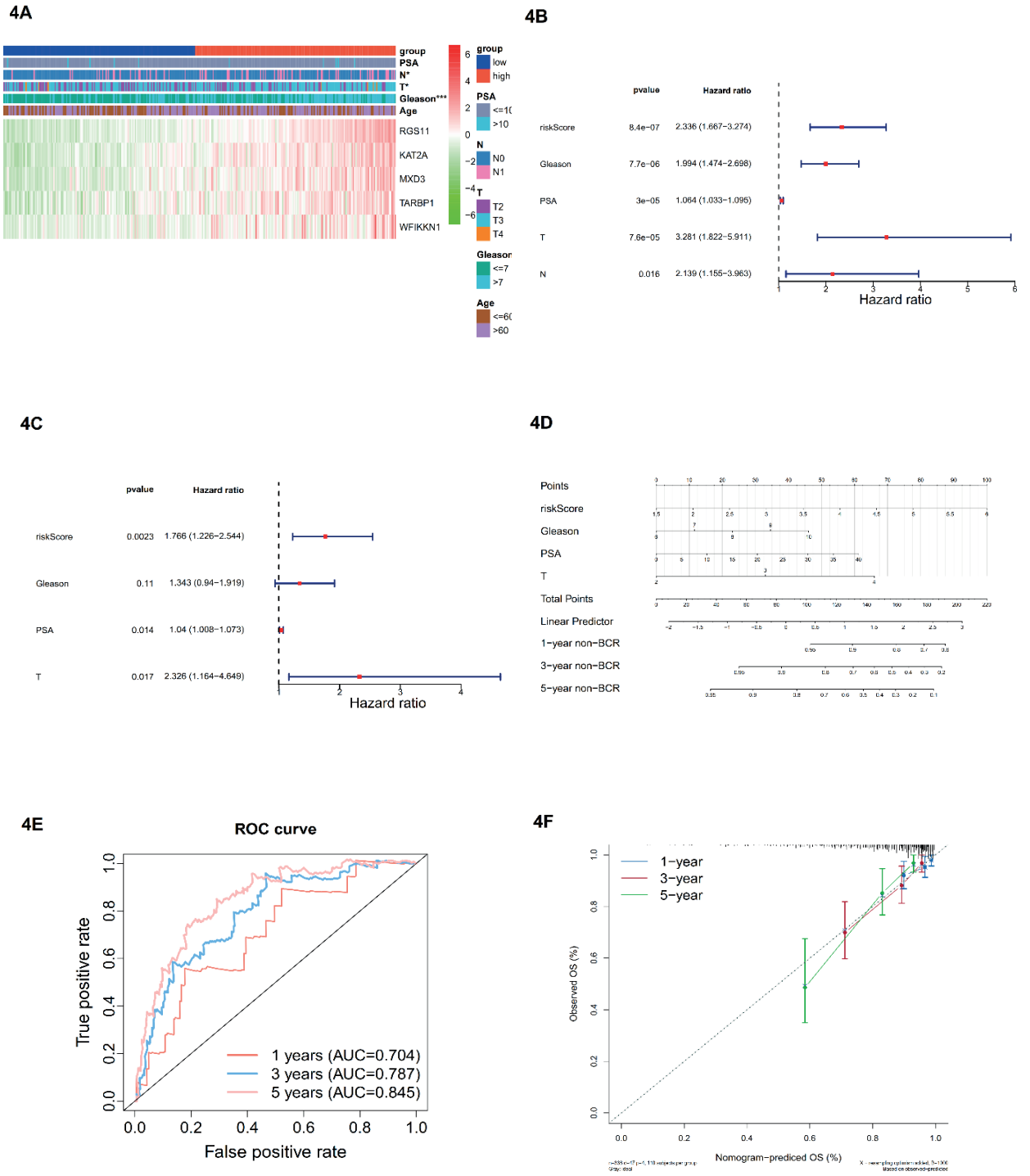
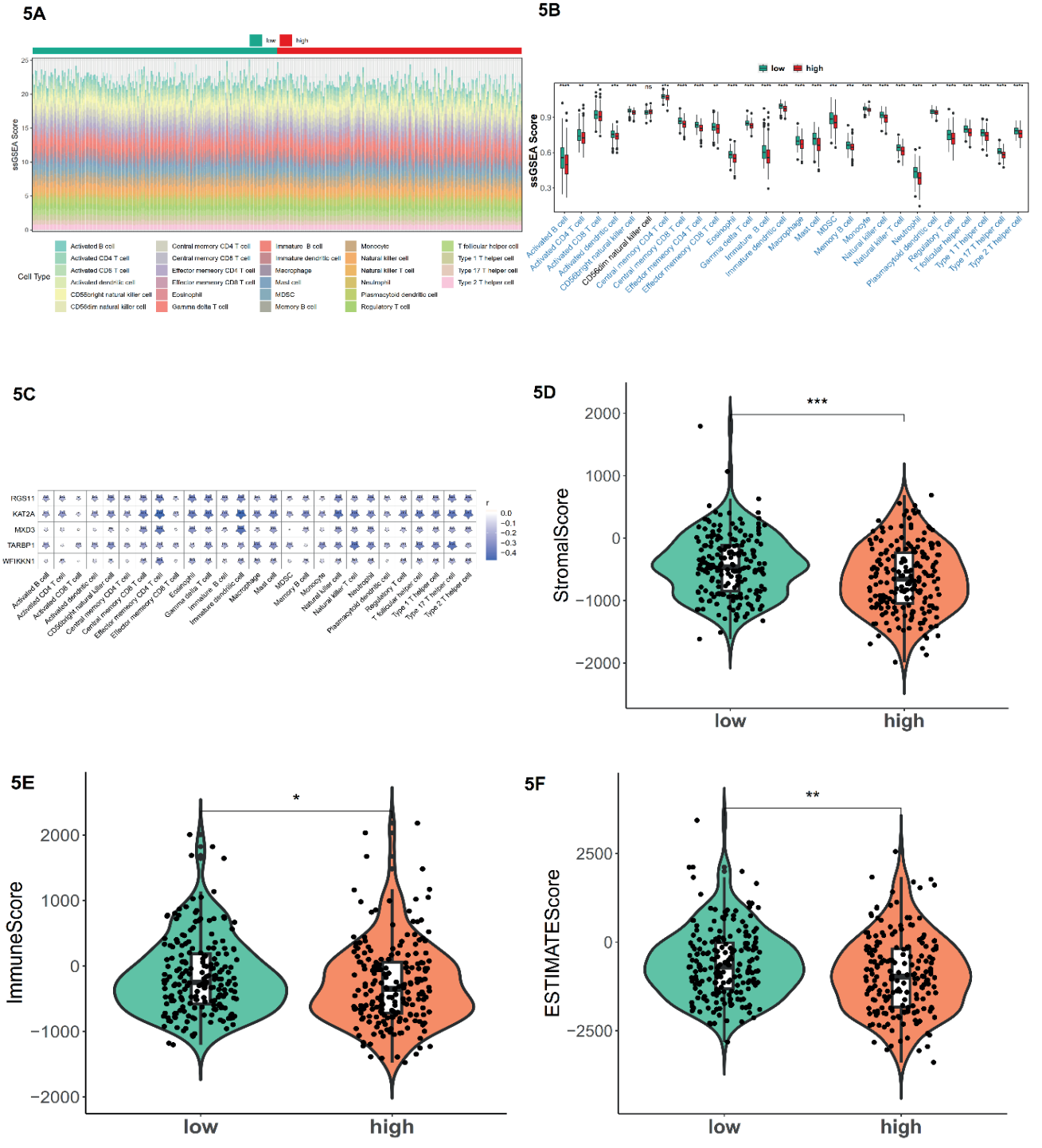


Figure 4. A. Subtypes of Gleason, N, and T, B-C. Multivariate Cox regression analysis, D. Nomogram of independent prognostic factors, E. ROC curve of Nomogram model, F. Nomogram Calibration Curve

https://doi.org/10.56294/saludcyt20251644

Immune mechanism was explored in Prostate cancer patients

In TCGA-PRAD, two groups of immune cell infiltration are shown in figure 5A. All immune cells, except CD56dim natural killer cells, exhibited significant differences between the two groups (P < 0,05) (figure 5B). Five prognostic genes showed significant correlations with most immune cells (figure 5C). Among them, the absolute value of the correlation coefficient between the effector memory CD4 T cell and KAT2A was the highest , which was | -0,476 | (P < 0,001). Furthermore, it turned out that Stoma's score, evaluation score, and high-risk group immune scores were significantly lower than low-risk groups (figure 5D-F). The Immuno-Phenoscore (IPS) score was used to predict patient response to immune checkpoint inhibitor (ICI) treatment. As shown in figure 5G-J, The IPS scores of CTLA4-/PD-1-, CTLA4-/PD-1+, and CTLA4+/PD-1- in the low-risk group were significantly higher than those in the high-risk group. This suggests that immune checkpoint inhibitors (ICIs) may be effective in treating low-risk Prostate cancer patients.



https://doi.org/10.56294/saludcyt20251644

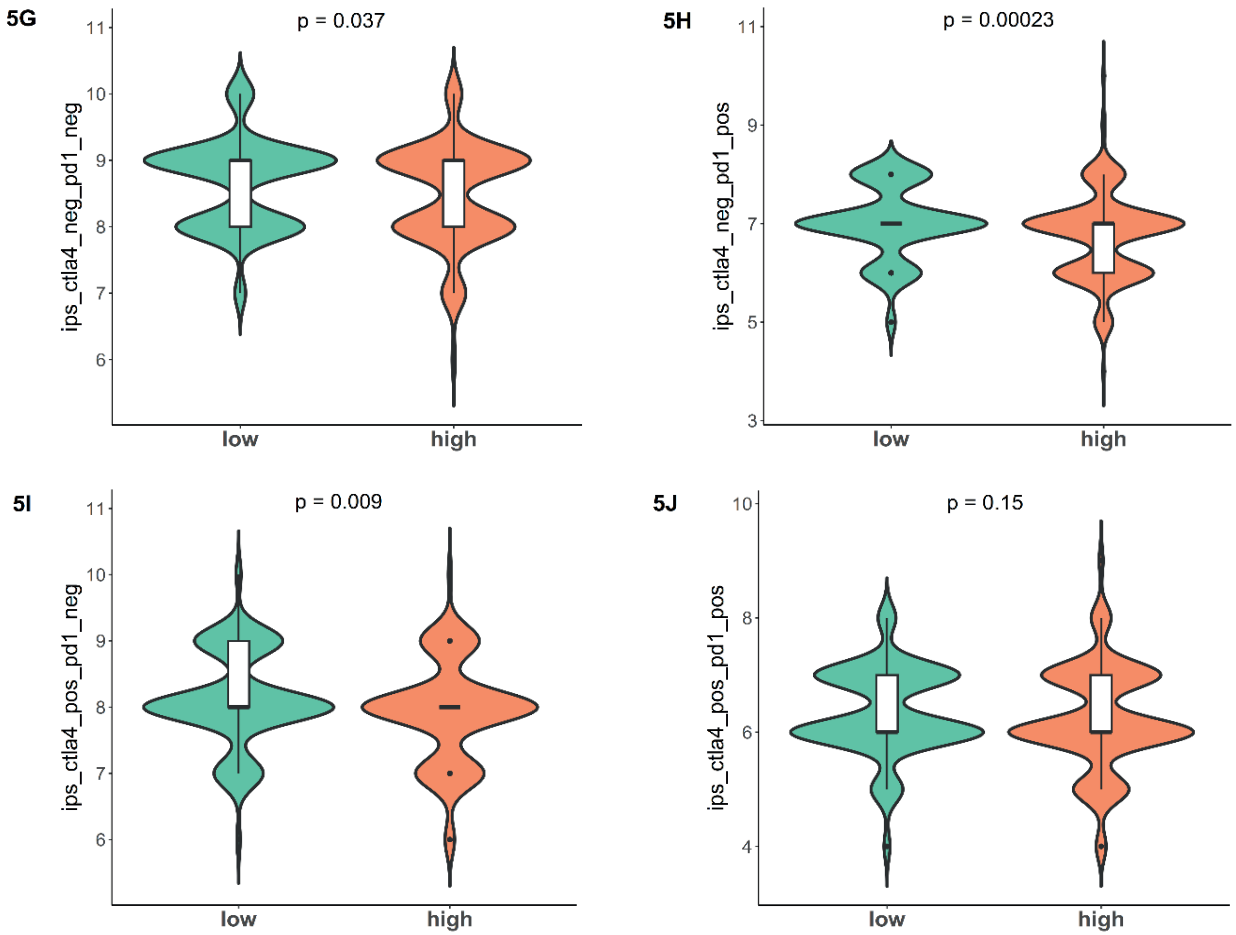
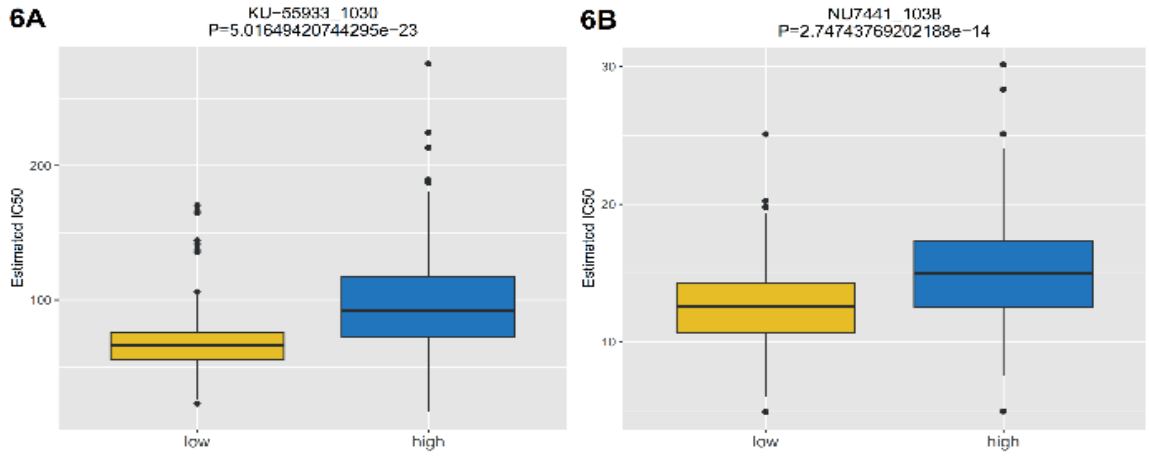


Figure 5. A. Immune cell score heat map, B. Box Plots of Immune Cell Scores in High and Low-Risk Groups, C. Correlation between prognostic genes and immune cells, D. Comparison between high-risk and low-risk groups based on Immune Score, E. Comparison between high-risk and low-risk groups based on Comprehensive Score, F. Stromal Score, G-J. Differences in IPS Between High and Low-Risk Groups

Figure 5A: Immune cell score heat map: The horizontal axis represents 28 types of immune cells (red indicates significantly up-regulated immune cells, while blue denotes significantly down-regulated immune cells), while the vertical axis represents the proportion of immune cells within the sample.

Figure 5C: Correlation between prognostic genes and immune cells: The horizontal coordinate shows immune cells, the vertical coordinate shows prognostic genes, and the five-pointed stars show positive correlation up and negative correlation down.

A total of 37 drugs were predicted related to risk score



https://doi.org/10.56294/saludcyt20251644

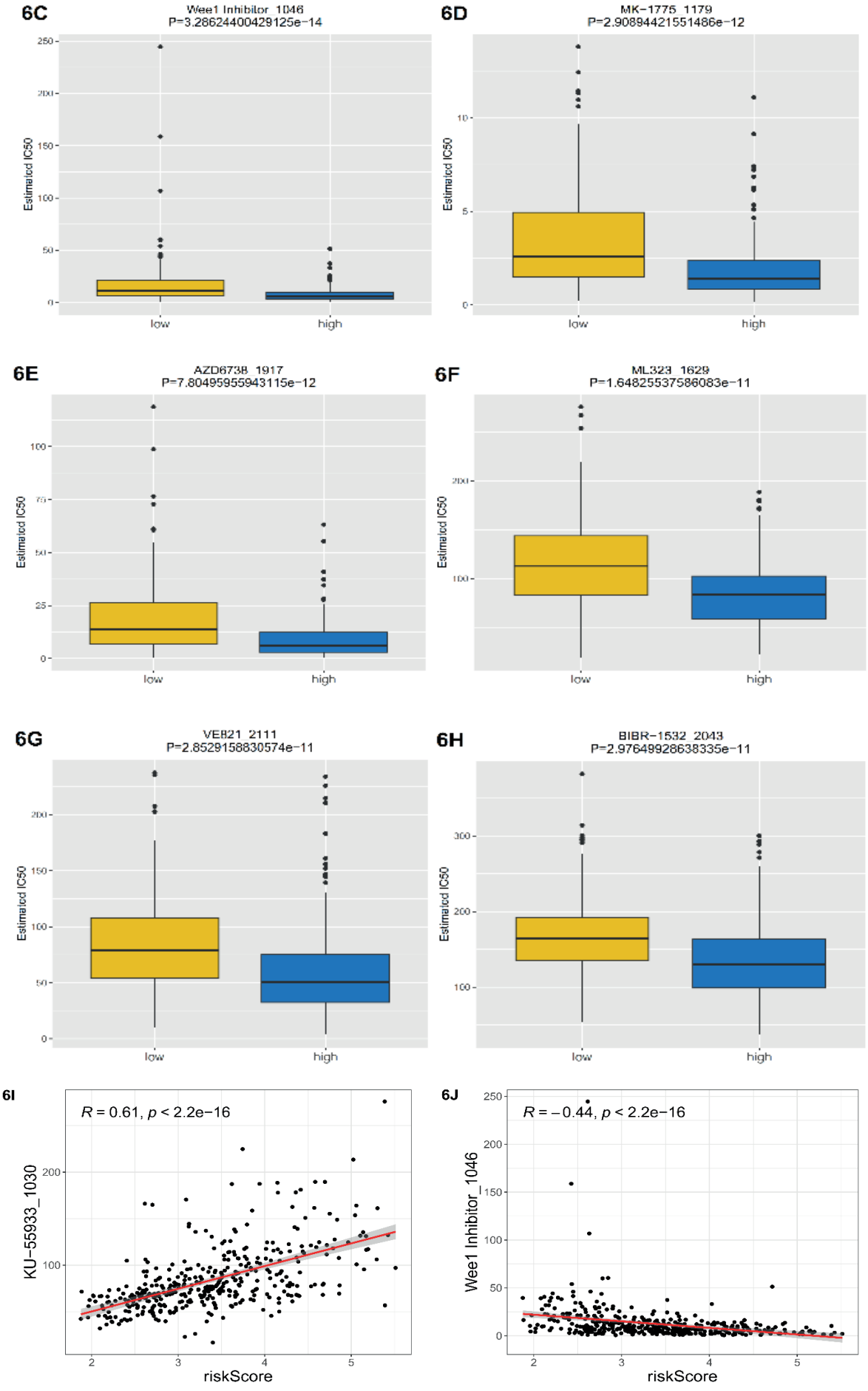


Figure 6. A-H. Top 8 drugs were displayed, I-J. Correlation between Risk-Score and IC50 values

https://doi.org/10.56294/saludcyt20251644

There were 126 drugs of significantly different IC50 levels between the high- and low-risk groups in TCGA-PARD. The top 8 drugs were displayed in figure 6A-H, including KU-55933_1030, NU7441_1038, Wee1 Inhibitor_1046, MK-1775_1179, AZD6738_1917, ML323_1629, VE821_2111, and BIBR-1532_2043. To explore the correlation between prognostic genes and drug sensibility, we analyzed the correlation between risk score and drug IC50. There were 37 drugs significantly associated with risk score (|cor| > 0,3, P < 0,05). Among them, KU-55933_1030 had a significantly positive correlation with risk score (cor = 0,608, P < 0,001), and Wee1 Inhibitor_1046 had a significantly negative correlation with risk score (cor = -0,441, P < 0,001) (figure 6I, J).

6A-6H: Top 8 drugs were displayed: The differences in the top 8 drugs between the high and low-risk groups, ordered by P-values from smallest to largest, were visualized using box plots.

6I-6J: The correlation between Risk-Score and IC50 values: Spearman correlation analysis was employed to examine the correlation between the expression of prognostic genes and susceptibility to common drugs.

A total of 90 miRNA-mRNA interaction pairs were predicted related to prognostic genes

We submitted the 5 prognostic genes on the online website, and then 264 miRNAs related to the 5 prognostic genes from the DIANA microT-CDS database and 208 miRNAs related to the 4 prognostic genes (except WFIKKN1) from the miRDB database were predicted. Next, the 89 miRNAs from the two databases were overlapped, which contained 90 miRNA-mRNA pairs. Then, the 23 miRNAs were obtained on the miRDB database at clipExpNum>70. Based on these 23 miRNAs, 12 LncRNAs were predicted on the miRDB database at clipExpNum>70, which generated a ceRNA network (figure 7). Thereinto, the interaction of MXAD was the most complex, including MIR17HG-hsa-miR-302b-3p-MXAD, NEAT1-hsa-miR-195-5p-MXAD, etc.

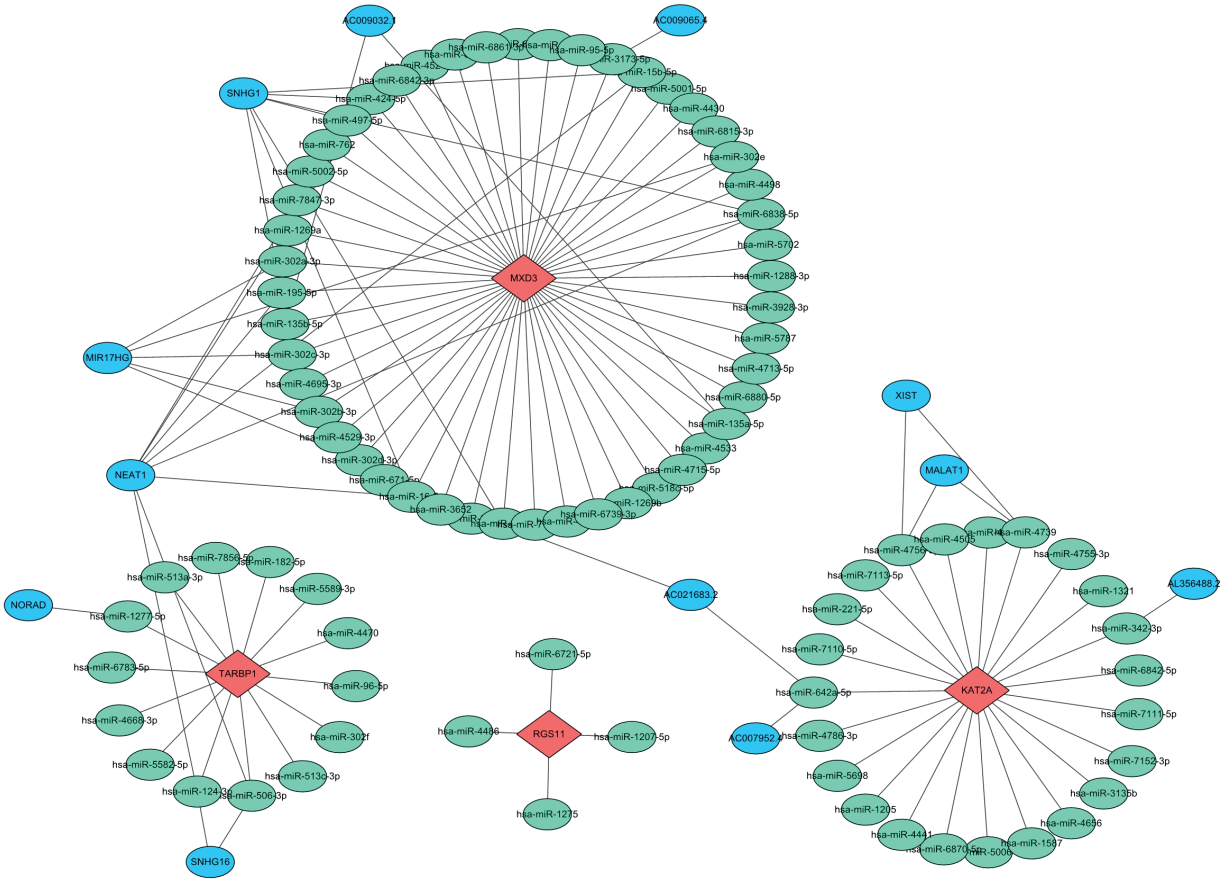


Figure 7. ceRNA regulatory network

In the network map, mRNA nodes are depicted in red, miRNA nodes in green, and lncRNA nodes in blue. The lines connecting them represent their interactions. The network comprises 4 mRNAs, 89 miRNAs, and 12 lncRNAs, resulting in a total of 90 interaction pairs.

qRT-PCR

The expression levels of 5 prognostic genes in Prostate cancer patients exhibited significantly higher values compared to those observed in the healthy controls within the TCGA-PARD dataset (figure 8).

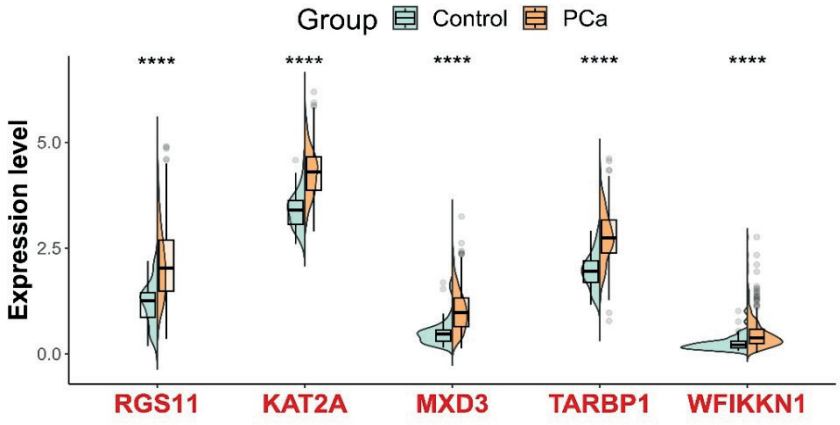


Figure 8. Expression of Prognostic Genes (Training Set)

DISCUSSION

Prostate cancer originates from the prostate gland, specifically in the gland cells, making it an adenocarcinoma. Diagnosis often involves a PSA test, digital rectal examination (DRE), and needle biopsy, which is the most commonly used test. Early detection and management are crucial for effective treatment and improving outcomes.

The relevant data information was downloaded from the GEO and TCGA databases, and bioinformatics techniques were used to screen for five highly expressed genes (RGS11, KAT2A, MXD3, TARBP1, and WFIKKN) associated with Prostate cancer. These five highly expressed genes have never been used in previous Prostate cancer studies, making this study innovative.

The application of gene expression analysis and tumor research is an important bioinformatics method, which can screen out genes with potential clinical significance in a large number of gene expression data. For example, the RGS11 gene has been found to be associated with tumor aggressiveness and metastasis in some studies and has been used as a novel tumor marker for lung cancer. The MXD3, TARBP1, and WFIKKN genes have also been mentioned in other cancer-related studies and have been shown to play a role in tumor cell growth, apoptosis, and metastasis. Using computer models to construct predictive models incorporating these genes to predict disease progression and treatment in Prostate cancer patients, thus providing a scientific basis for clinical treatment decision.

1.RGS11: Involved in signaling pathways by regulating G-protein activity. Potentially influences Prostate cancer cell behavior through effects on cell signaling, though more research is needed to establish a clear link. 2.KAT2A (GCN5 or PROSTATE CANCERF) A histone acetyltransferase that modifies chromatin structure, affecting gene expression. Overexpression or dysregulation of KAT2A may contribute to uncontrolled cell growth and tumorigenesis in Prostate cancer. 3.MXD3: A transcription factor that forms heterodimers with MAX, regulating genes involved in cell cycle and apoptosis. It is role in Prostate cancer is not well-established, but it may influence the expression of genes critical to cancer development. 4. TARBP1: Encodes a protein involved in RNA processing and stability. While primarily associated with neurodegenerative diseases, TARBP1 could theoretically impact Prostate cancer through effects on gene expression or RNA metabolism. 5.WFIKKN: Encodes a protein with anti-inflammatory properties that may modulate immune response and tumor microenvironment. Its specific role in Prostate cancer is not well-defined, but it could influence cancer progression through effects on inflammation and immunity.

Given the current state of research, it's crucial to approach the study of these genes in Prostate cancer with the understanding that the full extent of their involvement is not yet completely understood, and further investigation is necessary.

At present, there are no studies on RGS11, KAT2A, MXD3, TARBP1, and WFIKKN with Prostate cancer have been reported. For the first time, we have analyzed the association between Prostate cancer and RGS11, KAT2A, MXD3, TARBP1, and WFIKKN through biogenic analysis, but the specific functional mechanism needs to be further verified through follow-up experiment.

Therefore, in this study, five highly expressed genes were selected using bioinformatics methods, based on which a Prostate cancer risk prediction model was established, providing a new perspective and tool for personalized treatment and evaluation of Prostate cancer prognosis. Future studies are needed to further confirm the clinical utility of these genes and explore their specific roles in the development of Prostate cancer.

CONCLUSIONS

Overall, RGS11, KAT2A, MXD3, TARBP1, and WFIKKN have emerged as potential therapeutic targets for diagnosing and treating Prostate cancer. Their identification through the BCR-free risk model underscores their

significance in predicting biochemical recurrence and guiding personalized treatment strategies for Prostate cancer patients.

BIBLIOGRAPHIC REFERENCES

1. Schafer EJ, Jemal A, Wiese D, Sung H, Kratzer TB, Islami F, Dahut WL, Knudsen KE. Disparities and Trends in Genitourinary Cancer Incidence and Mortality in the USA. Eur Urol. 2023 Jul;84(1):117-126. doi: 10.1016/j.eururo.2022.11.023.

2. Bass EJ, Pantovic A, Connor MJ, Loeb S, Rastinehad AR, Winkler M, Gabe R, Ahmed HU. Diagnostic accuracy of magnetic resonance imaging targeted biopsy techniques compared to transrectal ultrasound guided biopsy of the prostate: a systematic review and meta-analysis. Prostate Cancer Prostatic Dis. 2022 Feb;25(2):174-179. doi: 10.1038/s41391-021-00449-7

3. Rauluseviciute I, Drabløs F, Rye MB. DNA hypermethylation associated with upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation. BMC Med Genomics. 2020 Jan 8;13(1):6. doi: 10.1186/s12920-020-0657-6.

4. Kanzi AM, San JE, Chimukangara B, Wilkinson E, Fish M, Ramsuran V, de Oliveira T. Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance. Front Genet. 2020 Oct 23;11:544162. doi: 10.3389/fgene.2020.544162.

5. Guo Y, Ziesch A, Hocke S, Kampmann E, Ochs S, De Toni EN, Göke B, Gallmeier E. Overexpression of heat shock protein 27 (HSP27) increases gemcitabine sensitivity in pancreatic cancer cells through S-phase arrest and apoptosis. J Cell Mol Med. 2015 Feb;19(2):340-50. doi: 10.1111/jcmm.12444.

6. Hamzeh O, Alkhateeb A, Zheng J, Kandalam S, Rueda L. Prediction of tumor location in prostate cancer tissue using a machine learning system on gene expression data. BMC Bioinformatics. 2020 Mar 11;21(Suppl 2):78. doi: 10.1186/s12859-020-3345-9.

7. Li J, Eu JQ, Kong LR, Wang L, Lim YC, Goh BC, Wong ALA. Targeting Metabolism in Cancer Cells and the Tumour Microenvironment for Cancer Therapy. Molecules. 2020 Oct 20;25(20):4831. doi: 10.3390/molecules25204831.

8. Peng X, Chen Z, Farshidfar F, Xu X, Lorenzi PL, Wang Y, Cheng F, Tan L, Mojumdar K, Du D, Ge Z, Li J, Thomas GV, Birsoy K, Liu L, Zhang H, Zhao Z, Marchand C, Weinstein JN; Cancer Genome Atlas Research Network; Bathe OF, Liang H. Molecular Characterization and Clinical Relevance of Metabolic Expression Subtypes in Human Cancers. Cell Rep. 2018 Apr 3;23(1):255-269.e4. doi: 10.1016/j.celrep.2018.03.077.

9. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8.

10. Maag JLV. gganatogram: An R package for modular visualisation of anatograms and tissues based on ggplot2. F1000Res. 2018 Sep 28;7:1576. doi: 10.12688/f1000research.16409.2.

11. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec 29;9:559. doi: 10.1186/1471-2105-9-559.

12. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011 Jan 26;12:35. doi: 10.1186/1471-2105-12-35.

13. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012 May;16(5):284-7. doi: 10.1089/omi.2011.0118

14. Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics. 2015 Feb 15;31(4):608-9. doi: 10.1093/bioinformatics/btu684

15. Houyun Huang. ggcor: Extended tools for correlation analysis and visualization. R package version 0.9 7. 2020. https://gitee.com/luerner/ggcor

16. Scire J, Huisman JS, Grosu A, Angst DC, Lison A, Li J, Maathuis MH, Bonhoeffer S, Stadler T. estimateR: an R package to estimate and monitor the effective reproductive number. BMC Bioinformatics. 2023 Aug 11;24(1):310. doi: 10.1186/s12859-023-05428-4.

17. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. Brief Bioinform. 2021 Nov 5;22(6):bbab260. doi: 10.1093/bib/bbab260.

18. Bader DA, McGuire SE. Tumour metabolism and its unique properties in prostate adenocarcinoma. Nat Rev Urol. 2020 Apr;17(4):214-231. doi: 10.1038/s41585-020-0288-x.

19. Galosi AB, Palagonia E, Scarcella S, Cimadamore A, Lacetera V, Delle Fave RF, Antezza A, Dell'Atti L. Detection limits of significant prostate cancer using multiparametric MR and digital rectal examination in men with low serum PSA: Up-date of the Italian Society of Integrated Diagnostic in Urology. Arch Ital Urol Androl. 2021 Mar 22;93(1):92-100. doi: 10.4081/aiua.2021.1.92.

20. Wei JT, Barocas D, Carlsson S, Coakley F, Eggener S, Etzioni R, Fine SW, Han M, Kim SK, Kirkby E, Konety BR, Miner M, Moses K, Nissenberg MG, Pinto PA, Salami SS, Souter L, Thompson IM, Lin DW. Early Detection of Prostate Cancer: AUA/SUO Guideline Part I: Prostate Cancer Screening. J Urol. 2023 Jul;210(1):46-53. doi: 10.1097/JU.0000000000003491. Epub 2023 Apr 25.

21. Huang L, Xiong W, Cheng L, Li H. Bioinformatics-based analysis of programmed cell death pathway and key prognostic genes in gastric cancer: Implications for the development of therapeutics. J Gene Med. 2024 Jan;26(1):e3590. doi: 10.1002/jgm.3590.

22. Loe AKH, Zhu L, Kim TH. Chromatin and noncoding RNA-mediated mechanisms of gastric tumorigenesis. Exp Mol Med. 2023 Jan;55(1):22-31. doi: 10.1038/s12276-023-00926-0.

23. Kondás K, Szláma G, Trexler M, Patthy L. Both WFIKKN1 and WFIKKN2 have high affinity for growth and differentiation factors 8 and 11. J Biol Chem. 2008 Aug 29;283(35):23677-84. doi: 10.1074/jbc.M803025200.

24. Hu Y, Zheng M, Wang S, Gao L, Gou R, Liu O, Dong H, Li X, Lin B. Identification of a five-gene signature of the RGS gene family with prognostic value in ovarian cancer. Genomics. 2021 Jul;113(4):2134-2144. doi: 10.1016/j.ygeno.2021.04.012.

FINANCING

This work was financially supported by the Secretary of the Academic School, Graduate School, Mongolian National University of Medical Sciences, Ulaanbaatar city.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Wei Liu, Shiirevnyamba Avirmed, Bilegtsaikhan Tsolmon.
Data curation: Wei Liu, Shiirevnyamba Avirmed, Bilegtsaikhan Tsolmon.
Formal analysis: Wei Liu, Shiirevnyamba Avirme.
Research: Wei Liu, Shiirevnyamba Avirmed, Bilegtsaikhan Tsolmon.
Methodology: Wei Liu, Shiirevnyamba Avirmed, Hong Zhi.
Project management: Shiirevnyamba Avirmed, Bilegtsaikhan Tsolmon.
Resources: Wei Liu.
Supervision: Shiirevnyamba Avirmed, Bilegtsaikhan Tsolmon.
Validation: Shiirevnyamba Avirmed, Bilegtsaikhan Tsolmon.
Display: Shiirevnyamba Avirmed.
Drafting - original draft: Wei Liu.
Writing - proofreading and editing: Wei Liu, Shiirevnyamba Avirmed, Bilegtsaikhan Tsolmon.