



OPEN ACCESS

EDITED BY

Nemat Ali,
King Saud University, Saudi Arabia

REVIEWED BY

Pravin Hivare,
The University of Utah, United States
Elad Mazor,
Soroka Medical Center, Israel

*CORRESPONDENCE

Ninghan Feng

✉ n.feng@jiangnan.edu.cn

Yuwei Zhang

✉ zhangyw0123@njmu.edu.cn

Menglu Li

✉ lmlbw1232@163.com

†These authors have contributed equally to this work

RECEIVED 26 November 2024

ACCEPTED 21 March 2025

PUBLISHED 07 April 2025

CITATION

Lv J, Zhou Y, Jin S, Fu C, Shen Y, Liu B, Li M, Zhang Y and Feng N (2025) WGCNA-ML-MR integration: uncovering immune-related genes in prostate cancer. *Front. Oncol.* 15:1534612. doi: 10.3389/fonc.2025.1534612

COPYRIGHT

© 2025 Lv, Zhou, Jin, Fu, Shen, Liu, Li, Zhang and Feng. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

WGCNA-ML-MR integration: uncovering immune-related genes in prostate cancer

Jing Lv^{1,2†}, Yuhua Zhou^{1,2†}, Shengkai Jin^{1,2†}, Chaowei Fu^{1,2}, Yang Shen^{2,3}, Bo Liu^{2,4}, Menglu Li^{2*}, Yuwei Zhang^{2,4*} and Ninghan Feng ^{1,2,4*}

¹Wuxi School of Medicine, Jiangnan University, Wuxi, China, ²Department of Urology, Jiangnan University Medical Center, Wuxi, China, ³Department of Clinical Medicine, Nanjing Medical University, Nanjing, China, ⁴Department of Clinical Medicine, Nantong University Medical School, Nantong, China

Background: Prostate cancer is one of the most common tumors in men, with its incidence and mortality rates continuing to rise year by year. Prostate-specific antigen (PSA) is the most commonly used screening indicator, but its lack of specificity leads to overdiagnosis and overtreatment. Therefore, identifying new biomarkers related to prostate cancer is crucial for the early diagnosis and treatment of prostate cancer.

Methods: This study utilized datasets from the Gene Expression Omnibus (GEO) to screen for differentially expressed genes (DEGs) and employed Weighted Gene Co-expression Network Analysis (WGCNA) to identify driver genes highly associated with prostate cancer within the modules. The intersection of differentially expressed genes and driver genes was taken, and Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses were performed. Furthermore, a machine learning algorithm was used to screen for core genes and construct a diagnostic model, which was then validated in an external validation dataset. The correlation between core genes and immune cell infiltration was analyzed, and Mendelian randomization (MR) analysis was conducted to identify biomarkers closely related to prostate cancer.

Results: This study identified six core biomarkers: SLC14A1, ARHGEF38, NEFH, MSMB, KRT23, and KRT15. MR analysis demonstrated that MSMB may be an important protective factor for prostate cancer. In q-PCR experiments conducted on tumor tissues and adjacent non-cancerous tissues from prostate cancer patients, it was found that: compared to the adjacent non-cancerous tissues, the expression level of ARHGEF38 in prostate cancer tumor tissues significantly increased, while the expression levels of SLC14A1, NEFH, MSMB, KRT23, and KRT15 significantly decreased. To further validate these findings at the protein level, we conducted Western blot analysis, which corroborated the q-PCR results, demonstrating consistent expression patterns for all six biomarkers. IHC results confirmed that ARHGEF38 protein was highly expressed in tumor tissues, while MSMB expression was markedly reduced.

Conclusion: Our study reveals that SLC14A1, ARHGEF38, NEFH, MSMB, KRT23, and KRT15 are potential diagnostic biomarkers for prostate cancer, among which MSMB may play a protective role in prostate cancer.

KEYWORDS

prostate cancer, machine learning, weighted gene co-expression network analysis, Mendelian randomization, biomarkers

1 Introduction

Prostate cancer (PCa) is the second most common cancer globally and the fifth leading cause of cancer death among men (1). Statistics show that the incidence of prostate cancer is gradually increasing every year, especially in developed countries, partly due to an aging population and changes in lifestyle (2). Early-stage prostate cancer is usually asymptomatic, and when symptoms appear, the disease often has progressed to a point where it is incurable. Over the past decade, the incidence of PCa has risen significantly while mortality has declined, mainly due to the use of prostate-specific antigen (PSA) as a biomarker for PCa screening (3). Therefore, routine screening and monitoring based on reliable biomarkers are crucial for the early detection and assessment of tumor progression in prostate cancer, which can significantly improve patient survival rates and quality of life.

PCa is typically diagnosed through digital rectal examination, PSA screening, and biopsy. Early studies have shown that PSA is more sensitive than palpable tumor assessment by digital rectal examination in the detection of prostate cancer, with a higher detection rate, and plays an important role in monitoring the prognosis and recurrence of tumor patients (4, 5). Data from the European Randomized Study of Prostate Cancer Screening (ERSPC) indicates that after 16 years of follow-up, PSA screening significantly reduces the mortality rate of PCa, and the longer the follow-up, the greater the reduction in PCa mortality (6). Although PSA as a routine screening indicator has been widely accepted, its lack of specificity may lead to excessive examination and treatment, or overlook some potential malignant lesions (7, 8). Biopsy is the “gold standard” for the diagnosis of prostate cancer, but it may bring complications, including infection, hematospermia, hematuria, urinary retention, and erectile dysfunction (9). Therefore, there is an increasing clinical need to identify new biomarkers for prediction, diagnosis, and prognosis. These biomarkers could complement or reduce the reliance on invasive diagnostic procedures like biopsies.

In recent years, advancements in immunological research have revealed the critical role of immune cells in the occurrence and development of PCa. The infiltration of immune cells not only significantly affects the formation and maintenance of the tumor microenvironment but also plays a decisive role in the biological behaviors of tumor growth, invasion, and metastasis (10). For

instance, regulatory T cells (Treg cells) play an important role in regulating tumor progression by modulating immune suppression. Flammiger et al. detected Tregs in PCa specimens through FOXP3 immunohistochemistry and found that Tregs within tumor tissues were positively correlated with PCa staging and the Ki67 index (11). Zhao et al.’s research indicated that in the bone marrow microenvironment of patients with bone metastases, the number of Treg significantly increased. In particular, CD4+CD25^{high} Treg cells may promote the transfer of tumor cells to the skeleton by suppressing the immune microenvironment within the bone marrow (12). Furthermore, Yang et al.’s study demonstrated that in patients with high-risk PCa who underwent radical prostatectomy, the infiltration level of CD8+ T cells within the tumor was independently associated with a significant increase in patient survival rates (13). Therefore, immunotherapy is extremely important in the treatment of PCa. However, compared to other malignant tumors, the clinical efficacy of immunotherapy in PCa has certain limitations (14). In summary, although immunotherapy has shown potential in the treatment of PCa, it still faces many challenges. Future research needs to further explore new biomarkers to provide new immune treatment plans for PCa.

In this study, we obtained 5 publicly available PCa datasets from the Gene Expression Omnibus (GEO) database. Two of these datasets were merged to create a meta-data cohort, which was used as the training group. The remaining three datasets were merged into another meta-data cohort and used as the validation group. We first compared 189 PCa cases with 63 normal control cases in the training group to identify differentially expressed genes (DEGs) and performed principal component analysis (PCA) to eliminate batch effects. Weighted Gene Co-expression Network Analysis (WGCNA) was used to select important module genes. DEGs were then intersected with module genes, and functional enrichment analysis and KEGG enrichment analysis were performed. Machine learning techniques were subsequently employed to screen and identify diagnostic biomarkers for PCa, which were validated in the trial group. Next, we assessed the diagnostic efficacy of the selected diagnostic markers using the receiver operating characteristic (ROC) curve. GSVA analysis was used to analyze the pathways that these genes might enrich, and immune cell infiltration analysis was conducted to screen for core immune-related diagnostic biomarkers for PCa. Finally, we used Mendelian randomization studies to clarify the causal relationship between core biomarkers

and PCa. This study aims to provide new molecular markers for the early diagnosis and treatment of PCa and to provide a scientific basis for the early diagnosis and personalized treatment of PCa.

2 Materials and methods

2.1 Data collection and processing

All PCa datasets were obtained from the GEO public database (<http://www.ncbi.nlm.nih.gov/geo>) including the GSE46602, GSE79021, GSE200879, GSE60329, and GSE71016 datasets. These datasets were selected based on their inclusion of both tumor and normal prostate tissue samples, as well as their relevance to prostate cancer biomarker discovery. These datasets were divided into a training group (GSE46602 and GSE79021) with 189 PCa cases and 63 normal control cases, and the test group (GSE200879, GSE60329, and GSE71016) with 216 PCa cases and 70 normal control cases. Detailed information about each dataset, including sample collection methods, sources, purposes, and database curators, is provided in Table 1. Hyperlinks to the original datasets are included for further reference.

2.2 Identification of DEGs in PCa

The raw data were processed into standardized data using the R package “limma”. We screened data from 189 PCa cases and 63 normal control cases in the training group and employed the “SVA” package to correct for batch effects. The threshold for DEG selection was set as $|\log\text{Fold Change (logFC)}| > 1$ and adjusted p-value < 0.05 . Volcano plots for all DEGs and clustering heatmaps were generated using the R packages “ggplot2” and “pheatmap”, respectively.

2.3 Weighted gene co-expression network analysis

The R package “WGCNA” is used to construct a gene co-expression network and identify modules related to PCa. A subset of genes with a standard deviation greater than 0.5 is selected for further analysis. The “goodSamplesGenes” function is employed to

check for missing values in the data, to remove genes or samples that do not meet the quality standards, and to build a scale-free co-expression network. The “pickSoftThreshold” function is utilized to determine the optimal soft threshold. On this basis, the gene expression data matrix is converted into the corresponding adjacency matrix, which is then transformed into a topological overlap matrix to determine the gene weights and similarities. Gene modules are identified through topological overlap clustering methods. The module eigengenes (ME) are calculated and merged among similar modules, followed by the generation of a hierarchical clustering dendrogram. Subsequently, the Gene significance (GS) and Module Membership (MM) are computed within the modules through intramodular analysis.

2.4 Functional enrichment analysis

Using the R package “clusterProfiler,” core genes were subjected to Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, with a p-value of < 0.05 set as the criterion. To further explore the signaling pathways involved in core differential genes, we employed the Gene Set Variation Analysis (GSVA) method. The R packages “GSVA” and “enrichplot” were used for GSVA analysis of the differential genes, referring to the gene set file “c2.cp.kegg.Hs.symbols.gmt”.

2.5 Construction of a diagnostic model using machine learning methods

Machine learning is a branch of artificial intelligence, and the core of machine learning is the algorithms, which can analyze and extract patterns from data and then use these patterns to make predictions or decisions without explicit programming (15). We divided the PCa dataset into a training group and a test group and included 12 machine learning methods, including SVM, Ridge, Enet, glmBoost, Lasso, plsRglm, Stepglm, RF, GBM, LDA, XGBoost, and NaiveBayes. We constructed models with 113 combinations of the 12 algorithms in the PCa training group dataset, and all constructed models were evaluated in the PCa test group dataset. For each model, we calculated its AUC values in the training and validation groups and extracted the feature genes screened by each

TABLE 1 Fundamental details of the GEO datasets utilized in this research.

GEO series	Platform	Country	Samples		Category
			PCa	Normal	
GSE46602	GPL570	Denmark	36	14	Train group
GSE79021	GPL19370	USA	153	49	Train group
GSE200879	GPL32170	France	115	9	Test group
GSE60329	GPL14550	Italy	54	14	Test group
GSE71016	GPL16699	USA	48	47	Test group

method. Then, we ranked the predictive performance of the models based on the average AUC values to select the optimal model and the optimal model genes.

2.6 Analysis of immune cell infiltration

The CIBERSORT algorithm was utilized to assess the composition and abundance of 22 types of immune-infiltrating cells in both PCa and control samples. The R package “Cibersort” was employed for the immune cell infiltration analysis. Bar charts were used to visualize the proportion of each type of immune cell in different samples. Comparisons in the proportions of different types of immune cells between the PCa and control groups were visualized using the R package “ggpubr”. A heatmap depicting the correlations of the 22 infiltrating immune cells was plotted using the R package “corrplot”.

2.7 Correlation analysis between gene identification and immune cell infiltration

We conducted Spearman’s rank correlation analysis using R software to examine the correlation between the expression levels of the identified biomarkers and the levels of infiltrating immune cells. Finally, the results of the analysis were visualized using the R packages “ggpubr” and the “linkET” function.

2.8 Mendelian randomization analysis

We obtained expression quantitative trait locus (eQTL) data from the IEU Open GWAS project (<https://gwas.mrcieu.ac.uk/>), which covers 31,684 blood and peripheral blood mononuclear cell samples involving 19,942 genes. Concurrently, we obtained protein quantitative trait locus (pQTL) data from <https://www.decode.com/summarydate/>, encompassing 4,907 proteins in the plasma of 35,559 Icelandic individuals, and identified 18,084 sequence variations associated with plasma protein levels. We used these two datasets as exposure data for subsequent Mendelian Randomization (MR) analysis. Furthermore, the PCa outcome data “Finngen_R10_C3_PROSTATE” for MR analysis was derived from <https://www.finngen.fi/en/>, including 15,199 PCa cases and 131,266 control cases. In our study, single nucleotide polymorphisms (SNPs) were defined as instrumental variables (IVs), with the selection criteria for IVs being: (a) demonstrating a genome-wide significant association ($P < 5 \times 10^{-8}$) (16); (b) showing an independent association (linkage disequilibrium (LD) $> 10,000$ kb, $\text{clump } r^2 < 0.001$) (17). Additionally, to avoid bias caused by weak instruments, we considered those IVs with an F statistic > 10 as strong instruments and reserved them for the following analysis (18). In the causal inference of Mendelian Randomization (MR), valid instrumental variables (IVs) must satisfy three fundamental assumptions: (1) the genetic variant is directly related to the exposure; (2) the genetic variant is unrelated to potential confounders between the exposure and the outcome; (3)

the genetic variant does not affect the outcome through pathways other than the exposure.

2.9 Ethical statement and tissue collection

This study was approved by the Ethics Committee of Wuxi No. 2 Hospital affiliated with Nanjing Medical University (2022-Y-80). The PCa specimens and their paired normal tissues were obtained from patients at Wuxi No. 2. Hospital and all participants have signed informed consent forms. All patients received no endocrine therapy before surgery, and all patients underwent radical prostatectomy. The tissues were immediately preserved in liquid nitrogen.

2.10 Reagents

FastPure Cell/Tissue Total RNA Isolation Kit V2, HiScript III RT SuperMix, and ChamQ Universal SYBR qPCR Master Mix Q711-02 were all acquired from Vazyme (Nanjing, China). RIPA lysis buffer was purchased from Beyotime (Shanghai, China). Protease Inhibitor Cocktails (Cat. HY-K0010) were sourced from MedChemExpress (Shanghai, China). Anti- ARHGEF38 (Cat. ab122345) was purchased from Abcam(Shanghai, China). Antibodies against UT-B (Cat. 25962-1-AP), NEFH (Cat. 18934-1-AP), MSMB (Cat. 15888-1-AP), KRT23 (Cat. 24049-1-AP), KRT15 (Cat. 10137-1-AP) and Beta Actin (Cat. 66009-1-Ig) were purchased from Proteintech (Wuhan, China). Secondary antibodies (Cat. SA00001-1and Cat. SA00001-2) were purchased from Proteintech (Wuhan, China).

2.11 Total RNA extraction and quantitative real-time PCR

Total RNA was isolated from cells with FastPure Tissue Total RNA Isolation Kit V2 and reverse-transcribed into complementary DNAs (cDNAs) with HiScript III RT SuperMix. The cDNAs were amplified based on the standard qPCR protocol with ChamQ Universal SYBR qPCR Master Mix in an Applied Biosystems 7500 Fast Real-Time PCR System (ThermoFisher, US). Quantitative PCR was conducted at 95 °C for 30 s, followed by 40 cycles of 95 °C for 10 s and 60 °C for 30 s. GAPDH was used as the internal control and the results were calculated using $2^{-\Delta\Delta CT}$ method. The sequences of primers can be seen in [Supplementary Table 1](#).

2.12 Western blot assay

RIPA lysis buffer containing protease inhibitor was used to extract tissue protein. Protein concentration was determined with BCA Protein Assay Kit. ColorMixed Protein Marker was applied as a protein size marker. Total proteins were separated by 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis and then

transferred to a polyvinylidene difluoride (PVDF) membrane (3010040001, Roche, Germany). The PVDF membrane was then blocked with 5% milk for 2 h followed by incubation with the indicated primary antibody overnight at 4°C. After incubation with the corresponding secondary antibody at room temperature for 1.5 h, the protein bands were visualized using enhanced chemiluminescence (Tanon, China).

2.13 Immunohistochemistry

Tissues were fixed with 4% paraformaldehyde, dehydrated, embedded in paraffin and sectioned at 4µm. The sections were deparaffinized, then boiled with citrate solution, and incubated with 3% H₂O₂ for 10 min. After antigen retrieval and blocking, the slides were incubated with the indicated antibodies at 4°C overnight. Subsequently, the slides were incubated with the corresponding secondary antibody for 30 min at room temperature and then incubated with the streptavidin peroxidase complex. Staining was performed using a 3,3-diaminobenzidine (DAB) substrate kit for peroxidase reaction and counterstained with hematoxylin. Sections were washed, dehydrated, and sealed with neutral balsam. Finally, an Olympus microscope was used to acquire images.

2.13 Statistical analysis

All statistical analyses were conducted using Perl version 5.38.2 and R software version 4.3.1. A P-value of less than 0.05 was used to determine statistical significance. In MR analysis, we employed five methods, including “MR Egger,” “Weighted Median,” “Inverse Variance Weighting (IVW),” “simple mode,” and “weighted mode.” Among these, IVW is the primary method for causal inference due to its superior precision and stability. A P-value of less than 0.05 was considered indicative of a significant correlation between exposure and outcome. We conducted sensitivity and heterogeneity tests based on the Q statistic. The MR-egger regression test was utilized to assess the presence of horizontal pleiotropy based on its intercept term. MR-PRESSO compares the observed distance of all variants to the regression line with the expected distance under the null hypothesis of no horizontal pleiotropy. To evaluate whether the removal of individual influential SNPs influenced the overall estimate of causal effects, leave-one-out analyses were performed.

3 Results

3.1 Identification of DEGs in PCa

DEGs in the GEO database (GSE46602 and GSE79021) were identified in 189 cases of PCa and 63 normal controls using the R

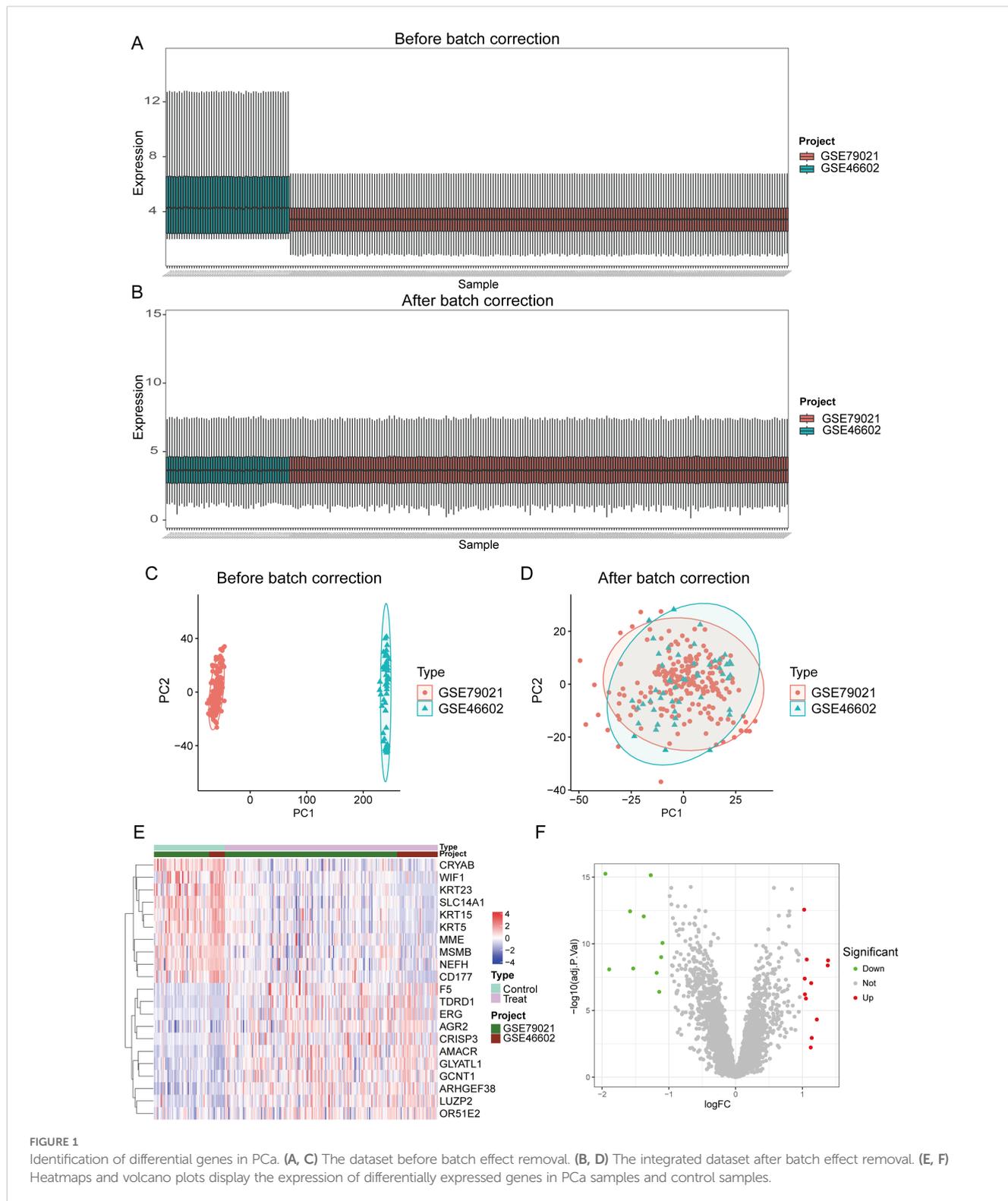
package “limma”. Due to batch effects present between different datasets (as shown in [Figures 1A, C](#)), after removal, the distribution of data across different datasets became significantly more uniform, with medians aligned and both mean and variance showing stronger consistency (as shown in [Figures 1B, D](#)). This indicates that after adjustment, the integrated data can be considered as a unified batch for subsequent analysis. Based on this, we identified a total of 21 DEGs in the training group, 11 of which were upregulated and 10 were downregulated. The expression levels of the 21 DEGs are displayed in heatmaps and volcano plots ([Figures 1E, F](#)).

3.2 Weighted gene co-expression network analysis and identification of key modules

We constructed a scale-free co-expression network using WGCNA to identify the most highly correlated modules in PCa. The optimal “soft” threshold $\beta = 4$ (scale-free $R^2 = 0.9$) was selected based on scale independence and mean connectivity ([Figures 2A, B](#)). A total of 5 modules with different colors were obtained through dynamic tree shearing ([Figure 2C](#)). Then, the modules were correlated with clinical characteristics to further identify the driver genes of the most strongly positively correlated module. We ultimately identified the ME turquoise module as being most highly correlated with DCM in PCa ($R = 0.39$, $p < 0.001$) ([Figure 2D](#)). A significant distribution of Gene Significance (GS) was observed across the five modules ([Figure 2E](#)). Therefore, we selected the turquoise module, which contains 1,195 genes, for further analysis. Additionally, we performed a correlation analysis between module membership (MM) and gene significance (GS), and found a significant positive correlation between them ($R = 0.84$, $p < 0.001$; [Figure 2F](#)).

3.3 Functional enrichment analysis

We intersected the DEGs with the WGCNA module genes to obtain a total of 20 genes ([Figure 3A](#)). To further assess the potential regulatory pathways of these genes in PCa, we conducted GO and KEGG enrichment analyses. The GO analysis indicated that these genes are primarily enriched in the biological process (BP) terms “intermediate filament organization,” “intermediate filament cytoskeleton organization,” and “intermediate filament-based process.” Regarding the cellular component (CC) ontology, these genes are involved in “intermediate filament” and “intermediate filament cytoskeleton.” The molecular function (MF) analysis showed that the genes are enriched in “structural constituent of cytoskeleton” and “scaffold protein binding” ([Figure 3B](#)). The KEGG enrichment analysis demonstrated that these genes participate in the “Staphylococcus aureus infection” and “Estrogen signaling pathway” ([Figure 3C](#)).



3.4 Identification of candidate core genes through machine learning methods

To construct a diagnostic model for PCa-related genes, we utilized a combination of 113 machine learning algorithms to analyze the previously selected 20 genes. In the training group

dataset, we calculated that “RF” produced the best predictive performance among the 113 tested algorithms, and this model also performed well in the test group dataset. Furthermore, we calculated the AUC values for each model in both the training and validation groups, with the RF machine learning algorithm achieving the highest AUC value (Figure 4A). Therefore, we

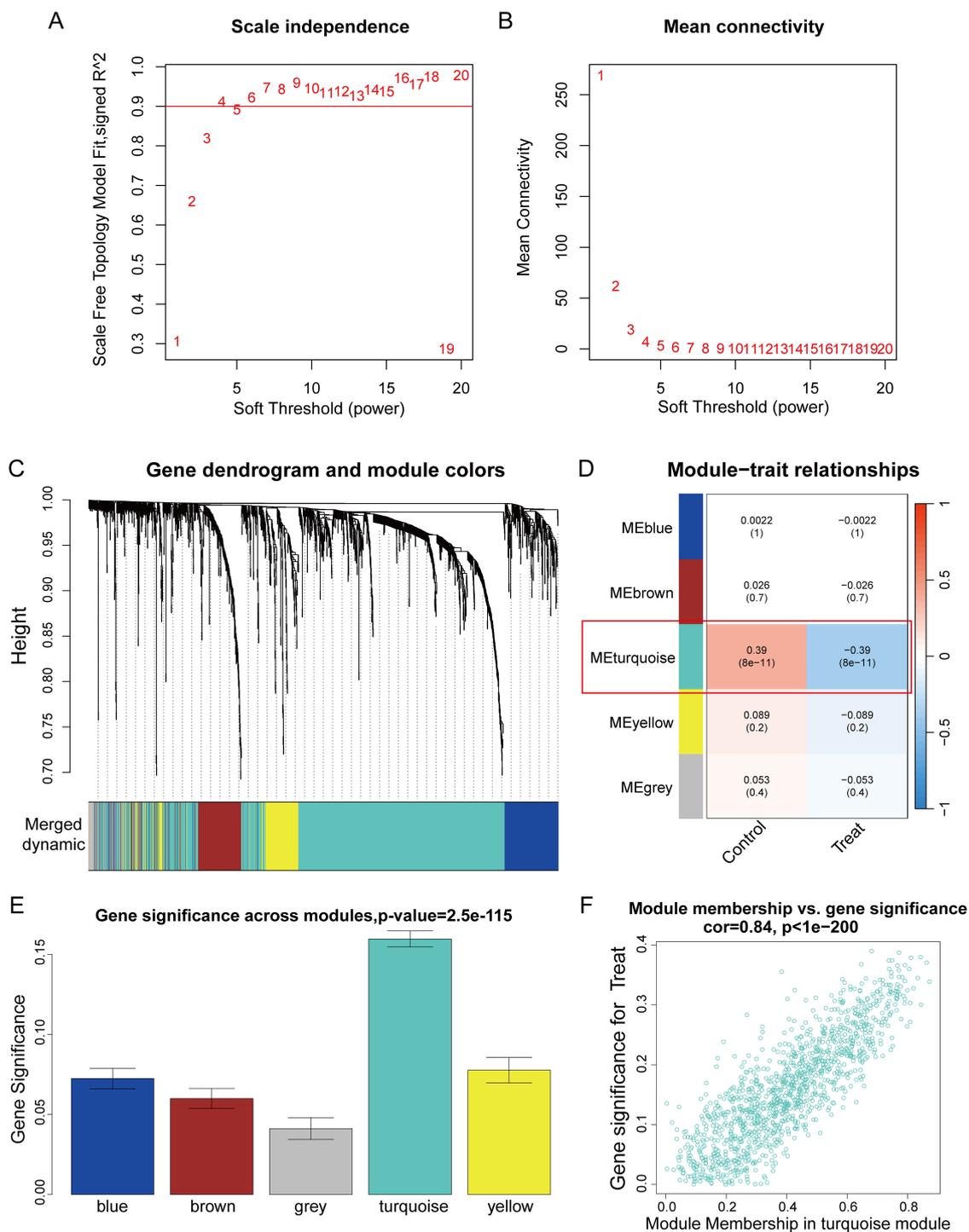


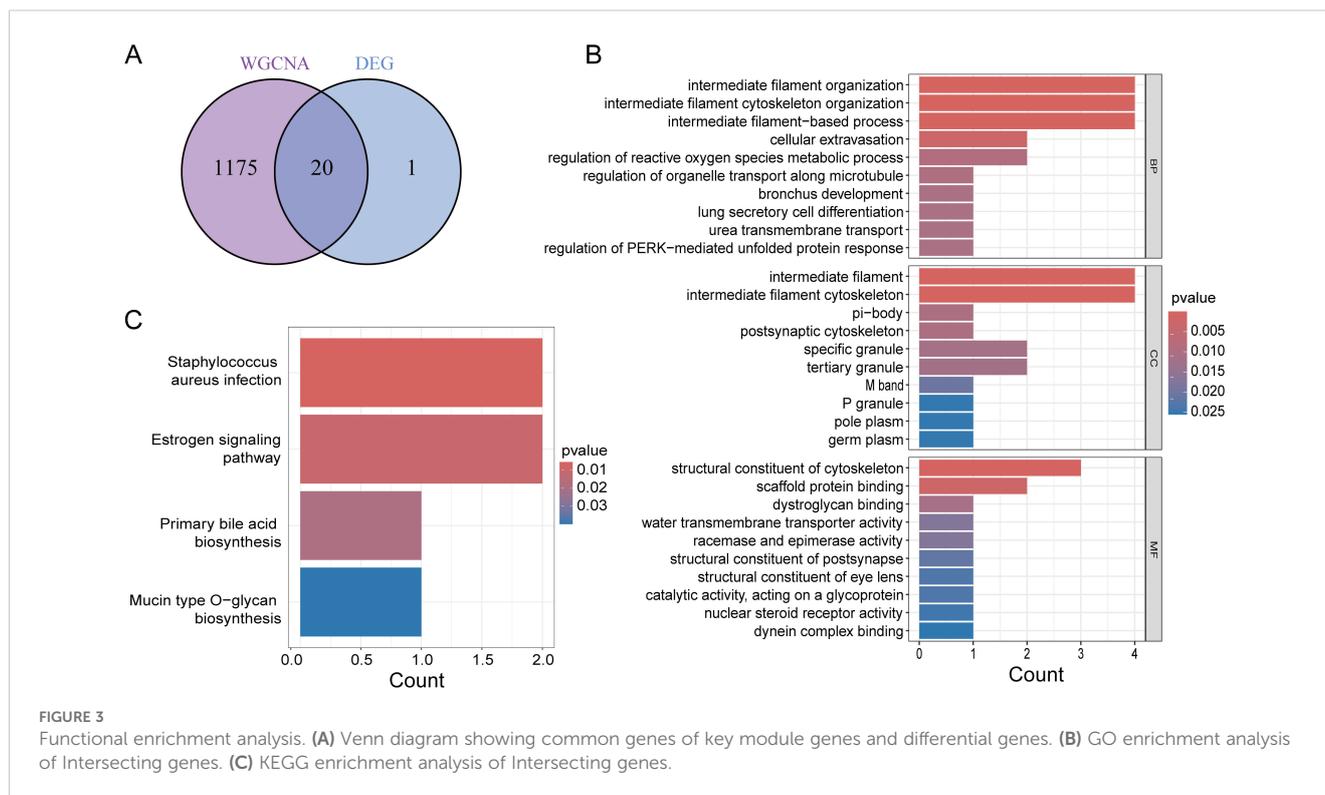
FIGURE 2

WGCNA analysis. (A) The scale-free topology fit index (R^2) for soft threshold powers (β) ranging from 1 to 20. (B) The mean connectivity for β from 1 to 20. (C) Dendrogram of all genes clustered based on topological overlap. (D) Correlation heatmap of modules and clinical traits. (E) Boxplot of Gene Significance (GS) distribution across five modules. (F) Scatterplot of Module Membership (MM) in the turquoise module versus GS for PCa.

identified RF as the optimal diagnostic model. Additionally, we plotted the ROC curves for the RF model in the training and validation groups across three datasets (Figures 4B–E). The RF model identified six genes (SLC14A1, ARHGEF38, NEFH, MSMB, KRT23, KRT15) as the final validated core genes (Figure 4F).

3.5 Diagnostic assessment and validation of core genes

To assess the diagnostic value of the six candidate core genes selected, we established ROC curves to evaluate the diagnostic



specificity and sensitivity of each gene. The results are as follows: SLC14A1 (AUC 0.862), ARHGEF38 (AUC 0.817), NEFH (AUC 0.805), MSMB (AUC 0.784), KRT23 (AUC 0.728), KRT15 (AUC 0.824), all candidate genes have a high diagnostic value for PCa (Figure 5A). To delve deeper into the biological significance of these six core genes, we utilized GeneMANIA to analyze the relationships and interaction patterns between these genes and other closely related genes. we found 20 genes that are closely related to these 6 genes (Figure 5B). Next, we applied the Gene Set Variation Analysis (GSVA) method to analyze the signaling pathways in the KEGG database to assess the enrichment of each gene in these pathways. We found that the genes SLC14A1, MSMB, KRT23, and KRT15 are primarily enriched in “prostate cancer” signaling pathway (Supplementary Figure S1). We also analyzed the expression level differences of these core genes in PCa patients and normal populations. The results showed that SLC14A1, NEFH, MSMB, KRT23, and KRT15 are underexpressed in PCa, while ARHGEF38 is highly expressed in PCa (Figure 5C). In addition, we also analyzed the correlations between these core genes and found that SLC14A1, NEFH, MSMB, KRT23, and KRT15 genes are all positively correlated with each other, while ARHGEF38 is negatively correlated with the other five genes (Figure 5D).

3.6 Immune cell infiltration analysis

We used the CIBERSORT algorithm to evaluate the composition and abundance of 22 types of immune cell infiltration in the training group (Figure 6A). In addition, the correlation analysis of the 22 types of immune cells showed that

Dendritic cells resting were positively correlated with Macrophages M1 ($r = 0.62$), T cells regulatory were positively correlated with NK cells activated ($r = 0.50$), while T cells CD4 memory resting were negatively correlated with T cells follicular helper ($r = -0.70$) (Figure 6B). The box plots showed that compared to the control group, PCa patients had higher levels of Plasma cells and Monocytes, while the levels of Mast cells resting were lower (Figure 6C).

3.7 Correlation analysis between PCa-associated core genes and immune infiltrating cells

To validate the relationship between the selected core genes and immune cell infiltration, we conducted a correlation analysis. The results indicated that SLC14A1 is positively correlated with Dendritic cells activated ($R = 0.48$, $P = 0.027$) and NK cells resting ($R = 0.45$, $P = 0.039$), and negatively correlated with Dendritic cells resting ($R = -0.45$, $P = 0.039$) and Macrophages M1 ($R = -0.52$, $P = 0.016$) (Figure 7A); ARHGEF38 is negatively correlated with T cells regulatory (Tregs) ($R = -0.52$, $P = 0.017$) and Mast cells resting ($R = -0.56$, $P = 0.009$) (Figure 7B); NEFH is positively correlated with NK cells activated ($R = 0.46$, $P = 0.038$) and negatively correlated with T cells CD4 memory activated ($R = -0.58$, $P = 0.006$) (Figure 7C); MSMB shows no significant correlation with immune cell infiltration (Figure 7D); KRT23 is positively correlated with Monocytes ($R = 0.47$, $P = 0.031$) and NK cells resting ($R = 0.45$, $P = 0.040$), and negatively correlated with Macrophages M1 ($R = -0.45$, $P = 0.039$), Dendritic cells resting ($R =$

-0.48, $P = 0.029$), T cells CD4 memory activated ($R = -0.49$, $P = 0.025$), and B cells memory ($R = -0.52$, $P = 0.015$) (Figure 7E); KRT15 is positively correlated with NK cells resting ($R = 0.48$, $P = 0.028$) and negatively correlated with Macrophages M1 ($R = -0.45$, $P = 0.039$) (Figure 7F). Additionally, we visualized the correlations between the expression of all core genes and the infiltration of 22 immune cells (Figure 7G). These findings suggest that the associated core genes may contribute to PCa by influencing immune cell infiltration.

3.8 MR analysis of the association between core genes and PCa

In our study, we initially conducted a two-sample MR analysis using eQTL data of six core genes combined with PCa data. Unfortunately, this analysis did not reveal any causal associations between these genes and PCa. Subsequently, we turned to another round of two-sample MR analysis using the pQTL data of these core genes along with PCa data. In this analysis, we found that only the

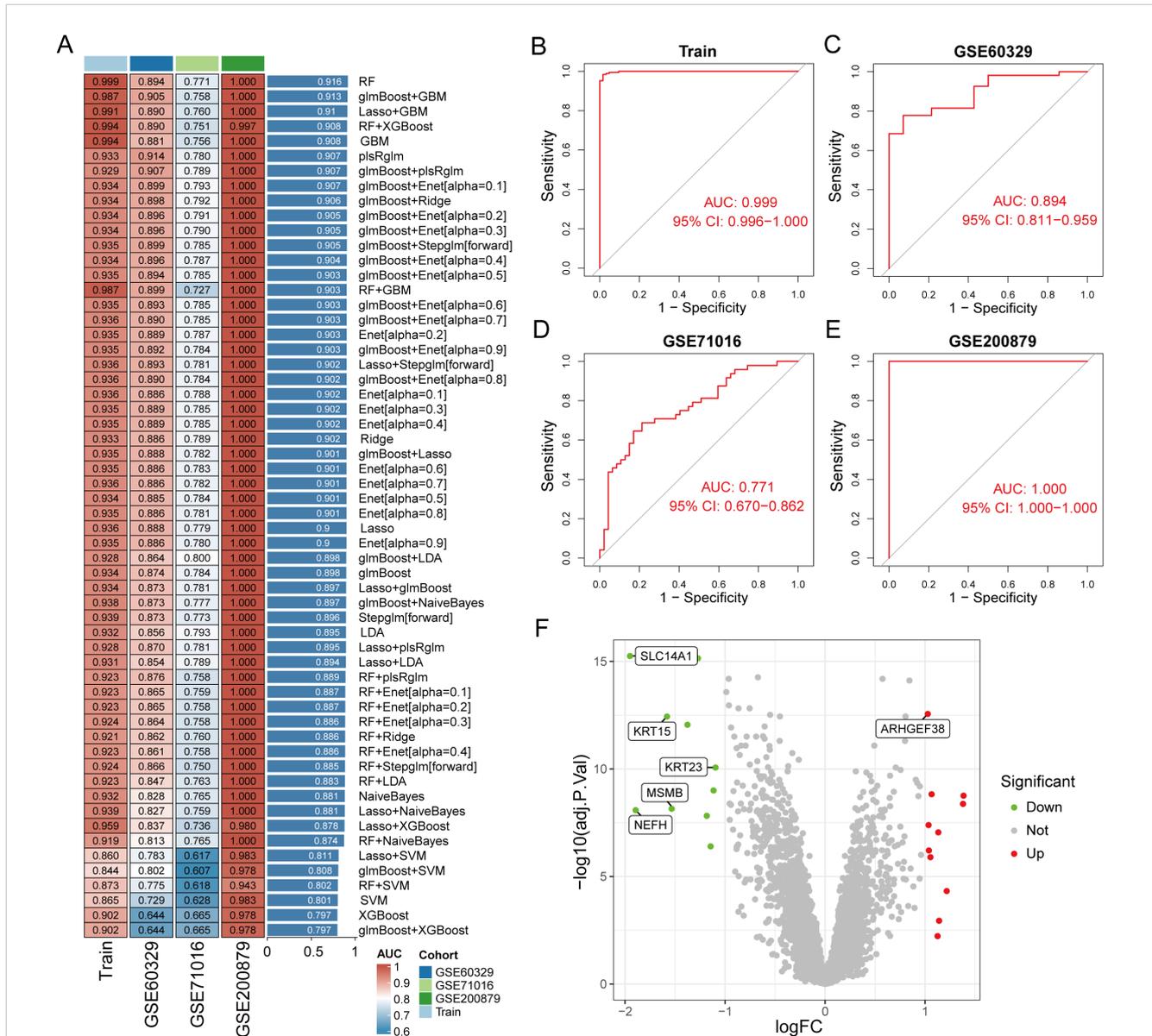


FIGURE 4 Identification of core genes in prostate cancer using machine learning methods. (A) Constructing a prostate cancer diagnostic prediction model with combinations of 113 machine learning algorithms in training and test group. (B–E) ROC curves of RF model for training and test group. (F) Volcano plot of the 6 core genes identified by the RF model.

MSMB pQTL had a significant causal relationship with the incidence of PCa (OR: 0.8589, 95% CI: 0.8085-0.9125, $P < 0.05$) (Figure 8A). This indicates that the expression of MSMB is a major protective factor in the pathogenesis of PCa, while the other core genes may only appear as biomarkers after the occurrence of PCa. The forest plot shows the effect sizes of various SNPs in the MSMB pQTL (Figure 8B). The scatter plot shows the regression trends of the five methods, all of which are consistent (Figure 8C). The funnel plot shows the distribution of causal effects (Figure 8D). The leave-one-out analysis reveals the impact of individual SNPs on the overall causal estimate (Figure 8E). The MR-Egger intercept analysis for the MSMB pQTL did not detect any potential horizontal pleiotropy ($P = 0.55$), indicating that the instrumental variables do not significantly affect the outcome through pathways other than the exposure. In the Cochran Q test for heterogeneity, heterogeneity was observed in the MSMB pQTL data ($P = 0.013$). This may be due to IVs originating from different analysis platforms, different studies or datasets, and different populations.

3.9 Validation of core gene expression in prostate cancer and adjacent non-cancerous tissues

We measured the mRNA expression levels of six core genes in PCa tumor tissues and their corresponding para-cancerous tissues. Our results revealed that, compared to para-cancerous tissues, the expression levels of SLC14A1, NEFH, MSMB, KRT23, and KRT15 were significantly downregulated in PCa tumor tissues, while ARHGEF38 expression was notably upregulated, consistent with our initial predictions (Figures 9A–F). To further validate these findings at the protein level, we conducted Western blot analysis on tumor tissues and adjacent normal tissues from PCa patients. The results demonstrated that ARHGEF38 protein expression was significantly elevated in tumor tissues. In contrast, the expression levels of UT-B (encoded by SLC14A1), NEFH, MSMB, KRT23, and KRT15 were reduced in tumor tissues compared to adjacent normal tissues, although the decrease in UT-B did not reach statistical

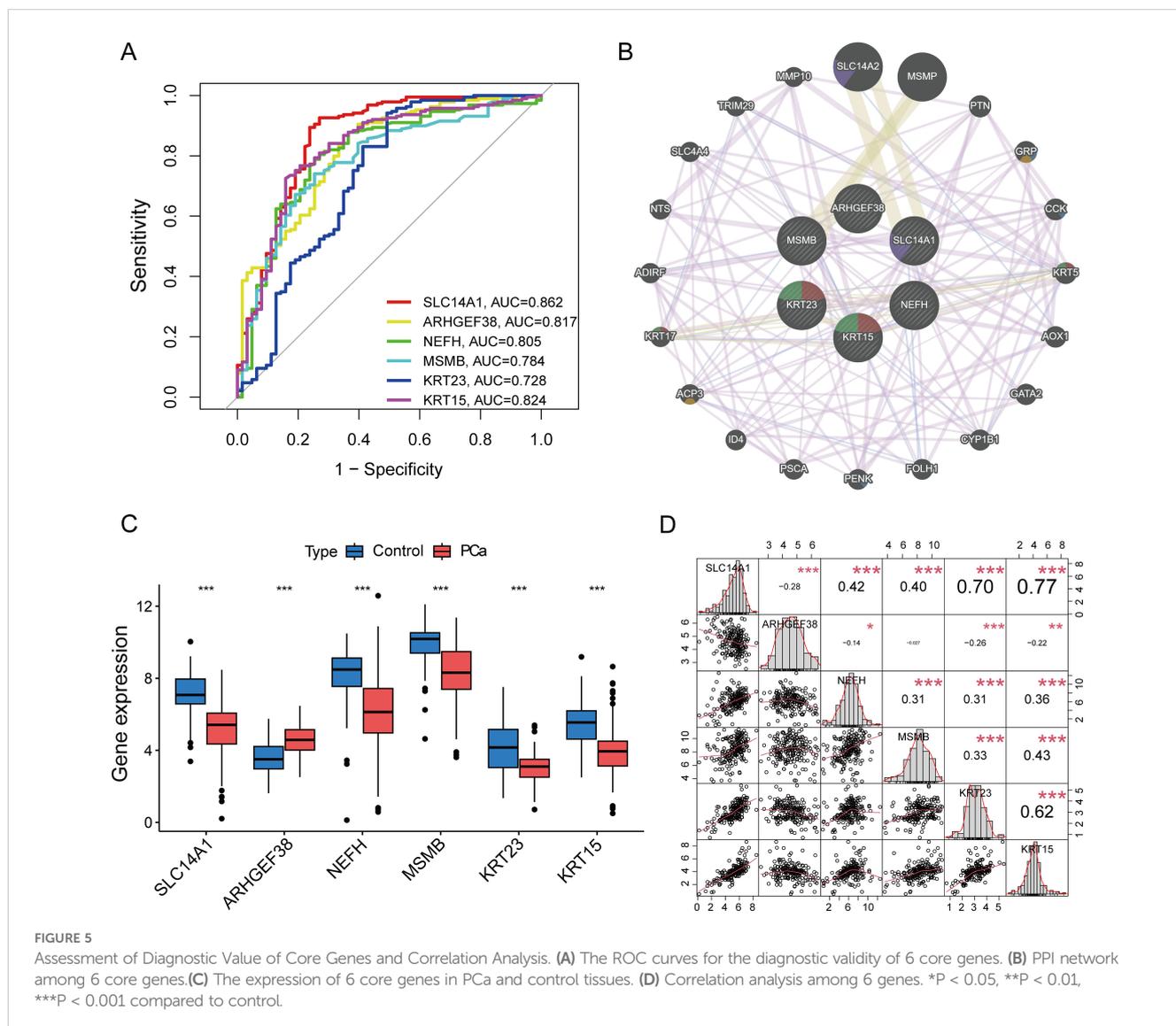


FIGURE 5 Assessment of Diagnostic Value of Core Genes and Correlation Analysis. (A) The ROC curves for the diagnostic validity of 6 core genes. (B) PPI network among 6 core genes. (C) The expression of 6 core genes in PCa and control tissues. (D) Correlation analysis among 6 genes. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ compared to control.

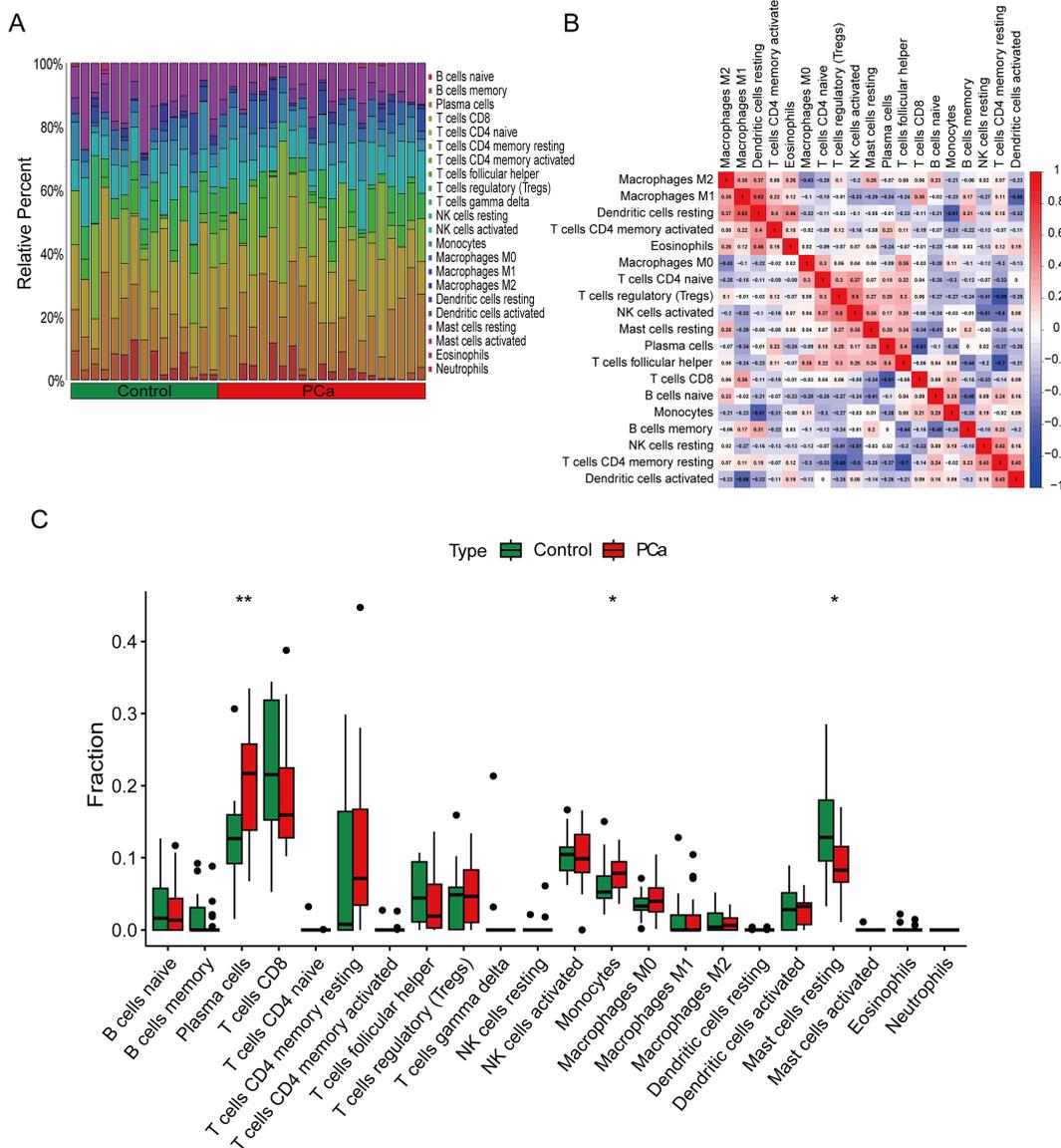


FIGURE 6 Immune cell infiltration analysis between PCa and control. **(A)** The box-plot diagram indicating the proportion of 22 kinds of immune cells between PCa and normal controls. **(B)** The heat map illustrates the correlation among 22 different immune cell populations. **(C)** The box plot shows the differences in immune infiltration between prostate cancer and normal controls. *P < 0.05, **P < 0.01 compared to control.

significance (Figure 9G). Quantitative analysis of the Western blot results is presented in Figure 9H, further supporting the observed protein expression trends.

Additionally, immunohistochemistry (IHC) was performed to evaluate the localization and expression levels of two key proteins in PCa and para-cancerous tissues. The IHC results revealed that ARHGEF38 protein was highly expressed in tumor tissues, while MSMB exhibited reduced expression in tumor tissues compared to para-cancerous tissues (Figure 9I). These findings were consistent with the mRNA and protein expression patterns observed in qRT-PCR and Western blot analyses.

4 Discussion

PCa is one of the most common malignant tumors in the male reproductive system and is the fifth leading cause of cancer-related deaths in men. Its pathogenesis is complex, involving a variety of factors such as genetics, environment, and lifestyle. Although the causes of PCa are not yet clear, early prevention and treatment remain key to improving prognosis and prolonging survival. PCa-related biomarkers play an important role in the early diagnosis, prognosis assessment, and monitoring of treatment response in PCa (19). In recent years, with the advancement of molecular biology

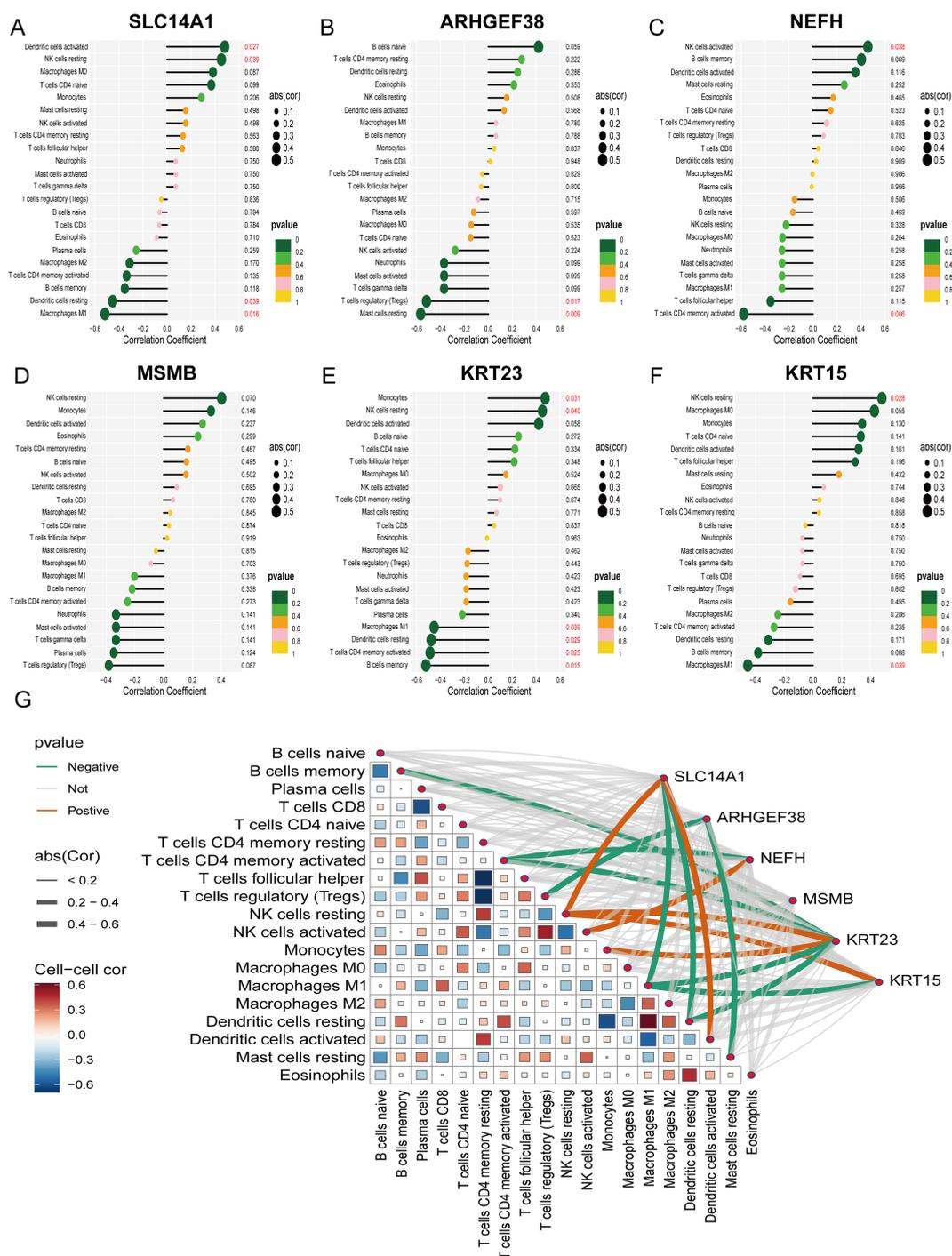


FIGURE 7 Correlation analysis of 6 PCA-related core genes with infiltrating immune cells. (A–F) Correlation between SLC14A1, ARHGEF38, NEFH, MSMB, KRT23, KRT15, and infiltrating immune cells. (G) A visualization chart of the correlations between 6 core genes and the infiltration of 22 types of immune cells.

and genetic technology, more and more PCA-related biomarkers have been discovered and applied in clinical and research settings. For example, Srikantan et al. (20) identified a prostate-specific gene PCGEM1 that is expressed only in human prostate tissue and is upregulated in the vast majority of tumor tissues by comparative analysis of differentially expressed genes between normal prostate

tissue and PCA tissue. This indicates that PCGEM1 may be involved in the occurrence and progression of PCA. Wei et al. (21) assessed the value of urine prostate cancer antigen 3 (PCA3) for screening PCA in men, and the results showed that urine PCA3 plays an important role in reducing the medical burden of men undergoing repeat prostate biopsies, and for men who have not undergone a

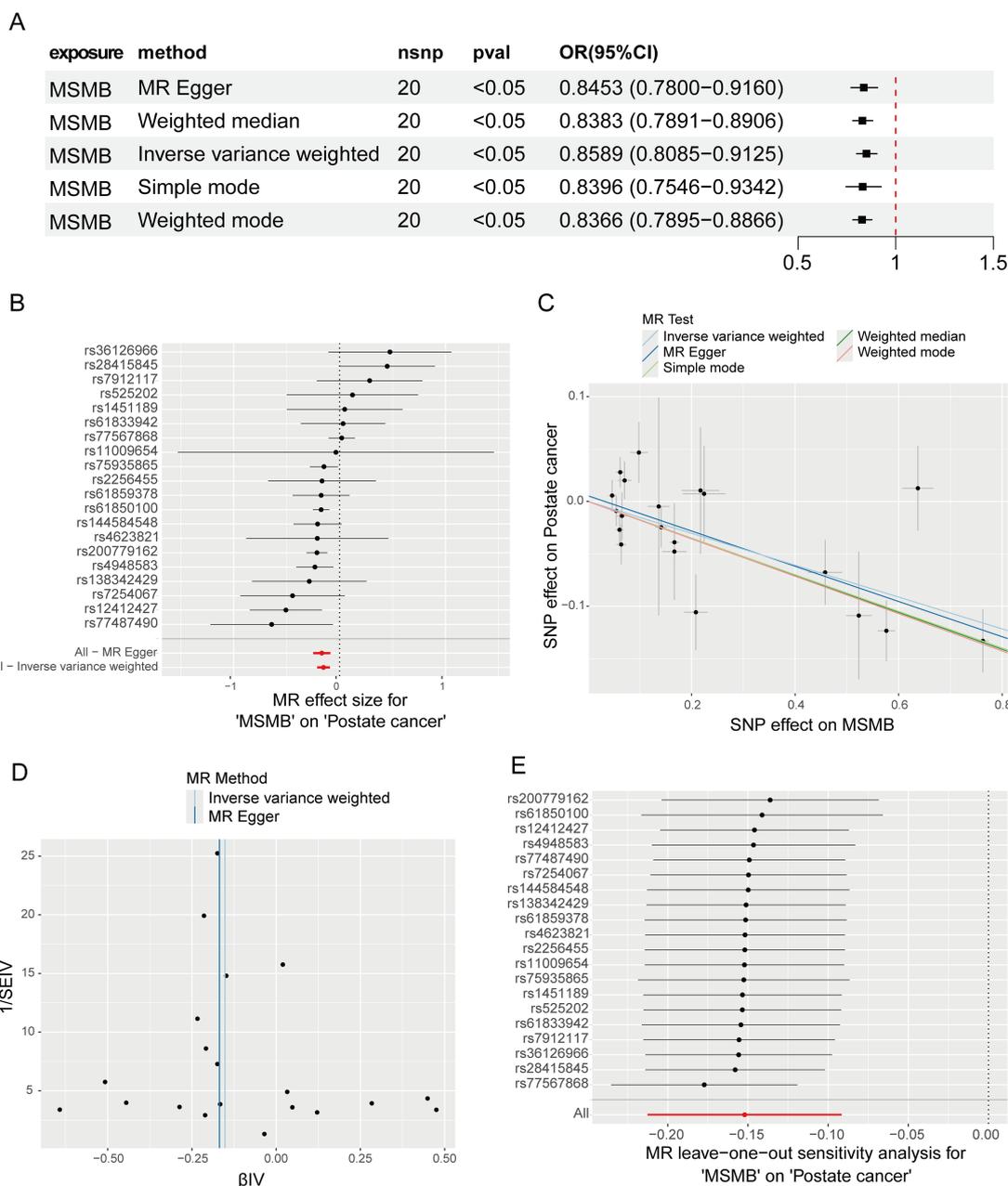


FIGURE 8 Mendelian randomization study results on the relationship between MSMB and PCa. **(A)** MR analysis demonstrated the causal relationship of MSMB pQTL on PCa. **(B–E)** Forest plot, Funnel plot, Scatter plot and Leave-one-out analysis of the causal association between MSMB pQTL and PCa.

biopsy, a high PCA3 score significantly increases the likelihood of detecting cancer in the initial prostate biopsy. We obtained five datasets from the GEO database, of which two were combined into the training group, and the remaining three were combined into the test group. In the training group, we identified 21 DEGs, of which 11 genes were upregulated and 10 genes were downregulated. Additionally, KEGG enrichment analysis revealed that these differentially expressed genes are involved in pathways related to “Staphylococcus aureus infection.” The identification of shared genetic pathways between PCa and Staphylococcus aureus infection suggests potential interactions between microbial

infections and PCa progression. For example, *S. aureus* may play a role in modulating the tumor immune microenvironment, thereby influencing prostate cancer pathogenesis (22, 23). For example, *S. aureus* may influence the progression and metastasis of prostate cancer by modulating the tumor immune microenvironment. These genetic overlaps may play an important role in the regulation of microbial infections and the immune microenvironment. Further studies are needed to explore these mechanisms in detail.

Subsequently, we used WGCNA and 12 machine learning methods to screen for PCa-related core genes. Ultimately, we

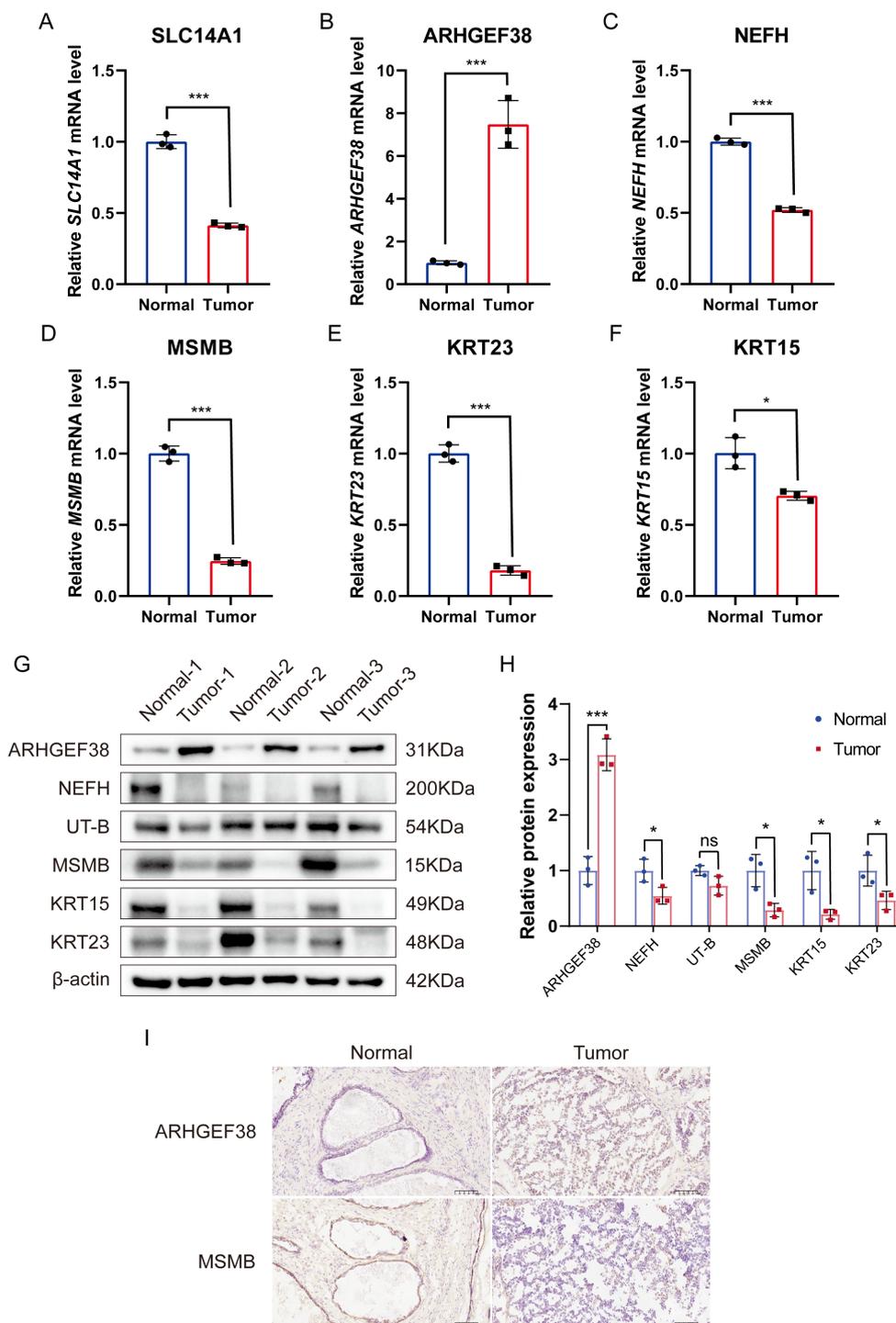


FIGURE 9 Validation of mRNA and protein expression levels of six core genes in PCa tumor tissues and adjacent non-cancerous tissues. **(A–F)** mRNA expression levels of SLC14A1, ARHGEF38, NEFH, MSMB, KRT23, and KRT15 in PCa tumor tissues and adjacent non-cancerous tissues, as determined by qRT-PCR. (n=3). **(G)** Protein expression levels of UT-B (encoded by SLC14A1), ARHGEF38, NEFH, MSMB, KRT23, and KRT15 in PCa tumor tissues and adjacent non-cancerous tissues, as determined by Western blot. **(H)** Quantitative analysis of Western blot results. **(I)** Immunohistochemistry (IHC) analysis of ARHGEF38 and MSMB protein expression in PCa tumor tissues and adjacent non-cancerous tissues. Data are presented as mean ± SEM. ns P>0.05, *P < 0.05, ***P < 0.001 compared to normal tissues.

identified Random Forest (RF) as the optimal machine learning method and discerned six core genes from it, including SLC14A1, ARHGEF38, NEFH, MSMB, KRT23, and KRT15. To validate whether these genes could act as diagnostic biomarkers, we analyzed their diagnostic value in PCa. The results demonstrated that all of these genes exhibited certain diagnostic potential, with AUC values all exceeding 0.7.

Previous studies have indicated that immune cell infiltration is closely related to the development and prognosis of PCa. In the tumor microenvironment of PCa, the mechanism and interaction of immune cells are very complex, and they play significant roles in the occurrence and progression of cancer (24). Our study used CIBERSORT to evaluate the infiltration of immune cells in PCa, and the results showed an increase in the expression of Plasma cells and Monocytes, while the expression of Mast cells resting decreased in PCa patients. In addition, we analyzed the relationship between the six core genes and PCa-related immune cells, and the results revealed that SLC14A1, NEFH, KRT15, and KRT23 were significantly correlated with NK cells. Furthermore, SLC14A1, KRT15, and KRT23 were significantly associated with Macrophages M1. Natural Killer (NK) cells are a subset of lymphocytes that play a central role in the innate immune response to tumors and infections. NK cells can detect changes in the expression of self MHC-I molecules on the surface of autologous cells and maintain cytotoxic activity against tumors with high expression of MHC-I (25). A follow-up study over 11 years indicated that low NK cell cytotoxic activity is associated with an increased risk of cancer (26). Pasero et al. (27) retrospectively analyzed natural killer (NK) cells in the peripheral blood of 39 patients with metastatic PCa and found that patients with longer survival and castration-resistant time showed high expression of NK cell receptors (NKP46 and NKP30). Seki et al. (28) demonstrated that NK cells can target PCa stem-like cells through a pathway, thereby inhibiting the occurrence, progression, metastasis, and recurrence of PCa. Tumor-associated macrophages are one of the major types of immune cells in the TME, and an increase in their density is associated with poor prognosis in PCa. Kainulainen et al. (29) showed that the secreted factors of M1 macrophages can upregulate the plasticity of PCa stem cells through the NF- κ B signaling pathway, thereby affecting the progression and drug resistance of PCa. Hadimani et al. (30) analyzed the expression of CD68 and CD163 in tumor tissues from 62 patients with PCa to evaluate the expression of tumor-associated macrophages M1 and M2 in PCa and explore their association with tumor stage. The results showed that CD68 expression was associated with a good prognosis with fewer lymph node and distant metastases, while CD163 expression was associated with a poor prognosis and an increased likelihood of lymph node and distant metastases. These studies, including our own, indicate that various types of immune cells play important roles in PCa, providing new insights for immunotherapy of PCa.

Although we have identified 6 core genes associated with PCa, the genetic associations of these genes with PCa still need to be studied in depth. Therefore, I employed Mendelian randomization analysis to further explore the potential causal effects of these

genetic variants on the risk of developing PCa. Our study did not find any significant association between the eQTL data of any genes and PCa; however, we did find that the pQTL data of the MSMB gene showed a significantly negative correlation with the occurrence of PCa. This finding suggests that the expression level of the MSMB gene may have a significant impact on the risk of developing PCa.

In this study, we systematically validated the expression patterns of six core genes (SLC14A1, ARHGEF38, NEFH, MSMB, KRT23, and KRT15) in PCa tissues and adjacent non-cancerous tissues using qRT-PCR, Western blot, and IHC. Our qRT-PCR results demonstrated that SLC14A1, NEFH, MSMB, KRT23, and KRT15 were significantly downregulated in PCa tissues, while ARHGEF38 was markedly upregulated. These findings were further corroborated at the protein level through Western blot analysis, which revealed consistent expression trends for ARHGEF38, NEFH, MSMB, KRT23, and KRT15. Although the reduction in UT-B (encoded by SLC14A1) protein expression did not reach statistical significance, the overall trend aligned with the mRNA data, suggesting a potential role of these genes in PCa pathogenesis. This lack of statistical significance for UT-B may be attributed to the relatively small sample size, which could limit the power to detect subtle differences in protein expression. Future studies with larger cohorts are needed to confirm these findings and further explore the functional implications of UT-B in PCa.

To strengthen our findings, we performed IHC analysis for two key proteins, ARHGEF38 and MSMB. The IHC results confirmed that ARHGEF38 protein was highly expressed in tumor tissues, while MSMB expression was markedly reduced, consistent with the qRT-PCR and Western blot data. These results not only validate the dysregulation of these genes at both mRNA and protein levels but also highlight their potential as biomarkers or therapeutic targets in PCa. The consistent downregulation of MSMB, a known tumor suppressor in prostate cancer, further supports its role in inhibiting tumor progression. Conversely, the upregulation of ARHGEF38, a gene implicated in cell signaling and cytoskeletal reorganization, suggests its potential involvement in promoting tumor growth and metastasis.

This study indicates that the genes SLC14A1, ARHGEF38, NEFH, MSMB, KRT23, and KRT15 hold significant diagnostic value in the onset and progression of PCa. SLC14A1, a member of the urea transporter family and a type B urea transporter (UT-B), plays a crucial role in regulating urine concentration (31). Furthermore, the SLC14A1 gene is also considered a novel tumor suppressor for urothelial carcinoma, and in PCa, the downregulation of its expression promotes tumor progression by activating the CDK1/CCNB1 and mTOR pathways and is closely associated with biochemical recurrence (BCR) of PCa (32–34). ARHGEF38 is a component of the human Rho guanine nucleotide exchange factor family, which plays an important role in cell migration and tumorigenesis. Studies have shown that knocking down the expression of ARHGEF38 can inhibit the proliferation, migration, and invasion of PCa cells, suggesting that ARHGEF38 may promote the progression of PCa (35). The protein encoded by the NEFH gene is a key component of the neurofilament cytoskeleton and belongs to the family of

intermediate filament proteins; its gene mutations are associated with various neurological diseases (36). MSMB (β -microseminoprotein), primarily expressed in the prostate, is an abundant component in semen and is considered one of the biomarkers for PCa, holding significant value in the diagnosis and treatment of PCa (37). KRT23 is a member of the keratin family and an acidic keratin expressed in various cancer types, including pancreatic, colorectal, and hepatocellular carcinomas (38–40). Research by Wang et al. (41) indicates that the knockout of the KRT23 gene can enhance the metastatic and invasive capabilities of PCa DU145 cells. KRT15, also a member of the keratin family, is an intermediate filament protein expressed in various epithelial cells and plays an essential role in maintaining cellular structural stability, proliferation, and differentiation. Research by Xiao et al. (42) shows that KRT15 is expressed at lower levels in PCa tissues and cell lines (LNCaP, DU145, and PC3), and the low expression of KRT15 is correlated with higher pathological staging and Gleason scores. Our study results are consistent with existing research, indicating that these genes may significantly impact the development of PCa and can be used for early diagnosis. In particular, MSMB shows a strong causal relationship in the Mendelian randomization analysis related to the occurrence of PCa and may become an important protective factor for PCa. However, it is important to clarify that these biomarkers are not intended to replace biopsy as the gold standard for definitive diagnosis. Instead, they aim to serve as non-invasive supplementary tools to reduce the reliance on biopsy, particularly in cases where PSA levels are ambiguous or patients are at high risk of complications from invasive procedures. By integrating these biomarkers into existing diagnostic workflows, clinicians can make more informed decisions about whether a biopsy is necessary, thereby minimizing unnecessary invasive interventions.

This study has certain limitations. Firstly, the sample size of the datasets used in this study is limited, which means that our findings need to be validated in larger datasets and clinical trials to enhance the reliability of the research results. Secondly, while our qRT-PCR, Western blot and IHC experiments confirmed the differential expression of the identified biomarkers, further functional studies are needed to elucidate their roles in prostate cancer progression and immune modulation. Future research should include *in vitro* experiments, such as gene knockdown or overexpression, and *in vivo* studies using animal models to validate the mechanistic roles of these biomarkers. Thirdly, our MR analysis suggests that MSMB is a protective factor for PCa; however, the presence of heterogeneity may limit the strength and reliability of the research conclusions. This could be due to the IVs originating from different analysis platforms, different studies or datasets, and different populations. Future studies should employ more stringent criteria for selecting instrumental variables to enhance the robustness of MR findings.

In summary, our study has made a significant contribution to prostate cancer research by combining advanced computational methods with experimental validation, especially in the context of personalized treatment strategies. The genes SLC14A1, ARHGEF38, NEFH, MSMB, KRT23, and KRT15 have been

identified as biomarkers associated with PCa. Mendelian randomization analysis indicates that MSMB is an important protective factor related to the occurrence of PCa. These biomarkers could be used to stratify patients based on their risk profiles, enabling more targeted and effective interventions. Additionally, these biomarkers may serve as non-invasive tools for early diagnosis, reducing the need for invasive procedures like biopsies. However, addressing the noted limitations, including expanding functional validation and exploring comparative analyses with PSA, will be essential for translating these findings into clinical practice. Future studies should include longitudinal analyses to assess changes in biomarker expression across different stages of prostate cancer. Such analyses could validate the prognostic utility of these biomarkers and provide insights into their roles in disease progression.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#). Further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by Ethics Committee of Wuxi No. 2 Hospital affiliated with Nanjing Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JL: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. YZ: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – review & editing. SJ: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – review & editing. CF: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Writing – review & editing. BL: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – review & editing. YS: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – review & editing. ML: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review

& editing. YZ: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing. NF: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the National Natural Science Foundation of China (Item No:82370777, and No. 82302654), Major Project of Jiangsu Commission of Health: Basic research and clinical application of biomarkers for early diagnosis of prostate cancer (Item No: ZD2021002), Wuxi Taihu Talent Plan, Team in Medical and Health High-level Talents Project: Basic and clinical study of urothelial carcinoma of upper urinary tract (Item No: THRCJH20200406), National Natural Science Foundation of Jiangsu Province (No. SBK2023044153), “San Ming” Strategic Talent Project of Wuxi Municipal Health Commission (No. HB2023032), Science and Technology Re-search Project of Wuxi Science and Technology Bureau (No. K20221029), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (Item No: KYCX24_3596).

Acknowledgments

We thank all the members who participated in this study. We sincerely acknowledges Professor Ke Wang from the Hospital of

Qingdao University for providing the critical antibodies and valuable technical guidance essential to this research. His support played a pivotal role in the successful execution and completion of the experiments. We are grateful to all the relevant online databases covered in the present study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2025.1534612/full#supplementary-material>

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin.* (2024) 74:12–49. doi: 10.3322/caac.21820
2. Bergengren O, Pekala KR, Matsoukas K, Fainberg J, Mungovan SF, Bratt O, et al. 2022 Update on prostate cancer epidemiology and risk factors-A systematic review. *Eur Urol.* (2023) 84:191–206. doi: 10.1016/j.eururo.2023.04.021
3. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2024) 74:229–63. doi: 10.3322/caac.21834
4. Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJ, et al. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N Engl J Med.* (1991) 324:1156–61. doi: 10.1056/NEJM199104253241702
5. Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS, Redwine E. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *N Engl J Med.* (1987) 317:909–16. doi: 10.1056/NEJM198710083171501
6. Hugosson J, Roobol MJ, Månsson M, Tammela TLJ, Zappa M, Nelen V, et al. A 16-yr follow-up of the european randomized study of screening for prostate cancer. *Eur Urol.* (2019) 76:43–51. doi: 10.1016/j.eururo.2019.02.009
7. Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst.* (2009) 101:374–83. doi: 10.1093/jnci/djp001
8. Loeb S, Bjurlin MA, Nicholson J, Tammela TL, Penson DF, Carter HB, et al. Overdiagnosis and overtreatment of prostate cancer. *Eur Urol.* (2014) 65:1046–55. doi: 10.1016/j.eururo.2013.12.062
9. Loeb S, Vellekoop A, Ahmed HU, Catto J, Emberton M, Nam R, et al. Systematic review of complications of prostate biopsy. *Eur Urol.* (2013) 64:876–92. doi: 10.1016/j.eururo.2013.05.049
10. Zhang Y, Zhang Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell Mol Immunol.* (2020) 17:807–21. doi: 10.1038/s41423-020-0488-6
11. Flammiger A, Weisbach L, Huland H, Tennstedt P, Simon R, Minner S, et al. High tissue density of FOXP3+ T cells is associated with clinical outcome in prostate cancer. *Eur J Cancer.* (2013) 49:1273–9. doi: 10.1016/j.ejca.2012.11.035
12. Zhao E, Wang L, Dai J, Kryczek I, Wei S, Vatan L, et al. Regulatory T cells in the bone marrow microenvironment in patients with prostate cancer. *Oncoimmunology.* (2012) 1:152–61. doi: 10.4161/onci.1.2.18480
13. Yang Y, Attwood K, Bshara W, Mohler JL, Guru K, Xu B, et al. High intratumoral CD8+ T-cell infiltration is associated with improved survival in prostate cancer patients undergoing radical prostatectomy. *Prostate.* (2021) 81:20–8. doi: 10.1002/pros.24068
14. Rizzo A, Mollica V, Cimadamore A, Santoni M, Scarpelli M, Giunchi F, et al. Is there a role for immunotherapy in prostate cancer? *Cells.* (2020) 9:2051. doi: 10.3390/cells902051

15. Erickson BJ. Basic artificial intelligence techniques: machine learning and deep learning. *Radiol Clin North Am.* (2021) 59:933–40. doi: 10.1016/j.rcl.2021.06.004
16. Zhang L, Xiong Y, Zhang J, Feng Y, Xu A. Systematic proteome-wide Mendelian randomization using the human plasma proteome to identify therapeutic targets for lung adenocarcinoma. *J Trans Med.* (2024) 22:330. doi: 10.1186/s12967-024-04919-z
17. Li Z, Zhang B, Liu Q, Tao Z, Ding L, Guo B, et al. Genetic association of lipids and lipid-lowering drug target genes with non-alcoholic fatty liver disease. *EBioMedicine.* (2023) 90:104543. doi: 10.1016/j.ebiom.2023.104543
18. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol.* (2016) 40:597–608. doi: 10.1002/gepi.21998
19. Lin Y, Zhao X, Miao Z, Ling Z, Wei X, Pu J, et al. Data-driven translational prostate cancer research: from biomarker discovery to clinical decision. *J Transl Med.* (2020) 18:119. doi: 10.1186/s12967-020-02281-4
20. S V, Z Z, P G, X L, A M, D L, et al. PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. *Proc Natl Acad Sci United States America.* (2000) 97. doi: 10.1073/pnas.97.22.12216
21. Wei JT, Feng Z, Partin AW, Brown E, Thompson I, Sokoll L, et al. Can urinary PCA3 supplement PSA in the early detection of prostate cancer? *J Clin Oncol.* (2014) 32:4066–72. doi: 10.1200/JCO.2013.52.8505
22. Thammavongsa V, Missiakas DM, Schneewind O. Staphylococcus aureus degrades neutrophil extracellular traps to promote immune cell death. *Science.* (2013) 342:863–6. doi: 10.1126/science.1242255
23. Pauli NT, Kim HK, Falugi F, Huang M, Dulac J, Henry Dunand C, et al. Staphylococcus aureus infection induces protein A-mediated immune evasion in humans. *J Exp Med.* (2014) 211:2331–9. doi: 10.1084/jem.20141404
24. Wu S-Q, Su H, Wang Y-H, Zhao X-K. Role of tumor-associated immune cells in prostate cancer: angel or devil? *Asian J Androl.* (2019) 21:433–7. doi: 10.4103/aja.aja_47_19
25. Waldhauer I, Steinle A. NK cells and cancer immunosurveillance. *Oncogene.* (2008) 27:5932–43. doi: 10.1038/onc.2008.267
26. Imai K, Matsuyama S, Miyake S, Suga K, Nakachi K. Natural cytotoxic activity of peripheral-blood lymphocytes and cancer incidence: an 11-year follow-up study of a general population. *Lancet.* (2000) 356:1795–9. doi: 10.1016/S0140-6736(00)03231-1
27. Pasero C, Gravis G, Granjeaud S, Guerin M, Thomassin-Piana J, Rocchi P, et al. Highly effective NK cells are associated with good prognosis in patients with metastatic prostate cancer. *Oncotarget.* (2015) 6:14360–73. doi: 10.18632/oncotarget.3965
28. Seki T, Shimizu Y, Ishii K, Takahama Y, Kato K, Yano T. NK cells can preferentially target prostate cancer stem-like cells via the TRAIL/DR5 signaling pathway. *Biomolecules.* (2021) 11:1702. doi: 10.3390/biom11111702
29. Kainulainen K, Niskanen EA, Kinnunen J, Mäki-Mantila K, Hartikainen K, Paakinaho V, et al. Secreted factors from M1 macrophages drive prostate cancer stem cell plasticity by upregulating NANOG, SOX2, and CD44 through NFκB-signaling. *Oncimmunology.* (2024) 13:2393442. doi: 10.1080/2162402X.2024.2393442
30. Hadimani SM, Das S, Harish KG. An immunohistochemical evaluation of tumor-associated macrophages (M1 and M2) in carcinoma prostate - An institutional study. *J Cancer Res Ther.* (2023) 19:S300–5. doi: 10.4103/jcrt.jcrt_497_22
31. Li J, Sun Y, Yan R, Wu X, Zou H, Meng Y. Urea transporter B downregulates polyamines levels in melanoma B16 cells via p53 activation. *Biochim Biophys Acta Mol Cell Res.* (2022) 1869:119236. doi: 10.1016/j.bbamcr.2022.119236
32. Grupp K, Höhne TS, Prien K, Hube-Magg C, Tsourlakis MC, Sirma H, et al. Reduced CD147 expression is linked to ERG fusion-positive prostate cancers but lacks substantial impact on PSA recurrence in patients treated by radical prostatectomy. *Exp Mol Pathol.* (2013) 95:227–34. doi: 10.1016/j.yexmp.2013.08.002
33. Ye B, Ding K, Li K, Zhu Q. Study on the role of SLC14A1 gene in biochemical recurrence of prostate cancer. *Sci Rep.* (2022) 12:17064. doi: 10.1038/s41598-022-20775-7
34. Hou R, Kong X, Yang B, Xie Y, Chen G. SLC14A1: a novel target for human urothelial cancer. *Clin Transl Oncol.* (2017) 19:1438–46. doi: 10.1007/s12094-017-1693-3
35. Sun Z, Mao Y, Zhang X, Lu S, Wang H, Zhang C, et al. Identification of ARHGEF38, NETO2, GOLM1, and SAPCD2 associated with prostate cancer progression by bioinformatic analysis and experimental validation. *Front Cell Dev Biol.* (2021) 9:718638. doi: 10.3389/fcell.2021.718638
36. Theunissen F, Anderton RS, Mastaglia FL, James I, Bedlack R, Akkari PA. Intronic NEFH variant is associated with reduced risk for sporadic ALS and later age of disease onset. *Sci Rep.* (2022) 12:14739. doi: 10.1038/s41598-022-18942-x
37. Whitaker HC, Warren AY, Eeles R, Kote-Jarai Z, Neal DE. The potential value of microseminoprotein-beta as a prostate cancer biomarker and therapeutic target. *Prostate.* (2010) 70:333–40. doi: 10.1002/pros.21059
38. Zhang JS, Wang L, Huang H, Nelson M, Smith DI. Keratin 23 (K23), a novel acidic keratin, is highly induced by histone deacetylase inhibitors during differentiation of pancreatic cancer cells. *Genes Chromosomes Cancer.* (2001) 30:123–35. doi: 10.1002/1098-2264(2000)9999:9999::AID-GCC1070>3.0.CO;2-W
39. Kim D, Brocker CN, Takahashi S, Yagai T, Kim T, Xie G, et al. Keratin 23 is a peroxisome proliferator-activated receptor alpha-dependent, MYC-amplified oncogene that promotes hepatocyte proliferation. *Hepatology.* (2019) 70:154–67. doi: 10.1002/hep.30530
40. Birkenkamp-Demtröder K, Hahn SA, Mansilla F, Thorsen K, Maghnooj A, Christensen R, et al. Keratin23 (KRT23) knockdown decreases proliferation and affects the DNA damage response of colon cancer cells. *PLoS One.* (2013) 8:e73593. doi: 10.1371/journal.pone.0073593
41. Wang W, Yuan D, Jiang K, Li R, Qu H, Jiang F-N, et al. Genome-wide CRISPR-cas9 screening and identification of potential genes promoting prostate cancer growth and metastasis. *Curr Cancer Drug Targets.* (2022) 23:71–86. doi: 10.2174/1568009622666220615154137
42. Xiao X, Qing L, Li Z, Ye F, Dong Y, Mi J, et al. Identification and validation of diagnostic and prognostic biomarkers in prostate cancer based on WGCNA. *Discovery Oncol.* (2024) 15:131. doi: 10.1007/s12672-024-00983-5