Discover Oncology

Analysis

Exploring the interaction between immune cells in the prostate cancer microenvironment combining weighted correlation gene network analysis and single-cell sequencing: An integrated bioinformatics analysis

Danial Hashemi Karoii[1] · Sobhan Bavandi[2] · Melika Djamali[3] · Ali Shakeri Abroudi[4]

Received: 29 August 2024 / Accepted: 25 September 2024
Published online: 30 September 2024

Abstract

Background The rise of treatment resistance and variability across malignant profiles has made precision oncology an imperative in today's medical landscape. Prostate cancer is a prevalent form of cancer in males, characterized by significant diversity in both genomic and clinical characteristics. The tumor microenvironment consists of stroma, tumor cells, and various immune cells. The stromal components and tumor cells engage in mutual communication and facilitate the development of a low-oxygen and pro-cancer milieu by producing cytokines and activating pro-inflammatory signaling pathways.

Methods In order to discover new genes associated with tumor cells that interact and facilitate a hypoxic environment in prostate cancer, we conducted a cutting-edge bioinformatics investigation. This included analyzing high-throughput genomic datasets obtained from the cancer genome atlas (TCGA).

Results A combination of weighted gene co-expression network analysis and single-cell sequencing has identified nine dysregulated immune hub genes (AMACR, KCNN3, MME, EGFR, FLT1, GDF15, KDR, IGF1, and KRT7) that are believed to have significant involvement in the biological pathways involved with the advancement of prostate cancer enviriment. In the prostate cancer environment, we observed the overexpression of GDF15 and KRT7 genes, as well as the down-regulation of other genes. Additionally, the cBioPortal platform was used to investigate the frequency of alterations in the genes and their effects on the survival of the patients. The Kaplan-Meier survival analysis indicated that the changes in the candidate genes were associated with a reduction in the overall survival of the patients.

Conclusions In summary, the findings indicate that studying the genes and their genomic changes may be used to develop precise treatments for prostate cancer. This approach involves early detection and targeted therapy.

Keywords Bioinformatics · Hub genes · Prostate cancer · Immune cells

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-024-01399-x.

Danial Hashemi Karoii, Sobhan Bavandi, Melika Djamali and Ali Shakeri Abroudi equal first authors.

✉ Danial Hashemi Karoii, d.hashemi.karoii@ut.ac.ir; danialhashemiac@gmail.com; Sobhan Bavandi, sobhanba.sb@gmail.com; Melika Djamali, mdjamali@ut.ac.ir; Ali Shakeri Abroudi, Alishakeriabroudi@gmail.com | [1]Department of Cell and Molecular Biology, School of Biology, College of Science, University of Tehran, Tehran, Iran. [2]Department of Biology, Qaemshahr Branch, Islamic Azad University, Qaemshahr, Iran. [3]Department of Biology, Faculty of Science, Tehran University, Tehran, Iran. [4]Department of Cellular and Molecular Biology, Faculty of Advanced Science and Technology, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran.

Abbreviations

PC	Prostate cancer
BP	biological process
CC	cellular components
DAVID	Annotation visualization and integrated discovery
DEGs	Differentially expressed genes
GEO	Gene expression omnibus
GO	Gene 0ntology
KEGG	Kyoto encyclopedia of gene and genome
MCODE	Molecular complex detection
WGCNA	Weighted correlation network analysis
MF	molecular function
PPI	protein-protein interaction
STRING	Search tool for the retrieval of interacting genes
TCGA	The Cancer genome atlas
TAMs	Tumor-associated macrophages

1 Introduction

A recent demographic research has shown that the growing elderly population, resulting from demographic and socio-economic changes, is leading to a significant rise in new cancer cases [1]. Among these instances, prostate cancer (PCa) continues to be the primary cause of cancer-related deaths in males globally. Due to its higher prevalence in older males, the care of cancer in this population might be complicated by the presence of other medical conditions and age-related differences, which can have substantial social and economic consequences [2]. Given the significant influence that the burden of prostate cancer has on both families and health systems, there is a pressing need to enhance cancer detection and therapy in order to guarantee proper disease management [3].

During the first phases of PCa progression, the proliferation of tumor cells relies on the presence of testosterone in the bloodstream. This is why androgen deprivation therapy (ADT) is used as a treatment for localized PCa [4]. Genetic and epigenetic changes lead to the development of highly aggressive PCa clones that eventually progress into the metastatic stage (mPCa), which occurs before the development of castration resistance [5]. The presence of clonal heterogeneity and gene instability in prostate cancer, exacerbated by the effects of antineoplastic treatment, may contribute to the lack of efficacy of traditional androgen deprivation therapy (ADT). The first phases of "castration-resistant" prostate cancer (CRPC) are marked by diverse populations of PCa cells formed due to the influence of anti-androgens [6]. The presence of stress circumstances enables the cells to adapt and reprogram themselves to survive with low levels of testosterone in the bloodstream and tissues. This syndrome results in the advancement of tumors and the rapid growth of cancerous cells, either by reactivating the androgen receptor (AR) pathway or by activating pathways that are independent of the androgen receptor [6]. Therefore, it is feasible to see significant variation in physical traits and characteristics within a single patient, which is characterised by the simultaneous existence of cell populations with distinct evolutionary backgrounds and varying sensitivity to drugs [7]. It has been suggested that this variety may come from non-mutational processes that cause an increase in populations with the same genes but varying sensitivity to ADT [8]. Recent data suggest that resistance to ADT is caused by the interaction of redundant genetic and epigenetic pathways, which engage in intricate communication with cellular plasticity to enable adaptation to long-term drug exposure [9].

Despite the advancements in multimodal methods and treatment resources, there is a lack of accurate prognostic and predictive criteria for choosing appropriate therapy for high-risk patients with metastatic castration-resistant prostate cancer (mCRPC) [10]. The absence of indicators capable of directing doctors towards the most suitable therapeutic option has resulted in little prognosis enhancement after the treatment of mPCa and mCRPC. The lack of standards for selecting target molecules based on clinical data, as described by the ESMO Scale for Clinical Action of Molecular Targets—ESCAT, is responsible for the insufficient precision medicine in many cancer therapies [11]. Furthermore, although several studies have emphasized the genetic changes linked to the progression of PCa, there is also a growing recognition of the crucial role played by epigenetic regulation in governing the ability of cancer cells to adapt and become resistant to androgen treatment. To enhance clinical procedures, further research is needed to discover certain critical genes and pathways that are linked to various populations of PCa and CRPC cells [11, 12].

Table 1 We are studying the gene expression patterns (GEO) related to PCa using mRNA, lncRNA, and miRNA

Sample /accession no.	Prostate Cancer	Benign prostate hyperplasia	Sample type	Platform	Included/ excluded	RNA type
GSE55945	13	8	Radical prostatectomy tissue	Affymetrix GPL570	Included	mRNA
GSE104749	4	4	Fine-Needle Aspiration tissue	Affymetrix GPL570	Included	mRNA
GSE46602	36	14	Laser micro-dissected tissue	Affymetrix GPL570	Included	mRNA
GSE32571	59	39	Fresh frozen tissue	Illumina GPL6947	Included	mRNA
GSE142288	48	-	Tissue	Agilent GPL13264	Excluded	mRNA
GSE155792	1	-	Tissue	Agilent GPL28148	Excluded	mRNA
GSE113153	10	-	Tissue	GPL21825	Excluded	mRNA
GSE134160	16	-	Fresh frozen tissue	Agilent GPL26898	Excluded	mRNA
GSE141806	3	-	Tissue	GPL16791	Included	mRNA
GSE123565	5	2	Tissue	GPL16791 Illumina HiSeq 2500	Included	GSE123565
GSE94577	12	5	Tissue	GPL10999 Illumina Genome Analyzer IIx	Included	mRNA/Single cell
GSE82225	13	6	Tissue	GPL11154 Illumina HiSeq 2000	Included	mRNA/Single cell
GSE176031	20	10	Tissue	GPL11154 Illumina HiSeq 2000	Included	mRNA/Single cell
GSE34933	6	6	Tissue	Agilent-014850 Whole Human Genome Microarray 4 × 44 K G4112F	Included	miRNA
GSE21036	50	20	Tissue	GPL8227 Agilent-019118 Human miRNA Microarray 2.0 G4470B	Included	miRNA
GSE48430	10	10	Tissue	GPL17391 Applied Biosystems Human MicroRNA Array A v2.0	Included	miRNA
GSE212215	5	5	Tissue	GPL20301 Illumina HiSeq 4000	Included	ciRNA

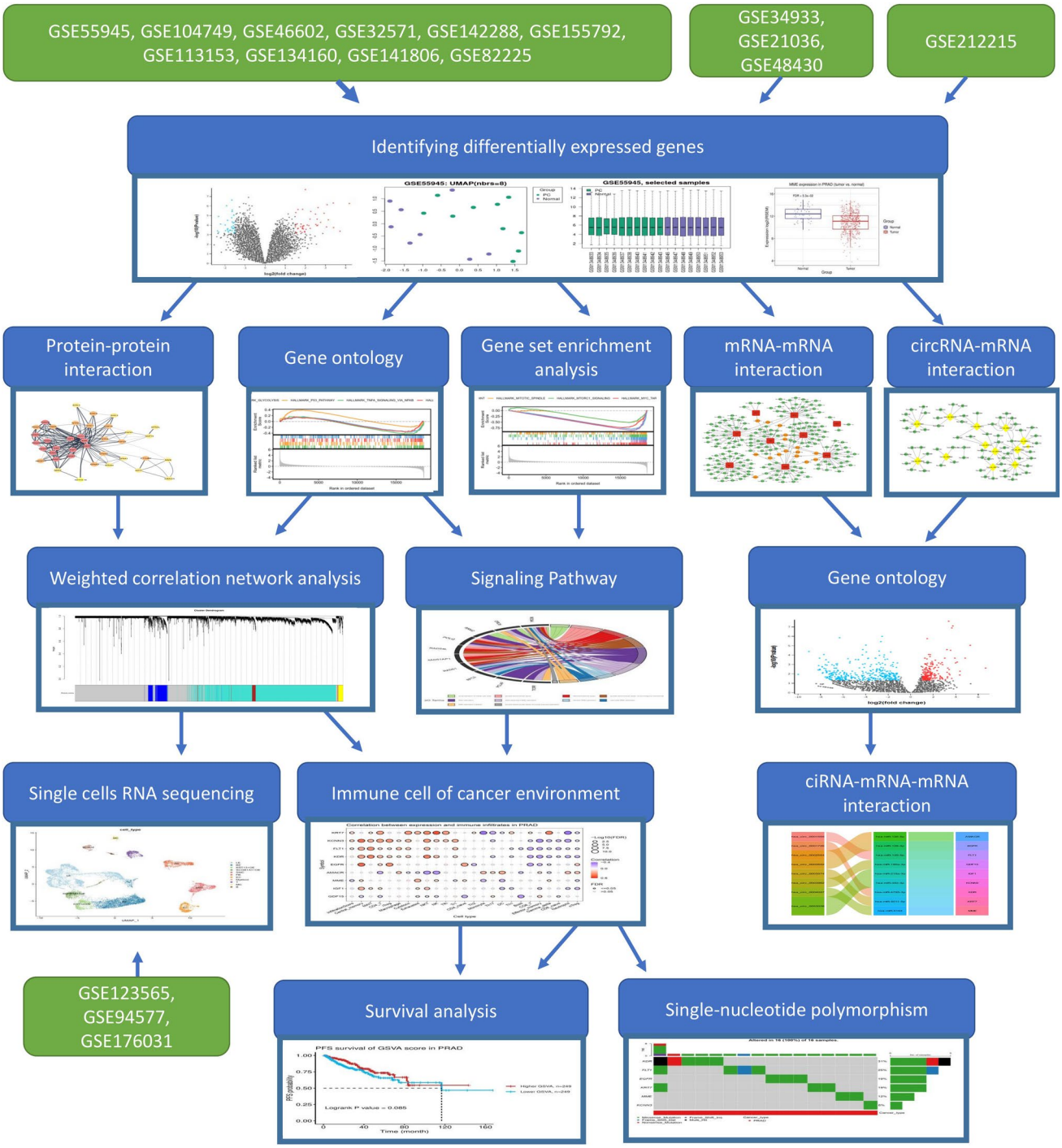


Fig. 1 Illustration depicting the design of the research. The datasets we considered in our analysis satisfied specified criteria: (1) they used human prostate tissue samples, (2) they included both case and control groups, and (3) they consisted of at least 10 samples each. The use of a substantial sample size was essential in order to accurately detect differentially expressed genes (DEGs) or non-coding RNAs since the use of small sample sizes is known to present difficulties in microarray research

The link between the immune system and the microenvironment of prostate cancer is marked by a dynamic interplay, in which immune cells may have both inhibitory and promotive effects on the growth of cancer. Immune surveillance systems have the ability to initially identify and remove cancer cells. However, prostate cancer often develops strategies

to avoid these defenses, such as reducing the expression of antigens and releasing chemicals that inhibit the immune system [12, 13]. The tumour microenvironment (TME) undergoes immunosuppression, facilitated by the presence of regulatory T cells and tumor-associated macrophages (TAMs), which promote tumor development and spread. Therapeutic approaches seek to address these pathways by using immune checkpoint inhibitors, cancer vaccines, and adoptive cell treatments to enhance the immune response against prostate cancer. Gaining comprehension and skillfully controlling these interactions are essential for progressing the therapy of prostate cancer [14, 15].

Studies have shown that the advancement of prostate cancer is affected by intricate interactions with the immune system. Research has shown that prostate cancer cells have the ability to avoid being detected by the immune system by reducing the expression of antigens and increasing the presence of immune checkpoint proteins such as PD-L1 [16, 17]. This indicates that checkpoint inhibitors might be effective in treating the disease. Tumor-associated macrophages (TAMs) and chronic inflammation in the prostate contribute to the formation of tumors. TAMs facilitate the evolution of cancer by secreting cytokines [18]. Increased infiltration of CD8 + T cells has been associated with a more favorable prognosis, highlighting the need of immunological profile. Therapeutic modalities, such as cancer vaccines and adoptive cell treatments, seek to augment the immune response against prostate cancer [19]. Moreover, the microbiome of the prostate might potentially impact the immune system in that area, suggesting the need for more investigation into how microbes can influence the development of prostate cancer.

Multiple crucial genes have an effect on the immunological interactions occurring inside the microenvironment of prostate cancer. PD-L1 and CTLA-4 play crucial roles in immune checkpoint pathways that cancer cells use to avoid immune responses [20]. FOXP3 is linked to regulatory T cells, which inhibit immunological function. CSF1R is related with tumor-associated macrophages, which facilitate tumor proliferation. IL-6 and IL-8 are cytokines that stimulate inflammation and facilitate the development of tumors [21, 22]. The presence of ARG1 in myeloid-derived suppressor cells has a role in inhibiting the immune system. TGF-β is a cytokine that promotes the ability of tumors to avoid detection by the immune system and facilitates their proliferation [23]. Cancer cells often suppress the expression of MHC class I molecules in order to evade detection by T lymphocytes, which are responsible for recognising and responding to antigens. IDO1 enzymatically breaks down tryptophan to suppress T cell activity [24]. Targeting these specific genes may facilitate the development of effective medicines for prostate cancer.

Nevertheless, the majority of study conducted so far has mostly concentrated on examining individual molecules, with very little attention given to the interactions that occur between molecules [25]. The current research specifically examines the genetic makeup of prostate cancer and aims to provide a comprehensive understanding of how genes interact in individuals with high and low grade Gleason scores, utilizing microarray data [26]. Microarray data often undergo a two-step filtering process to identify genes that are differentially expressed. This involves applying strict criteria such as fold change or p-value limits. Nevertheless, when using these approaches, genes located downstream are given higher rankings compared to genes located upstream. Consequently, genes that are really responsible for causing the illness may go unnoticed and remain unidentified [27]. Therefore, we used weighted gene coexpression network analysis (WGCNA) to maintain connections between genes and emphasize key genes in high-risk prostate cancer. WGCNA has been effectively used to develop biomarkers for castration-resistant prostate cancer and lncRNA markers for prostate cancer. The current work has discovered hub genes that potentially function as both prognostic indicators and possible therapy targets in high-risk prostate cancer.

2 Materials and methods

2.1 Prostate cancer datasets

The current study relied on the amalgamation of many published datasets to get a unified list and a reliable source for confirming the involvement of genes in the progression of PCa and the development of hormone-refractory prostate cancer (HRPC). The datasets from research that have supplied gene lists related to PC have been examined and combined with our results. The summary of these datasets can be seen in Table 1; Fig. 1. The tissue-specific expression was confirmed separately in the research obtained from cBioPortal. (i) The PCa-gene set consists of data from two sources: Prostate Adenocarcinoma (PRAD-TCGA PanCancer Atlas) and Prostate Adenocarcinoma (MSK, available at https://datacatalog.mskcc.org/dataset/10426).

2.2 Differential expression analysis

The GEO2R tool from the GEO site was used to conduct differential expression analysis of the GEO datasets. This study relied on the Limma R package [28]. The following comparisons were made: The GSE3325 [29], GSE3933 [30], and GSE68882 [31] datasets were compared between metastatic and primary prostate cancer groups. The GSE32269 [32], GSE6811 [33], GSE70770 [34], GSE6752 [35], and GSE35988 [36] datasets were compared between mCRPC and prostate cancer groups. The GSE101607 dataset was used to compare AR-driven and non-AR-driven groups. Lastly, the GSE77930 [37] dataset was used to compare CHGA negative and CHGA positive/SYP positive/SR negative groups. Significance was attributed to genes with an absolute log2fold-change more than or equal to 1.5 and a Benjamini-Hochberg adjusted p-value less than or equal to 0.05.

2.3 Identification of degs

DEGs were identified by screening BC and normal prostate tissue samples using the GEO2R web programme. DEGs meeting the requirements of having an absolute log fold change of more than 1.5 and an adjusted p-value less than 0.05 were evaluated as standard. The profiles of DEGs were downloaded and compared using Venn diagram software. Genes having a log fold change (FC) less than 1.5 were considered down-regulated, and vice versa.

2.4 Data preprocessing and analysis

The microarray gene expression data was analysed using R/Bioconductor, available at http://www.R-project.org/. The raw data in .CEL format was processed and standardised using the Bioconductor-Affy programme for probe data normalisation, which also performed background correction to provide expression values. The limma algorithm was then used to assess the statistical significance of the differentially expressed genes between normal and COPD samples using the t-test statistical method. The Benjamini-Hochberg method was used to calculate the false discovery rate (FDR) for all genes that showed statistical significance, enabling the identification and removal of false positive genes. The criteria for identifying differentially expressed genes (DEGs) were set as a false discovery rate (FDR) below 0.01 and an absolute log2 fold change (FC) over 1.5. A p-value less than 0.05 was considered to have statistical significance. The expression levels of differentially expressed genes (DEGs) were classified into two categories: up-regulated genes and down-regulated genes. The categorizations were then represented using the Heatmap online web tool found at http://www.heatmapper.ca.

2.5 Sorting and comparing groups of proteins

The comparison of differentially expressed genes across the three study groups was conducted using the web software ArrayMining. The gene list was analysed using PANTHER, a method for gene ontology analysis (http://www.pantherdb.org/).

2.6 Gene ontology and functional enrichment analyses

Enrichment analysis of DEGs was performed using the Gene Ontology (GO) and KEGG pathway databases. The STRING database was used for this research. The important GO keywords and pathways were se

2.7 Construction of protein-protein interaction (PPI) map

lected based on a criterion of adjusted p-value less than 0.05 and a FDR of 0.05. The GO annotation networks were visualized using the Cytoscape network style plugin, which may be found at http://www.cytoscape.org/.

2.8 Construction of protein-protein interaction (PPI) map

The putative PPI networks derived from the DEGs in the lung and blood were created using Bisogenet, a Cytoscape plugin (version 3.4.0). Moreover, the network clusters resulting from PPI were determined using a network analyzer

programmed. The threshold for input nodes and their neighbours was set to a maximum distance of 1 edge. During the development of PPIM, only interactions between proteins were chosen, whereas interactions between proteins and DNA, as well as interactions involved in microRNA silencing, were excluded. Our recent studies shows that each node in the network represents a gene, while the edges between the nodes reflect physical or functional connections. Consequently, a few nodes exhibit a high degree of connectedness, while many nodes have a low level of connectivity [38–50].

2.9 Integration of the PPI network

The identified DEGs were analyzed using the online Search Tool for the Retrieval of Interacting Genes database to assess their interacting associations. Interactions that had a cumulative score greater than 0.4 were considered to be statistically significant. The interconnected regulatory networks were visualised using Cytoscape software (version 3.5.1). The Cytoscape plugin Molecular Complex Detection (MCODE; version 1.31) [51] was used to identify more intricate interconnected areas inside the PPI network. The hub genes in the Cytoscape plugin cytoHubba (version 0.1) were determined based on the top five rated genes according to their degree level

2.10 Hub gene subnetwork construction

PPI is classified as a network of significant magnitude. By using the principles of network biology, the PPI complex was broken down into several subnetwork clusters known as the Significant Protein Interaction Network (SPIN). Several genes were retrieved based on the characteristics of degree centrality (DC) and betweenness centrality (BC). Every protein that was included in the network was integrated and standardized into Cytoscape 3.2.1. The Network Analyzer tool was then used to determine the local DC and global BC characteristics of the network.

2.11 Further authentication of hub genes using other open databases

In order to confir

2.12 Evaluating the effectiveness of prognostic signature

m the importance of four hub genes, mRNA-seq data for PCa was obtained from The Cancer Genome Atlas (TCGA) database. The Prostate Cancer Gene-Expression Miner v4.7 was used to visualize the expression of these four genes in healthy, BC-adjacent, and BC samples.

2.13 Calculation and classification of risk score

The risk score for each sample was computed using the prognostic model, and the prognostic signature was assessed using a receiver operating characteristic (ROC) curve. The R package's timeROC function was used to evaluate the receiver operating characteristic (ROC) and compute the area under the ROC curve (AUC) as a measure of the prognostic signature's efficacy. A z-score of 0 is the threshold for classifying samples into the low-risk and high-risk groups.

2.14 Evaluating the effectiveness of prognostic signature

The consistency of the test dataset's signature was assessed by comparing its performance with that of the training dataset. The independent datasets GSE141806 [52] and GSE123565 [53] were selected as the validation datasets for further validation. The test dataset and validation dataset were used to analyse the correlation between the prognostic signature and clinical information, such as age, stage (I, II, III, and IV), pathological stage (T, N, and M stages), and status (PR status, ER status, and HER2 status). This analysis was conducted using univariate and multivariable Cox regressions, ROC analysis, and Kaplan-Meier survival curves.

Fig. 2 This study demonstrates the diversity in gene expression data between persons with PCa and healthy individuals. The analysis ▶ includes five datasets: **A** GSE55945, **B** GSE29079, **C** GSE104749, **D** GSE46602, and (**E**) GSE32571. In this depiction, genes that exhibit an increase in expression, with a fold change more than 1.5 and a corrected P-value less than 0.05, are shown as red dots. Conversely, genes that show a decrease in expression and match the same criteria are represented as blue points. Genes exhibiting no significant variation in expression are indicated by black data points

2.15 Analyzing the correlation between risk score and functional pathways

An investigation was conducted to examine the link between the risk score and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways. The gene expression profile of each sample was analysed using single-sample gene set enrichment analysis (ssGSEA) in the R package. The ssGSEA score was generated for distinct functional pathways in each sample to evaluate the association between KEGG pathways and the risk score. KEGG pathways were considered to be correlated with the risk score when the ssGSEA score exceeded 0.25.

2.16 Statistical analysis

The association between RBP expression and overall survival was evaluated using both Cox regression analysis and Kaplan-Meier estimate technique, using the "survival" package in R. In the Cox regression analysis, the RBP was assessed as a continuous variable, while age and gender were included as extra covariates. The high-expression group and low-expression group for the Kaplan-Meier estimates were determined by utilizing the median RBP expression value as a threshold. The log-rank test was used to evaluate a notable difference in the survival curves of the two groups. A significance level of $P < 0.05$ was used to determine statistical significance.

2.17 Validation of the target genes in TCGA

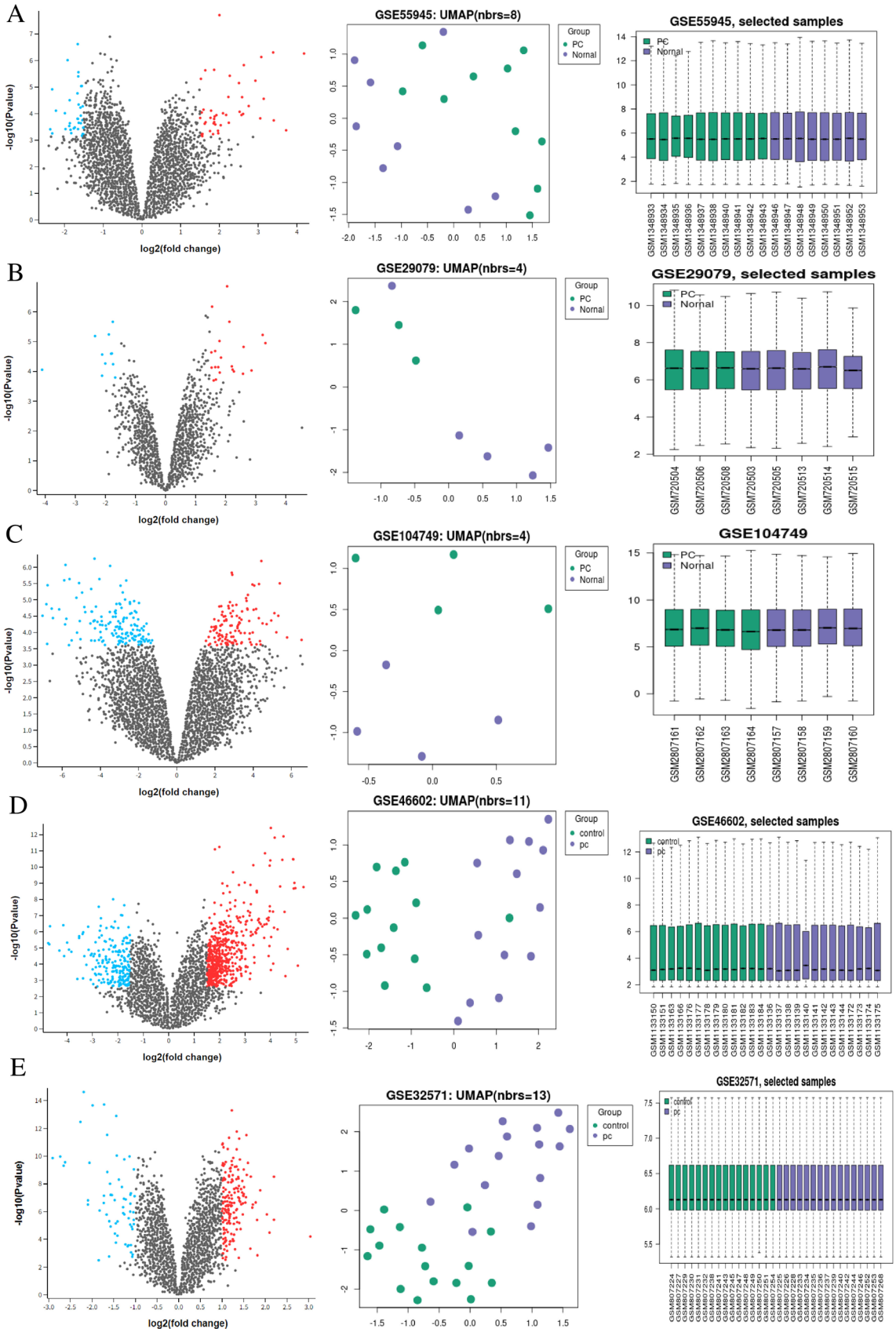
In order to verify the methylation and expression levels of the genes of interest, a dataset was obtained from TCGA (https://www.cancer.gov/ccg/research/genome-sequencing/tcga). The TCGA database contains detailed and extensive information on important genetic alterations in different forms of cancer. In addition, the translational levels of the hub genes were confirmed using the Human Protein Atlas (HPA) database. The cBio Cancer Genomics Portal, a publicly accessible platform for analyzing extensive cancer genomics datasets across different types of cancer, was used to investigate the genetic changes associated with the central genes and to assess the relationships between mRNA expression and DNA methylation in PCa.

2.18 Immune infiltration in pca

CIBERSORT used gene expression data to determine the relative percentage and abundance of several immune cell types in both tumor and normal groups. In this particular case, we used CIBERSORT to evaluate the proportions of 22 different kinds of immune cells in PCa samples. Only occurrences with CIBERSORT [54] outcomes that had a p-value of less than or equal to 0.05 were included for further analysis. The Wilcoxon rank-sum test was used to differentiate crucial immune cells in tumor samples from non-malignant ones. Subsequently, K-M survival analysis was used to examine the potential correlation between the quantity of specific immune cells and the overall survival of patients with PCa. A Cox proportional hazard model was constructed by including several immune cells, using LASSO regression analysis. Subsequently, a nomogram was developed to predict the prognosis of PCa. The bias and accuracy of the nomogram were assessed by calculating the concordance index of the Cox model.

2.19 Association between selected rnas and immune cells

Univariate Cox regression and Kaplan-Meier survival analysis were used to investigate the immune cells associated with survival. The final immune cell model was created using multivariate Cox regression analysis and Lasso regression, which were performed simultaneously. A nomogram was used to visually display an equation that predicts medical outcomes. Nomograms provide points to patients depending on the severity of their risk factors. The disparity in


A
-log10(Pvalue)
log2(fold change)
GSE55945: UMAP(nbrs=8)
Group
PC
Normal
GSE55945, selected samples
PC
Normal
GSM1348933
GSM1348934
GSM1348935
GSM1348936
GSM1348937
GSM1348938
GSM1348940
GSM1348941
GSM1348942
GSM1348943
GSM1348946
GSM1348947
GSM1348948
GSM1348949
GSM1348950
GSM1348951
GSM1348952
GSM1348953
B
-log10(Pvalue)
log2(fold change)
GSE29079: UMAP(nbrs=4)
Group
PC
Normal
GSE29079, selected samples
PC
Normal
GSM720504
GSM720506
GSM720508
GSM720503
GSM720505
GSM720513
GSM720514
GSM720515
C
-log10(Pvalue)
log2(fold change)
GSE104749: UMAP(nbrs=4)
Group
PC
Normal
GSE104749
PC
Normal
GSM2807161
GSM2807162
GSM2807163
GSM2807164
GSM2807157
GSM2807158
GSM2807159
GSM2807160
D
-log10(Pvalue)
log2(fold change)
GSE46602: UMAP(nbrs=11)
Group
control
pc
GSE46602, selected samples
control
pc
E
-log10(Pvalue)
log2(fold change)
GSE32571: UMAP(nbrs=13)
Group
control
pc
GSE32571, selected samples
control
pc

gene expression between tumor tissues and normal tissues was used as the foundation for the risk factors in the two nomograms that we developed for the research. Next, we predict the survival rate of patients for 1, 3, or 5 years.

2.20 WGCNA

A scale-free co-expression network was constructed using the R package WGCNA [55] based on RNA-Seq data. The Pearson's correlation coefficient values were used to illustrate the relationships between pairs of genes in this network. Subsequently, the matrix of network topological overlap measure (TOM) was calculated to quantify the relationship between pairs of genes. The gene co-expression modules were detected using hierarchical clustering with average linkage, which was based on the overlap of network architecture. The choice of subtype representative modules was determined by examining the relationship between the modules and external clinical characteristics (eigengene-sample subtype correlation) as well as the connection between gene significance (GS) and module membership (MM) values. The identification of subtype representative modules was determined by a significant correlation, with a P-value of less than 0.01, between GS and MM values, as well as a correlation of more than 0.75 between the module eigengene (ME) and the external trait. The hub genes in a certain module were determined by selecting those with a GS value above 0.2 and a module membership (MM) value greater than 0.8.

2.21 Data collection and preprocessing of single cells

The transcriptome data and clinical data of PCa patients who had neoadjuvant chemotherapy (NAC) were obtained from the Gene Expression Omnibus (GEO) database. The data was collected using the IDs GSE94577, GSE82225, and GSE176031, which can be found at https://www.ncbi.nlm.nih.gov/geo/. The discovery cohort included the most extensive collection of BC samples, totaling 306. The independent validation cohorts consisted of 15 BC samples from GSE94577 and 20 BC samples from GSE82225. Specimens without complete survival data were excluded from the research. The scRNA-seq data from 14 BC samples was obtained from the study done by Qian et al. The data was obtained from the website of lambrechtslab - Laboratory of Translational Genetics (vib.be). The scRNA-seq data was preprocessed using the Seurat [56] R tool (version 4.0.0). Cell samples that had an expression of more than 200 genes and a mitochondrial gene expression rate of less than 5% were kept. We used the "NormalizedData" tool to normalize the scRNA-seq dataset and the "FindVariableFeatures" method to identify 2000 genes with significant variability. We used the R package Harmony to mitigate the batch impacts. Following the normalization of the data, the principal component analysis (PCA) was conducted to group and visualize cells using uniform manifold approximation and projection (UMAP). Subsequently, the "DotPlot" tool was used to visually represent the expression level of marker genes inside a specific cluster. The assignment of these clusters to recognized cell lineages is determined by marker genes. The K-nearest neighbour (KNN) method and the "FindClusters" function with a resolution of 0.2 are used to detect clusters of cells.

2.22 Identification of candidate pca marker genes and their cell expression activity

In the discovery cohort (GSE176031 [57]) consisting of bulk transcriptome data, the R package limma was used to identify the DEGs across the NAC resistant and sensitive groups. The DEGs were determined based on a criteria of $|\log2FC| > 0.585$ and $FDR < 0.05$. The identification of cell-specific genes in single-cell data was performed using the "FindAllMarkers" function in the R package Seurat. Next, we identified the overlapping genes between NAC-associated differentially expressed genes (DEGs) and cell-specific genes, which we consider as the potential NAC marker genes.

We used the R package AUCell [58] (Version 1.12.0) to compute the expression score of potential NAC marker genes in every individual cell. The AUC value quantifies the fraction of genes with high expression levels in the gene set for each individual cell, and determines a ranking of gene expression for each cell. The "AUCell_explore Thresholds" technique was used to compute the threshold.

2.23 Modular analysis for prostate cancer progression

Three crucial phases were undertaken to perform modular analysis for the examination of disease progression. The gene expression data gained from comparing various disease stages of prostate cancer was integrated with the reconstructed

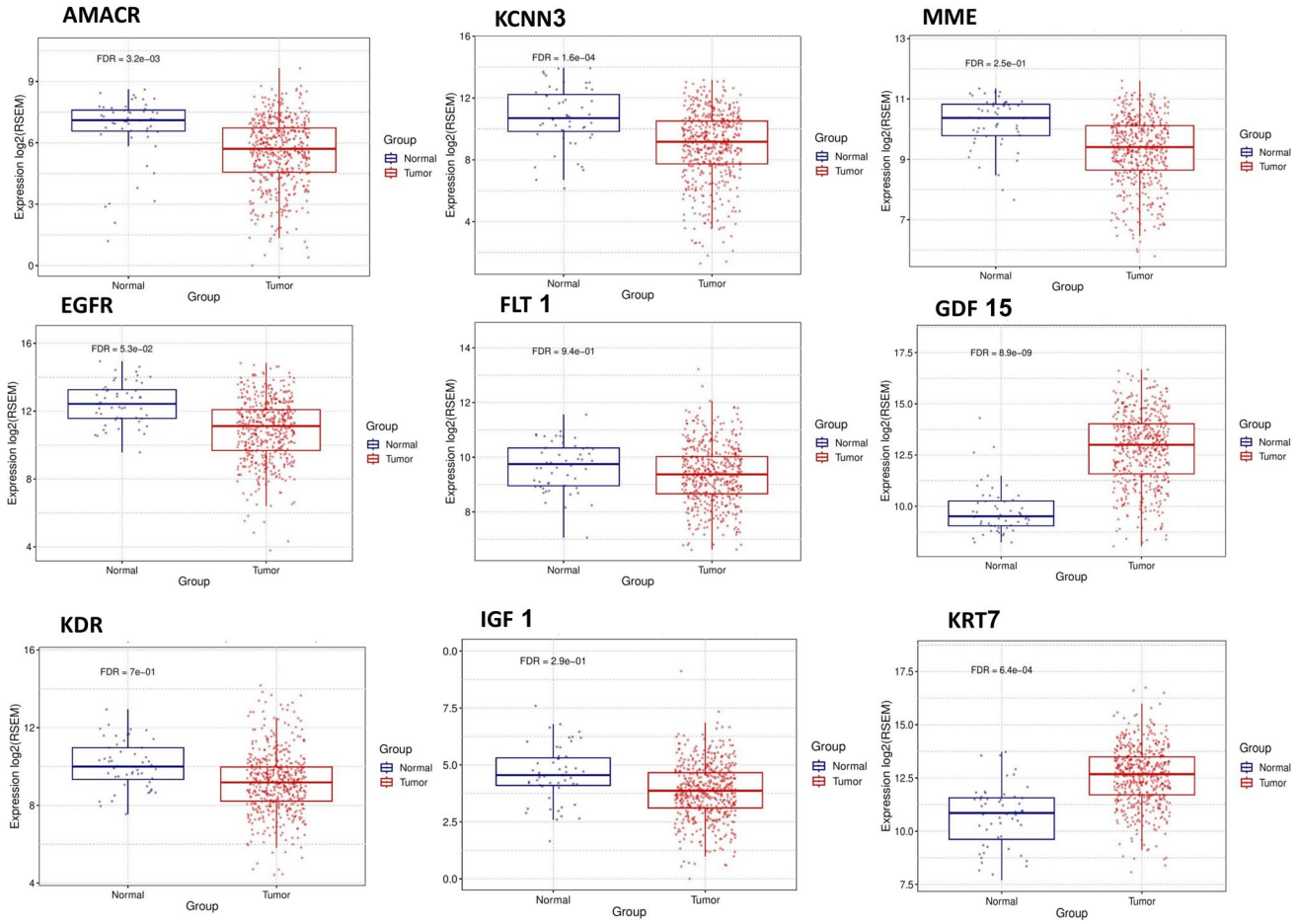


Fig. 3 Identify immune cells present in the microenvironment of prostate cancer and determine the levels of mRNA expression for the genes AMACR, KCNN3, MME, EGFR, FLT1, GDF15, KDR, IGF1, and KRT7 in tissue samples of prostate cancer using the GSE database

PCa at the beginning stage. Following that, the examination of core sub-networks and the investigation of overlapping were carried out as the second and third procedures, respectively. Regarding the analysis of the central sub-networks, we focused on those that had notably high activity scores and prominent ranks utilising the greedy approach. The study focused on using the greedy algorithm to locate key sub-networks within a vast network of connections. These sub-networks are interconnected sub-graphs of the interactome that exhibit a significant correlation between differential expression levels when comparing paired illness stages. To clarify the use of the greedy technique, it is crucial to mention that the adjusted p-value obtained from comparing paired illness stages was converted into a z-score using the inverse normal cumulative distribution (θ-1). The z-score was subsequently used for the purposes of scoring and ranking. Afterwards, the jActiveModules (jAM) plug-in, which employs a greedy algorithm, was used in Cytoscape to analyse and extract the statistically significant core sub-networks. The study was conducted with a threshold of three iterations and by taking into account the top ten ranks. After the completion, the top ten rankings were merged to form a definitive core sub-network. This sub-network depicts the juxtaposition of illness phases and gene expression patterns in a matched manner. More precisely, the choice of using jAM as the basis for this study was based on its widespread use and the support it has received in previous research.

To accomplish the modular analysis, the analysis of overlaps was performed. The gene-level overlapping analysis was used to illustrate the number of enriched genes that are common to all gene expression patterns. The investigation was performed via core sub-networks analysis. The proportion of overlapping genes was measured between any two final core sub-networks generated from pairwise disease phase comparison analysis across all gene expression profiles. Regarding the examination of overlapping, we have determined the number of genes using a specified formula.

Table 2 AMACR, KCNN3, MME, EGFR, FLT1, GDF15, KDR, IGF1, and KRT expression in PCa

Genes	Function	Log2FC	*p*-value
AMACR	alpha-methylacyl-CoA racemase (AMACR) enzyme interconverts pristanoyl-CoA and C27-bile acylCoAs between their (R)- and (S)-stereoisomers.	−3.5	0.003
KCNN3	potassium calcium-activated channel subfamily N member 3 (KCNN3) regulates neuronal excitability by contributing to the slow component of synaptic AHP.	−3.7	0.0026
MME	membrane metalloendopeptidase (MME) is a type II transmembrane glycoprotein and a common acute lymphocytic leukemia antigen	−2.9	0.0045
EGFR	Epidermal Growth Factor Receptor (EGFR) is a transmembrane glycoprotein that is a member of the protein kinase superfamily.	−2.87	0.01
FLT1	Fms Related Receptor Tyrosine Kinase 1 (FLT1) are receptor tyrosine kinases (RTKs) that contain an extracellular ligand-binding region with seven immunoglobulin (Ig)-like domains, a transmembrane segment, and a tyrosine kinase (TK) domain within the cytoplasmic domain.	−1.86	0.021
GDF15	Growth Differentiation Factor 15 (GDF15) encodes a secreted ligand of the TGF-beta (transforming growth factor-beta) superfamily of proteins.	2.7	0.003
KDR	Kinase Insert Domain Receptor (KDR) known as kinase insert domain receptor, is a type III receptor tyrosine kinase. It functions as the main mediator of VEGF-induced endothelial proliferation, survival, migration, tubular morphogenesis and sprouting.	−2.5	0.021
IGF1	Insulin-Like Growth Factor 1 (IGF1) is similar to insulin in function and structure and is a member of a family of proteins involved in mediating growth and development.	−1.76	0.03
KRT7	Keratin 7 (KRT7) is a type II cytokeratin consisting of basic or neutral proteins which are arranged in pairs of heterotypic keratin chains coexpressed during differentiation of simple and stratified epithelial tissues.	3.9	0.014

2.24 Mapping of regulatory elements

Various bioinformatics techniques and publicly accessible datasets were used to associate functional regulatory elements with the peak expression quantitative trait loci (eQTL) areas that were found for nine specific genes, AMACR, KCNN3, MME, EGFR, FLT1, GDF15, KDR, IGF1, and KRT7. The custom track built by Hazelett et al. [59] in the UCSC Genome Browser included prostate-specific mapping information. This information was collected from the GEO database. The DNase hypersensitivity tracks utilized for regions of open chromatin were GSE32970 and GSE29692. The histone modification tracks used in the study were H3K4me1 and H3K4me3 from dataset GSE27823, and H3K27Ac from dataset GSE51621, as well as H3K4me3 from samples GSM686935 and GSM503906. The transcription factors analyzed were those of interest, namely (GSE51621).

The data was gathered from prostate-specific cell lines, namely PrEC and PC3. Taberlay et al. utilized Chip-seq to collect data on H3K4me1, H3K27ac, H3K4me3, H3K27me3, CTCF, and RNAPolII. They then applied this data to the multivariate hidden Markov model ChromHMM52 in order to classify the epigenome into nine distinct chromatin states: poised promoter, promoter, promoter + CTCF, insulator, enhancer, enhancer + CTCF, transcribed, repressed, and heterochromatin.

Bioinformatics tools used for the purpose of mapping areas of interest include FunciSNP and Regulome (http://regulome.stanford.edu/). Regulome integrates many sources of information to provide a causal score for a single variation by associating SNPs with known and expected regulatory elements. There is little evidence suggesting that scores ranging from 1 to 3 are potentially responsible for gene dysregulation, but scores ranging from 4 to 6 do not contain such evidence. In addition to the score, the output includes information on transcriptional factor and histone modification binding sites, PWM (position weight matrix), DNase (deoxyribonuclease) activity, and eQTL (expression quantitative trait loci) data.

2.25 Prediction of mirna-mrna and circrna-mrna target genes and their regulatory networks

The TargetScan (https://www.targetscan.org/vert80/), miRTarBase (https://mirtarbase.cuhk.edu.cn/), and StarBase databases (https://starbase.sysu.edu.cn/) were used to determine the targets of differentially expressed genes and miRNAs. To find potential target miRNAs, search for the common elements across all three databases and the differentially expressed genes on microarrays. CircRNA-mRNA networks are constructed using the regulatory linkages between circRNAs and mRNAs. Cytoscape was used to visualize and identify the hub signalling route of the miRNA-mRNA regulatory network in order to determine the targets of differentially expressed genes and circRNAs. The GEO2R online analytic tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/) was used to identify DEGs and differentially expressed circular RNAs (DE-circRNAs). The selection criterion consisted of an absolute value of logFC more than 1.0 and an adjusted p-value less than 0.05. The chosen mRNAs and circRNAs were classified as DEGs and differentially expressed circRNAs (DE-circRNAs). Subsequently, Venny 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/) was used to generate a Venn diagram to identify the intersection.

3 Results

3.1 Degs screening

Through the use of the edgeR package, we successfully detected the differentially expressed genes (DEGs) among the various subgroups of prostate cancer. The threshold values for the adjusted P-value (FDR) were set to be less than 0.01, and a fold change of at least 2 was required. A total of 752 differentially expressed genes (DEGs) were identified via five pairwise comparisons of data from four subtypes of prostate cancer: luminal A, luminal B, her2 positive, and basal-like (Supplementary 1 and Fig. 2).

3.2 Protein class sorting and finding immune cell and their related genes

Significant differences in gene expression were detected in both prostatic and normal cell groups. Transcript analysis utilising PANTHER showed that differentially expressed RNAs consist of a diverse set of gene sequences that are spread throughout the immune cell. The sequences are linked to different molecular functions, including binding (GO:0005488),

Fig. 4 Using Enrich, we conduct an analysis of PPI, biological processes, molecular functions, cellular components, signaling pathways, and weighted gene co-expression network analysis (WGCNA) in the genes. The text includes the following topics: **A** PPI analysis of the prostate cancer microenvironment, **B** PPI analysis of immune cells in the prostate cancer microenvironment, **C** analysis of biological processes, **D** analysis of molecular functions, **E** analysis of cellular components, **F** WGCNA analysis, **G** heat map visualization of co-expression in WGCNA, and (**H**) identification of specific modules involved in immune cells in the prostate cancer microenvironment

ATP-dependent activity (GO:0140657), catalytic activity (GO:0003824), cytoskeletal motor activity (GO:0003774), molecular function regulation (GO:0098772), gene-specific transcriptional regulation (PC00264), molecular transducer activity (GO:0060089), structural molecule activity (GO:0005198), transcription regulation activity (GO:0140110), and transporter activity (GO:0005215). Moreover, they participate in the biological process of biological adhesion (GO:0022610). The genes AMACR, KCNN3, MME, EGFR, FLT1, GDF15, KDR, IGF1, and KRT7 have been identified as key genes that have a substantial impact on the microenvironment of prostate cancer (Fig. 3; Table 2).

3.3 Identification of key degs through PPI analysis

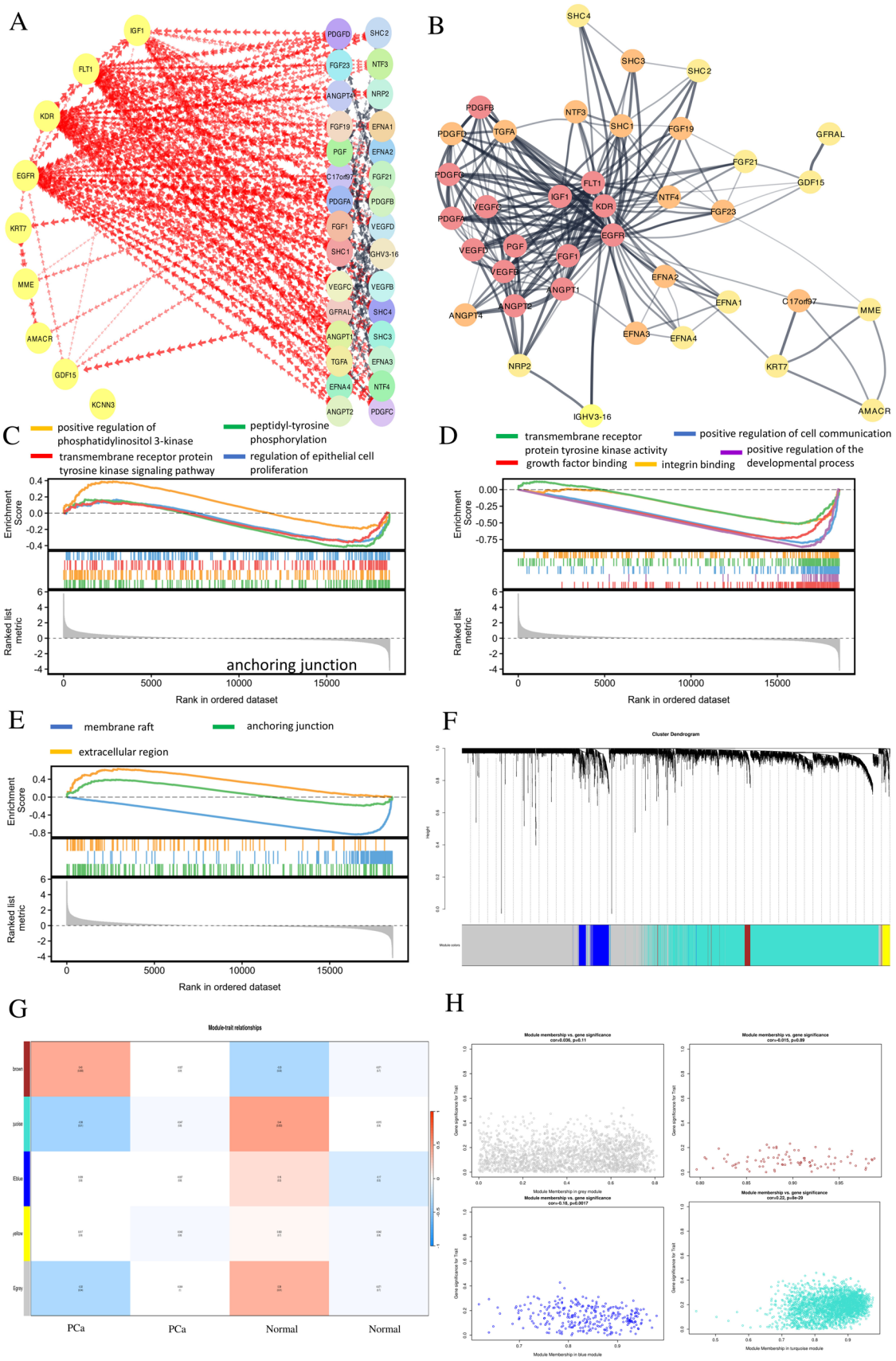
The master genes were identified by network analysis conducted using Cytoscape. Figure 4 used the yFiles radial structure to import the genes that were found in Fig. 4. The network determined the DEGs by analysing their interconnections. Based on the image, it can be seen that the 45 genes that have the highest level of interaction also demonstrate the most robust link. CD8A, GABRA3, HSP90BA, PECAM1, HLA-G, and Sect. 13 are the primary regulators of several genes, exerting unique control over them. Notable gene connections were observed in the different configurations of the Cytoscape network. The Cytoscape network analysis tool was used to authenticate the yfiles radial design. The Centiscape plugin has identified AMACR, ANGPT1, ANGPT2, ANGPT4, C17orf97, EFNA1, EFNA2, EFNA3, EFNA4, EGFR, FGF1, FGF19, FGF21, FGF23, FLT1, GDF15, GFRAL, IGF1, IGHV3-16, KCNN3, KDR, KRT7, MME, NRP2, NTF3, NTF4, PDGFA, PDGFB, PDGFC, PDGFD, PGF, SHC1, SHC2, SHC3, SHC4, TGFA, VEGFB, and lLUZP1 as the most significant hub genes based on their degree and betweenness centrality, as shown in Fig. 4A-B. The PPI network graph was visually examined using Cytoscape, with data sourced from the STRING database. Through the use of the CytoHubba plug-in, we successfully filtered genes that exhibited either up-regulation or down-regulation. As a result, we found a total of 50 differential genes inside the gene regulatory network. Out of them, a total of 14 hub target genes were identified as common, namely ANGPT1, ANGPT2, EGFR, FGF1, FLT1, IGF1, KDR, PDGFA, PDGFB, PDGFC, and PGF.

3.4 The biological process of enrichment analysis and its molecular functions are as follows

During the study using the Enrich programme, three Gene Ontology (GO) keywords were linked to up-regulated DEGs, whereas three GO terms were linked to down-regulated DEGs. The functional enrichment analysis showed that the biological processes of positive regulation of phosphatidylinositol 3-kinase, transmembrane receptor protein tyrosine kinase signalling pathway, peptidyl-tyrosine phosphorylation, and regulation of epithelial cell proliferation were significantly more active in the DEGs that were either upregulated or downregulated (Fig. 4C-D and Supplementary 2). The up-regulated DEGs exhibited an overrepresentation of molecular function (MF) keywords such as transmembrane receptor protein tyrosine kinase activity, growth factor binding, integrin binding, positive control of cell communication, and positive regulation of the developmental process. Curiously, co-expression analysis of genes that are down-regulated showed an excessive expression in the membrane raft, anchoring junction, and extracellular area. The up-regulated DEGs were found to have an overrepresentation of phrases such as membrane raft, anchoring junction, and extracellular area (Fig. 4E).

3.5 Construction of the co-expression network and identification of modules

The WGCNA method was used to find modules that are linked to various subtypes of prostate cancer. An analysis was conducted on the 6315 DEGs to determine their co-expression. We performed an investigation on the soft threshold power of the network design, using a range of β values spanning from 1 to 20. Next, we assessed the degree to which the co-expression network is unaffected by changes in size and measured the average level of connectivity within the network. Upon assessing the research data, a criteria of 5 was selected as the most suitable (Fig. 4F). The node degree distribution was analysed using a power-law distribution, indicating that the network exhibits a scale-free structure (Fig. 4G). Afterwards, the value of β, which was equal to 5, was used to create a gene tree using the process of hierarchical


A
B
C
positive regulation of phosphatidylinositol 3-kinase
peptidyl-tyrosine phosphorylation
transmembrane receptor protein tyrosine kinase signaling pathway
regulation of epithelial cell proliferation
Enrichment Score
Ranked list metric
anchoring junction
Rank in ordered dataset
D
transmembrane receptor protein tyrosine kinase activity
positive regulation of cell communication
growth factor binding
integrin binding
positive regulation of the developmental process
E
membrane raft
anchoring junction
extracellular region
F
Cluster Dendrogram
G
Module-trait relationships
PCa
PCa
Normal
Normal
H
Module membership vs. gene significance
Module Membership in grey module
Module Membership in blue module
Module Membership in turquoise module

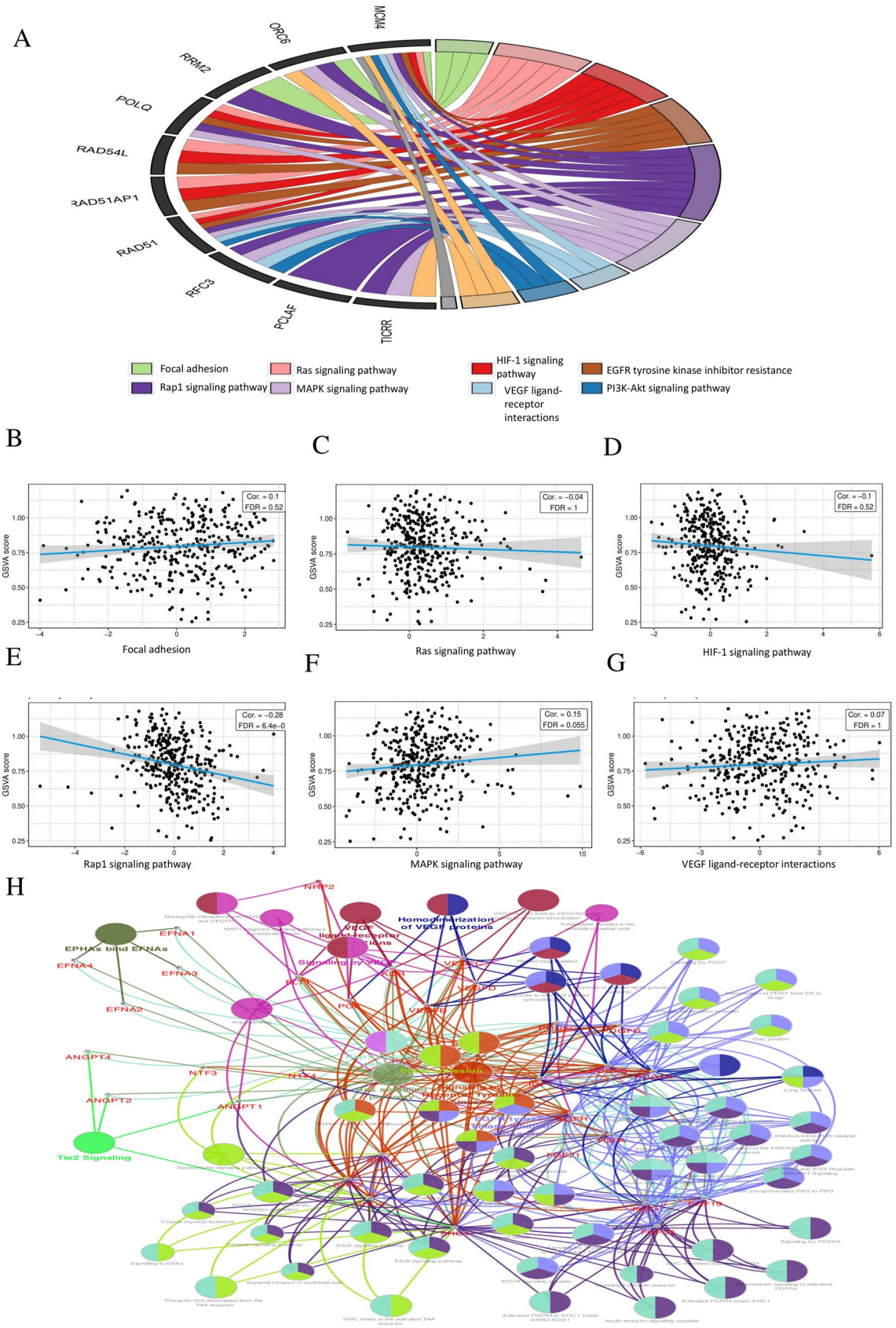


Fig. 5 Identification of the signaling pathway linked to immune cells in the prostate cancer microenvironment. **A** The signaling pathways involved in the prostate cancer environment include **B** Focal adhesion, **C** Ras signaling pathway, **D** HIF-1 signaling pathway, **E** Rap1 signaling pathway, **F** MAPK signaling pathway, **G** VEGF ligand-receptor interactions, and **H** immune cell activation and their associated signaling pathways

clustering. The expression level of genes inside a module is denoted by ME, which functions as the primary component for that specific module. There are two methodologies for investigating the correlation between each module and the corresponding subtypes of prostate cancer. Two modules were discovered that had significant associations with certain cancer subtypes ($P < 0.01$ and absolute value of correlation > 0.75). The modules associated with the subtypes of prostate cancer, namely the blue ($R = 0.8$; $P <$ 1e-168) and turquoise ($R = -0.71$; $P =$ 1e-113) modules, were found as significant using the Spearman correlation method in the analysis of module-trait associations. Figure 4 depict the interactions between the turquoise and blues modules, emphasising their genetic significance. Furthermore, our analysis of single-cell data uncovered nine key genes - AMACR, KCNN3, MME, EGFR, FLT1, GDF15, KDR, IGF1, and KRT7 - that have a significant correlation with PCa. Among them, GDF15, KRT76, and LSM1 were shown to have higher levels in PCa samples (Fig. 4G), whereas AMACR, KCNN3, MME, EGFR, FLT1, KDR, and IGF1 showed reduced expression (Fig. 4H).

3.6 KEGG pathway enrichment re-analysis for 15 hub genes

To examine the possible signalling pathways related to these 23 hub genes, we conducted a re-analysis of the KEGG pathway using the DAVID tool ($P < 0.05$). The genes are strongly linked to six signalling pathways: Focal adhesion, Rap1 signalling route, Ras signalling system, EGFR tyrosine kinase inhibitor resistance, MAPK signalling pathway, HIF-1 signalling pathway, PI3K-Akt signalling pathway, and VEGF ligand-receptor interactions (Fig. 5A). We performed a study on the expression levels of these four genes using the TCGA database. The findings demonstrated a significant rise in expression levels in both BC-adjacent and PCa patients as compared to those who are in good health (Fig. 5B-G). To get a more comprehensive knowledge of the underlying mechanism of these four genes in prostate cancer (PCa), we performed co-expression data mining using the pc-GenExMiner programme. Figure 5H and Supplementary 3 provide evidence that all nine genes showed either up-regulation or down-regulation in PCa tissues, indicating a signalling network.

3.7 Investigation of immune infiltration and prostate cancer

Figure 6A-B depicts the abundance of immune cells in the samples. The research revealed that tumour cells had a higher abundance of immature B cells, resting memory CD4 T cells, and several kinds of macrophages (M0, M1, M2). In contrast, tumour cells exhibited lower levels of memory B cells, plasma cells, CD8 T cells, activated memory CD4 T cells, and monocytes. Figure 6A-B. Figure 5C demonstrates the variation in expression levels of many immune cells inside tumour cells, as shown by the heatmap. The activation of CD4 memory T cells was shown to be associated with the chance of survival. A significant proportion of CD4 memory-activated T cells had a high probability of survival, as seen in Fig. 6A-B. The study demonstrated the link between 22 different immune cell types, as seen in Fig. 6A-B. Younger people had higher amounts of naive B cells, quiescent Mast cells, and activated NK cells, in comparison to older persons. In contrast, younger people had lower amounts of active Mast cells, Neutrophils, and resting NK cells in comparison to older persons. The quantities of Macrophage M0, active Mast cells, and Plasma cells were higher in G1/2 in comparison to G3 grade. In contrast, the numbers of Macrophage M1, resting Mast cells, Monocytes, activated T cells CD4 memory, CD8 T cells, and follicular helper T cells were lower in G1/2 compared to G3 grade.

3.8 Monocytes and neutrophils in prostate cancer

Following that, we saw no significant difference in the probability of survival between the group at high risk and the group at low risk, as seen in Fig. 6C-D. The AUC curve data demonstrated that the nomogram survival prediction model achieved a good degree of accuracy. The AUC values at 1 year, 3 years, and 5 years were 0.569, 0.525, and 0.605, respectively, as seen in Fig. 6C-D. The Cox regression analysis discovered two immune cell types, Monocytes and Neutrophils, which exhibited predictability and interconnection. The Nomogram was created using immune cell data to predict survival rates during the first 3 to 5 years. The immune cells used in this nomogram model were Monocytes and Neutrophils, as shown in Fig. 6C-D. In addition, the calibration curve for nomogram models consistently showed a strong correlation between the actual survival rate and the predicted 3-year overall survival rate, demonstrating the accuracy of this model in generating forecasts (Fig. 6C-D).

Fig. 6 The correlation between immune cells and survival in PCa. **A** The left side displays the survival probability of both high and low-risk ▶ categories. The area under the curve (AUC) with time, seen on the right side. **B** Nomogram for forecasting patients' prognosis based on immunological cells (left). **C** The calibration curves are used to evaluate the ability of the nomogram (shown on the right) to accurately and effectively discriminate between different values. The correlation between immune cells and prognostic genes. Immune cell heatmap comparing monocytes and neutrophils in high-risk and low-risk groups. **D** The correlation matrix examines the relationship between immune cells and genes. **E** It investigates the connection between the survival of patients with prostate cancer and certain genes

3.9 Survival and expression analysis of the hub genes

A total of 96 hub genes were found throughout the study process. These genes were selected based on their gene significance (GS) values being more than 0.2 and their module membership (MM) values being greater than 0.8 in the blue module. The Gene Expression Profiling Interactive Analysis (GEPIA) tool was used to examine the correlation between overall survival and statistical significance, with a significance level of $P < 0.05$. Eight genes had strong correlations with patient prognosis and demonstrated elevated expression levels in relation to the severity of the disease (Fig. 6E). The essential genes are AMACR, KCNN3, MME, EGFR, FLT1, GDF15, KDR, IGF1, and KRT7.

3.10 Survival and expression analysis of the hub genes

A total of 96 hub genes were found throughout the study process. These genes were selected based on their gene significance (GS) values being more than 0.2 and their module membership (MM) values being greater than 0.8 in the blue module. The Gene Expression Profiling Interactive Analysis (GEPIA) tool was used to examine the correlation between overall survival and statistical significance, with a significance level of $P < 0.05$. Eight genes had strong correlations with patient prognosis and demonstrated elevated expression levels in relation to the severity of the disease (Fig. 6E). The essential genes are AMACR, KCNN3, MME, EGFR, FLT1, GDF15, KDR, IGF1, and KRT7.

3.11 Validation of hub genes in a single-cell microenvironment, cell-cell communication, and Characteristics of PC microenvironment in single-cell RNA sequencing data sets

Subsequently, we analysed the exact distribution of the genes in PCa by using our data sets acquired from single-cell RNA sequencing. We used biomarkers to categorise cell kinds and visualised them in an Umap plot. We used a pie diagram (Fig. 7A-B) to graphically represent the prevalence of each cell type in CRPC and HSPC. The PCa of individual cells was assessed using the same nomogram used for bulk RNA-seq in CRPC and HSPC samples. The conclusions we made are consistent with the previous findings acquired from bulk RNA sequencing, suggesting that PCa cells often exhibited greater PRS values compared to hematopoietic stem and progenitor cells (HSPC). Violin plots were used to enhance comprehension of the divergences in PRS across several cell types. The plots consistently demonstrated significantly higher PRS values in luminal cells, myeloid cells, and endothelial cells of CRPC compared to HSPC (Fig. 7C-D). Notably, there was a substantial difference in PRS (proliferation rate score) between CRPC (castration-resistant prostate cancer) and HSPC (hormone-sensitive prostate cancer) within the luminal cell population. Further analysis, specifically focusing on luminal cells, revealed that most of these cells had a low PRS. In contrast, cells exhibiting a high PRS were mostly seen in the CRPC samples, as shown in Fig. 7C and D. To examine the relationship between the diversity of luminal cells within cells and PRS in CRPC and HSPC, we classified luminal cells into high-risk and low-risk score groups. Subsequently, we performed irGSEA (Iterative Gene Set Enrichment Analysis) on CRPC (Castration-Resistant Prostate Cancer) and HSPC (Hormone-Sensitive Prostate Cancer), both in the high-risk and low-risk score groups, in order to evaluate the divergent activation of essential pathways in luminal cells. The ssGSEA analysis revealed a significant upregulation of the androgen response pathway in the high PRS group compared to the low PRS group, as shown in Fig. 5F and Supplementary 4. The consistency of this finding was seen when comparing the group diagnosed with castration-resistant prostate cancer (CRPC) to the group diagnosed with hormone-sensitive prostate cancer (HSPC) (Fig. 7E and F). In order to validate the robustness and dependability of our findings, we conducted an identical study on an additional dataset of 25 samples. The second data set confirmed the ongoing rise in activity of the androgen response pathway, first identified by ssGSEA. These data suggest that there may be a link between increased PRS and an increased probability of getting CRPC (Fig. 8A-G).

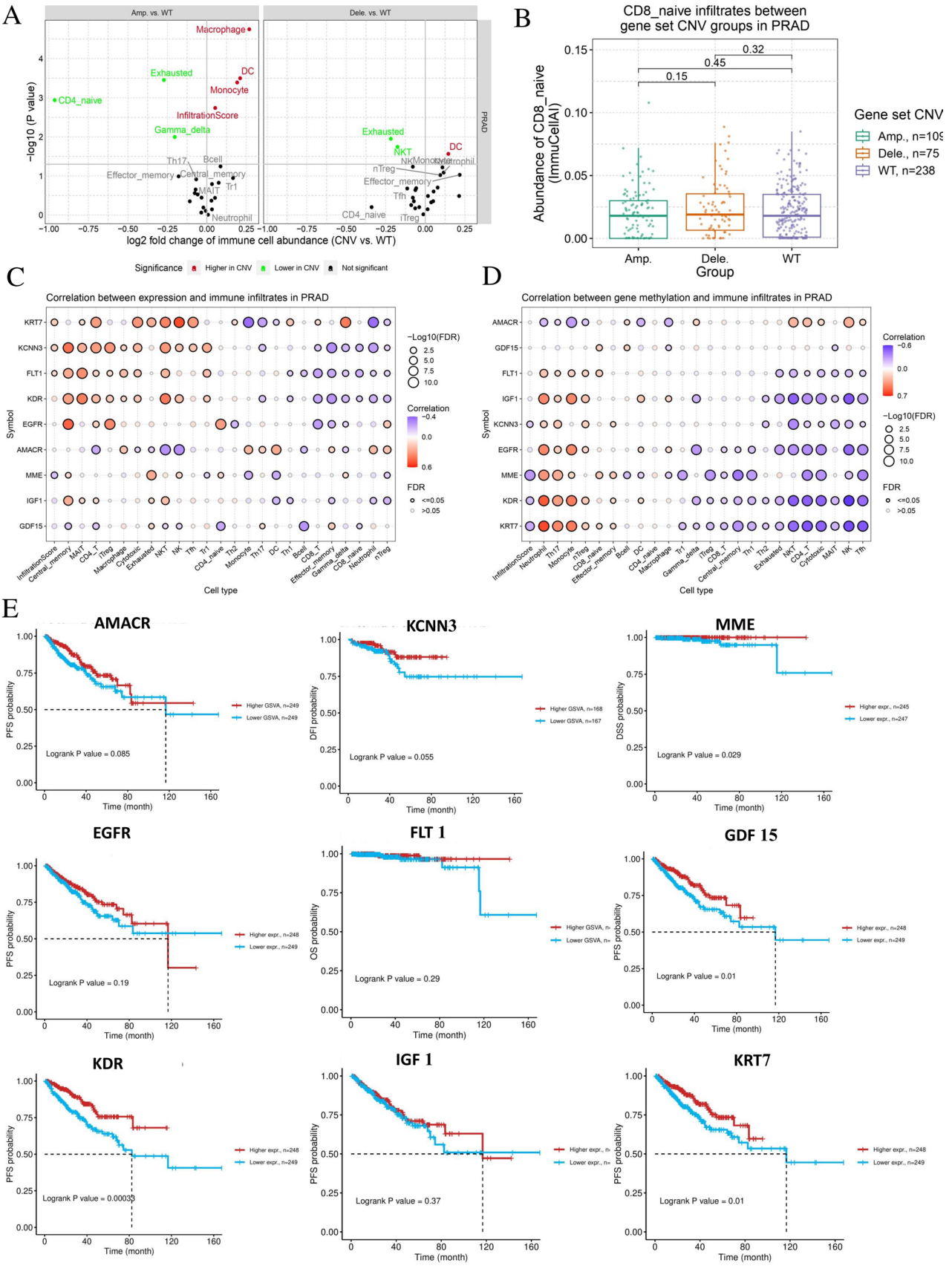

A
Amp. vs. WT
Dele. vs. WT
PRAD
−log10 (P value)
log2 fold change of immune cell abundance (CNV vs. WT)
Macrophage
Exhausted
DC
Monocyte
CD4_naive
InfiltrationScore
Gamma_delta
Th17
Bcell
Effector_memory
Central_memory
Tr1
MAIT
Neutrophil
NKT
nTreg
Tfh
iTreg
B
CD8_naive infiltrates between gene set CNV groups in PRAD
Abundance of CD8_naive (ImmuCellAI)
0.32
0.45
0.15
Amp.
Dele.
WT
Group
Gene set CNV
Amp., n=109
Dele., n=75
WT, n=238
C
Significance Higher in CNV Lower in CNV Not significant
Correlation between expression and immune infiltrates in PRAD
KRT7
KCNN3
FLT1
KDR
EGFR
AMACR
MME
IGF1
GDF15
Symbol
−Log10(FDR)
Correlation
FDR
<=0.05
>0.05
Cell type
D
Correlation between gene methylation and immune infiltrates in PRAD
E
AMACR
KCNN3
MME
EGFR
FLT 1
GDF 15
KDR
IGF 1
KRT7
PFS probability
DFI probability
OS probability
DSS probability
Time (month)
Logrank P value = 0.085
Logrank P value = 0.055
Logrank P value = 0.029
Logrank P value = 0.19
Logrank P value = 0.29
Logrank P value = 0.01
Logrank P value = 0.00033
Logrank P value = 0.37
Logrank P value = 0.01
Higher GSVA, n=249
Lower GSVA, n=249
Higher GSVA, n=168
Lower GSVA, n=167
Higher expr., n=245
Lower expr., n=247
Higher expr., n=248
Lower expr., n=249

Fig. 7 Identification of a PCa-specific hub gene in single-cell. (**A**, **B**) The 25 samples were split into eleven clusters. (**C**, **D**) The analysis of 12 ▶ normal and 13 PCa samples revealed four distinct clusters, highlighting the diversity of cell types involved in the prostate. These clusters included luminal epithelia (LE), basal epithelia (BE), other epithelia (OE), smooth muscle cell (SMC), fibroblast (FB), endothelial cell (EC), T cell (T), mast cell (MC) and B cell (BC). Single-cell transcriptomic analysis identified nine hub genes – (**E**) AMACR and (**F**) KCNN3 - as significantly associated downregulated with prostate cancer

3.12 Gene-based eqtl

The 9 target gene regions of interest were classified into three groups according to the Linkage disequilibrium (LD) between the PrCa-risk SNP and the strongest eQTL signal. The classification was conducted by taking into account both primary and secondary analyses, as seen in Fig. 9A-B. The Pearson correlation coefficient was used to measure the linkage disequilibrium (LD) between the single nucleotide polymorphism (SNP) associated with the risk of prostate cancer (PrCa-risk SNP) and the SNP that shows the strongest connection with the level of gene expression (peak eQTL signal). To determine the existence of multiple independent regulatory single nucleotide polymorphisms (SNPs) in the target genes, we performed a statistical analysis to evaluate the correlation of each SNP within the target gene region, taking into consideration the primary expression quantitative trait locus (eQTL) signal SNP.

3.13 Characteristics of significant cis-eqtl findings

Among the 51 risk intervals that exhibited a significant expression quantitative trait locus (eQTL) signal, 33 demonstrated significant associations between a single gene and a single nucleotide polymorphism (SNP). In the remaining 18 risk intervals, the SNPs associated with risk were shown to be connected to two or more genes. The number of genes linked to each risk SNP exhibited variation, spanning from 2 to 7 genes per region.

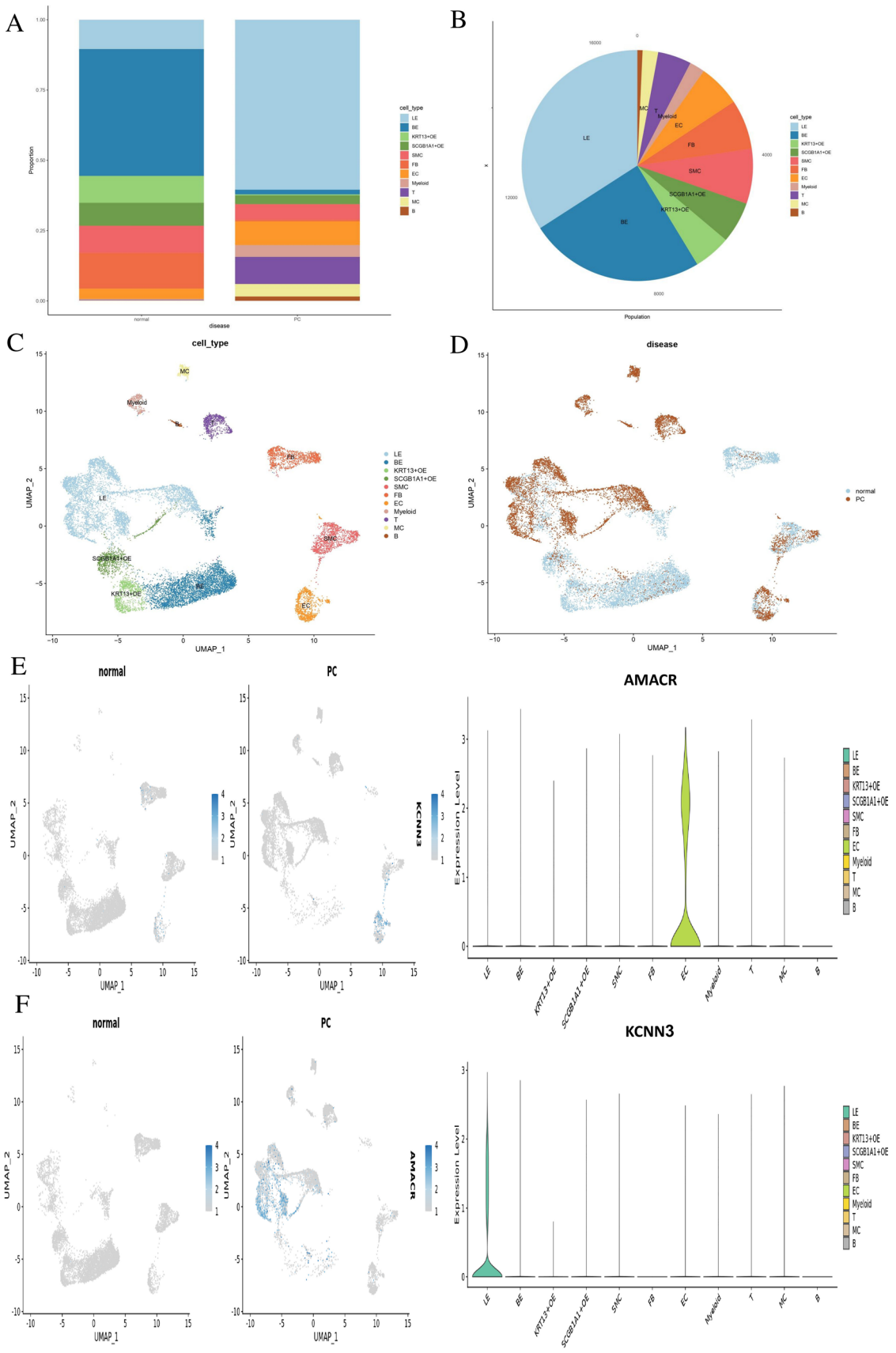
In addition, out of these 12 sites, 10 of them showed multiple identified risk SNPs in close proximity. The number of SNPs associated with risk ranged from 2 to 5 per area. Specifically, there were 7 regions with 2 risk SNPs, 2 regions with 3 risk SNPs, and 1 region with 5 risk SNPs. The linkage disequilibrium (LD) between the risk SNPs varied, with r2 values ranging from below 0.2 to 1. More precisely, there were 15 SNPs with a linkage disequilibrium (r2) value less than 0.25, 3 SNPs with a r2 value ranging from 0.5 to 0.7, and 4 SNPs with a r2 value ranging from 0.7 to 1.0. Figure 9C-E illustrate six instances where the correlation coefficient (r2) between the risk SNPs was below 0.7. More precisely, the single nucleotide polymorphisms (SNPs) associated with risk in the KRT7 and KDR genes showed a correlation value of 0.54, but the SNPs related to FLT1 had a correlation coefficient lower than 0.2. In addition, the correlation coefficient for KCNN3 was 0.67, but for the other areas, it was below 0.25 (Fig. 9F-H).

3.14 Micrornas and circrna candidate

During this stage, we have discovered a total of 15 genes, including AMACR, KCNN3, MME, EGFR, FLT1, GDF15, KDR, IGF1, and KRT7. Subsequently, we isolated and selected the most relevant microRNAs, as seen in Fig. 10A-C. The microRNAs hsa-miR-518c, has-miR-518 b, hsa-miR-518a-3p, has-miR-518f, hsa-miR-518d-3p, has-miR-718, mmu miR-126-3p, hsa-miR-4304, has-miR-598, and has-miR-1471 shown greater clarity in comparison to other microRNAs, as seen in Fig. 10. The following circular RNAs are detected: hsa_circ_0003962, hsa_circ_0001726, hsa_circ_0072697, hsa_circ_0001555. The circular RNAs hsa_circ_0003574, hsa_circ_0079136, hsa_circ_0006213, hsa_circ_0053028, hsa_circ_0001730, hsa_circ_0003522, and hsa_circ_0002554 are being mentioned.Ribonucleic acids (RNAs) have the capacity to regulate the AMACR, KCNN3, MME, EGFR, FLT1, GDF15, KDR, IGF1, and KRT7 genes, as seen in Fig. 10D-H.

4 Discussion

In this study, we conducted a comprehensive analysis to DEGs among various subtypes of prostate cancer, uncovering 752 DEGs through five pairwise comparisons using edgeR. These DEGs were further examined to reveal significant insights into the molecular landscape of prostate cancer. The PANTHER analysis elucidated the wide-ranging molecular functions of these genes, highlighting their involvement in critical biological processes and functions, such as binding, ATP-dependent activity, and gene-specific transcriptional regulation. Notably, genes like AMACR, KCNN3, MME, and EGFR


A
B
C
D
E
F
cell_type
disease
normal
PC
AMACR
KCNN3
UMAP_1
UMAP_2
Expression Level
Proportion
Population
LE
BE
KRT13+OE
SCGB1A1+OE
SMC
FB
EC
Myeloid
T
MC
B

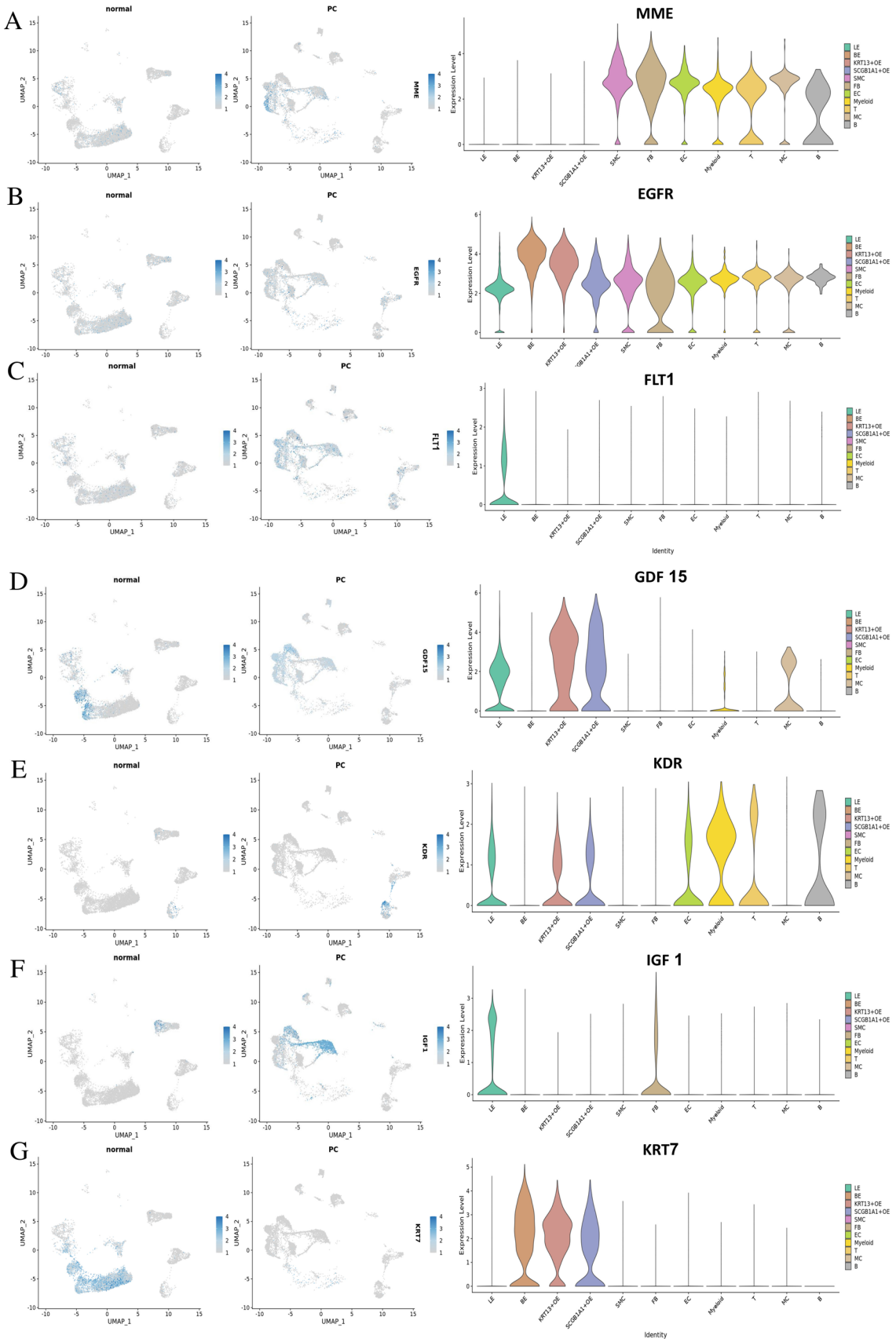


Fig. 8 Characteristics of PCa microenvironment in single-cell RNA sequencing data sets. **A** MME, **B** EGFR, **C** FLT1, **D** KDR, **E** IGF1, and were found to be downregulated and (**F**) GDF15 and (**G**) KRT7 were upregulated, suggesting their potential roles in the pathogenesis of prostate cancer

emerged as pivotal players within the prostate cancer microenvironment, suggesting their potential roles as therapeutic targets or biomarkers. The integrated genomic profiling of prostate cancers has yielded extensive insights and innovative findings that enhance our comprehension of the illness. Several metastatic castration-resistant prostate cancers include clinically identifiable genetic abnormalities, including modifications in DNA damage repair pathways and PTEN/PI3K signaling. Genomic changes in TP53, RB1, AR, and the cell cycle pathway correlate with worse clinical results, whereas SPOP mutations are linked to improved clinical outcomes. Numerous genetic profiling assays are being developed to identify individuals who may benefit from targeted treatment. Consequently, genetic sequencing of prostate cancer offers valuable insights for diagnosis and therapy in the contemporary landscape of precision medicine.

Our findings were substantiated through multiple analytical approaches, including PPI network analysis using Cytoscape and WGCNA, which identified significant hub genes such as ANGPT1, EGFR, and IGF1. The enrichment analyses underscored the relevance of pathways like the PI3K-Akt and MAPK signaling pathways, which are known to be crucial in cancer progression. Additionally, the investigation of immune cell infiltration revealed distinct patterns of immune cell presence and their correlation with patient prognosis, indicating the intricate relationship between the tumor microenvironment and prostate cancer progression. The validation of these hub genes in single-cell RNA sequencing data sets further confirmed their critical roles in prostate cancer pathogenesis, providing a robust foundation for future research and potential clinical applications.

The majority of patients with prostate cancer often have extended life periods and positive prognoses due to the efficacy of therapeutic interventions such as androgen deprivation therapy (ADT) or aggressive surgery [60]. However, some people who acquire advanced PCa or CRPC have worse outcomes, which are marked by the return of the disease or the development of separate metastases. Therefore, it is essential to explore novel biomarkers that can accurately forecast the return of PCa. Our investigation included the identification of eight prognostic differentially expressed genes (prDEGs) and the development of a prognosis risk model (PRM) based on progression-free survival (PFS). This model offers a reasonably precise clinical prognostic tool that may assist in making therapy choices for individuals with PCa. In this study, we first found 1454 differentially expressed genes (DEGs) that are regulated together. Furthermore, we observed that these patients may be categorized into three distinct subgroups based on the expression patterns of these DEGs. The Kaplan-Meier curves of the three subtypes exhibited statistically significant variations in progression-free survival (PFS).

The gene AMACR, which is responsible for repairing DNA damage via homologous recombination and DNA repair mechanisms, has been discovered to enhance the advancement of CRPC [61]. KCNN3 has been identified as a tumor suppressor that hinders the development of tumors in prostate cancer [62]. However, abnormal expression of MME has been seen to be increased in PCa and is associated with aggressive advancement [63]. EGFR is linked to the advancement of prostate cancer via the stimulation of c-MET signaling [64]. Varkaris et al. showed that the levels of NRP1 were increased after androgen target therapy and identified NRP1 as a predictive biomarker of mCRPC [64]. According to Huo et al., BCYRN1 enhances the growth, glucose processing, and survival of prostate cancer cells by upregulating the expression of HDAC11 in PCa. NDUFV1 deficiency is a prevalent factor in mitochondrial dysfunction [65]. PXN may function as a regulator in both castration-sensitive prostate cancer and CRPC. LCN2 has been discovered as a tumor suppressor in colon cancer and pancreatic cancer, whereas it promotes carcinogenesis in PCa [66].

Using the eight genes stated earlier, we performed calculations to determine the Polygenic Risk Score (PRS) and created the Polygenic Risk Model (PRM) to provide prognostic predictions for patients. Significantly, both the training and validation sets showed that individuals classed as high-risk had a notably poorer progression-free survival (PFS), which indicates a greater chance of PCa recurrence. In addition, the AUC values for 1-, 3-, and 5-year progression-free survival (PFS) showed very positive results in both the training and validation sets, highlighting the strong and consistent prediction power of our risk model. By conducting both univariate and multivariate Cox regression studies, we have shown that the PRS (Polygenic Risk Score) functions as an accurate and superior independent clinical prognostic predictor for PCa. This model shows great promise as a valuable addition to the current GS system.

Our extensive research yields convincing data that supports the crucial function of SYK dysregulation in the advancement of PCa, thereby establishing it as a prospective target for therapeutic intervention. The increase in SYK expression in tumor tissues, together with its association with worse progression-free survival (PFS), highlights its clinical importance as a prognostic indicator. Moreover, our laboratory tests provide additional evidence of the significant impact of SYK on the growth and movement of PCa cells, further confirming its role in the development of the disease. The mechanisms by which GDF15 and KRT7 function in malignancies need further elucidation. Tumor-associated neutrophils facilitate cancer progression via GDF15 and KRT7-mediated immunosuppression, with alterations in the tumor microenvironment resulting from TAN invasion. Additionally, GDF15 and KRT7 operate as tumor eliminators by enhancing T cell-mediated tumor eradication or by releasing cytotoxic agents. The dual impacts of GDF15 and KRT7 may

Fig. 9 A regional association map is shown for the gene area encompassing the FLT3, KRT7, KCNN3, MME, KDR genes. This region includes many single nucleotide polymorphisms (SNPs) associated with prostate cancer risk, which have different levels of linkage disequilibrium (LD) between them. The x-axis represents the chromosomal position of the SNPs, with the analyzed gene displayed below. The y-axis represents the -log10(P value) obtained by regressing normalized expression levels for the gene listed in the panel title on the number of minor alleles of each SNP genotype, adjusted for histologic characteristics and 14 expression principal components. The PrCa-risk single nucleotide polymorphism (SNP) is shown by a dotted red vertical line, and the expression quantitative trait locus (eQTL) result is illustrated as a diamond. A regional association map is shown for the FLT3 gene (**C**), KRT7 gene (**D**), KCNN3 genes (**E**), MME gene (**F**), and KDR gene (**G**). The gene area contains many single nucleotide polymorphisms (SNPs) associated with prostate cancer risk, with different levels of linkage disequilibrium (LD) between them. Cell-cell contact between prostate cancer cells is mediated by intercellular junctions and (**H**, **I**). Figure 9H-I shows that these genes are highly communicated with immune cells

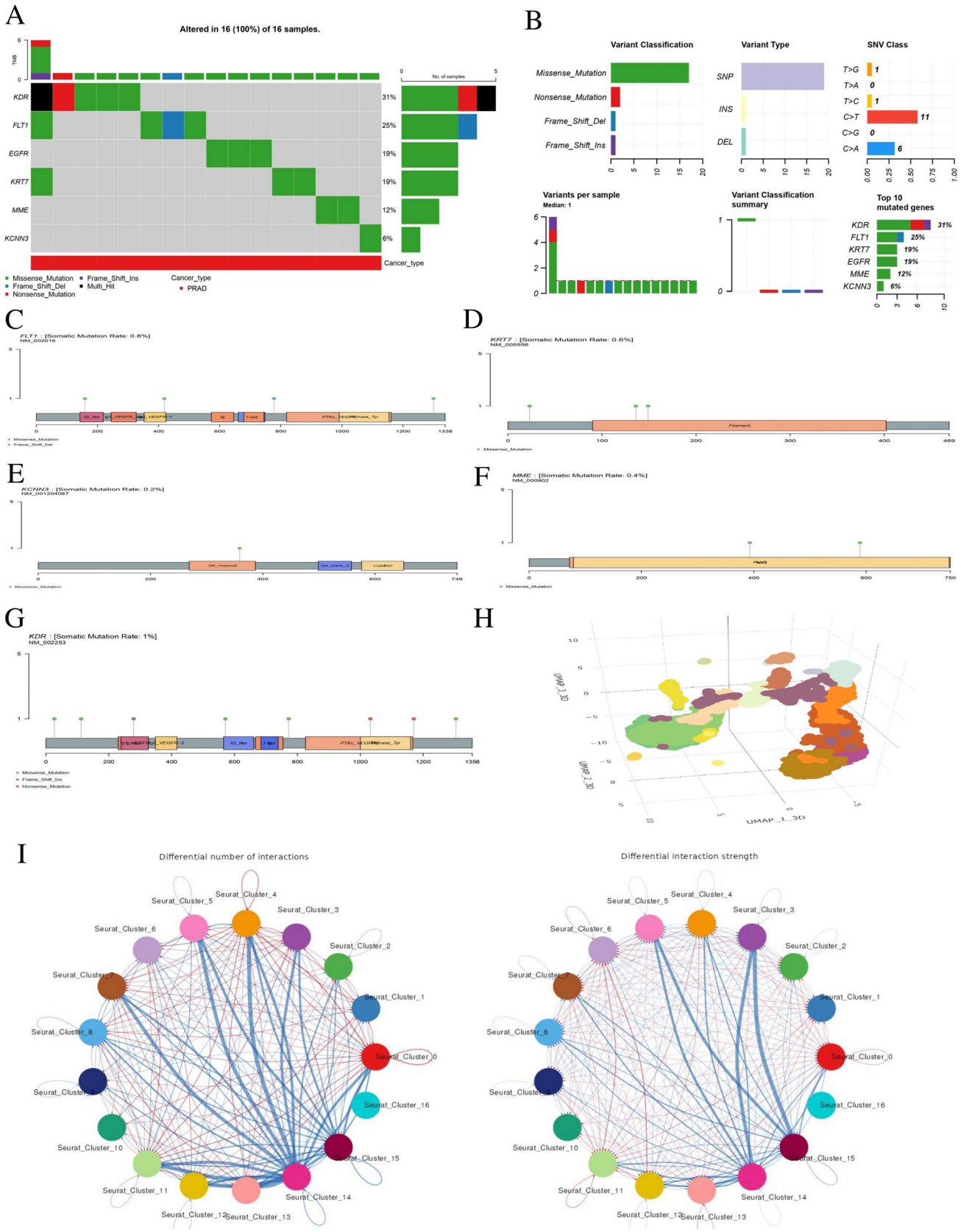
contribute to the significant heterogeneity or flexibility of neutrophils. Furthermore, the distinct microenvironment at various tumor stages influences the diverse functions of GDF15 and KRT7.

Despite the notable discoveries of our current research, it is crucial to recognize certain constraints. Although our comprehensive analysis has yielded useful insights into the PRM and its consequences for PCa, it is crucial to supplement these results with direct experimental data in future research. Furthermore, our findings have shown significant changes in the molecular properties of PCa, which need further targeted research to confirm these observed associations. Moreover, it is essential to carry out further investigation in order to determine the effectiveness of this prediction model for uncommon pathological subtypes, such as neuroendocrine carcinoma and ductal adenocarcinoma, which make up a tiny percentage of cases. However, we are dedicated to attentively observing new clinical data to further confirm the suggested PRM in future efforts.

The limitations of our study include several key factors that need to be considered. Firstly, the sample size used for the analysis was relatively small, which may limit the generalizability of our findings. Future studies should aim to include larger patient cohorts to validate the robustness of the identified DEGs and hub genes. Secondly, while we identified significant DEGs and pathways, experimental validation of these findings is necessary to confirm their roles in prostate cancer progression, immune cell infiltration, and the tumor microenvironment. Without such validation, the functional relevance of these genes remains speculative. Thirdly, the lack of comprehensive clinical data, such as detailed treatment histories and patient outcomes, restricts our ability to correlate gene expression profiles with clinical progression and response to therapy. Finally, while bioinformatics tools like PANTHER, Cytoscape, and WGCNA were used to identify key genes and pathways, these computational predictions require further biological verification to ensure their accuracy and relevance in clinical settings. Addressing these limitations will enhance the reliability and applicability of our findings.

5 Conclusion

To summaries, we have suggested an integrative study that combines gene expression profiles and a rebuilt protein-protein interaction network for prostate cancer. This approach differs from standard approaches that just examine differential gene expression or protein expression. This research specifically focused on doing a thorough examination of modular analysis to investigate the course of various disease stages in prostate cancer. The successful identification of key modules led to the discovery of possible disease-associated genes, which were then considered as prospective biomarkers for the modules. It may be used efficiently as a distinguishing characteristic to differentiate between control and illness samples. Regarding the functional study of putative disease-associated genes, it is worth noting that a significant proportion of these genes were found to be concentrated in the nucleus. Furthermore, these genes were shown to have diverse activities, with a particular emphasis on encoding transcription factors. The interaction between immune cells in the prostate cancer environment is crucial for maintaining gene stability and serves as a central hub for regulating several genes involved in immune gene expression. In our future work, we will conduct a more extensive


A
Altered in 16 (100%) of 16 samples.
TMB
KDR
FLT1
EGFR
KRT7
MME
KCNN3
31%
25%
19%
19%
12%
6%
No. of samples
Cancer_type
PRAD
Missense_Mutation
Frame_Shift_Ins
Frame_Shift_Del
Multi_Hit
Nonsense_Mutation
B
Variant Classification
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Del
Frame_Shift_Ins
Variant Type
SNP
INS
DEL
SNV Class
T>G 1
T>A 0
T>C 1
C>T 11
C>G 0
C>A 6
Variants per sample
Median: 1
Variant Classification summary
Top 10 mutated genes
KDR 31%
FLT1 25%
KRT7 19%
EGFR 19%
MME 12%
KCNN3 6%
C
FLT1 : [Somatic Mutation Rate: 0.8%]
NM_002019
D
KRT7 : [Somatic Mutation Rate: 0.6%]
NM_005556
E
KCNN3 : [Somatic Mutation Rate: 0.2%]
NM_001204087
F
MME : [Somatic Mutation Rate: 0.4%]
NM_000902
G
KDR : [Somatic Mutation Rate: 1%]
NM_002253
H
UMAP_1_3D
UMAP_2_3D
UMAP_3_3D
I
Differential number of interactions
Differential interaction strength
Seurat_Cluster_0
Seurat_Cluster_1
Seurat_Cluster_2
Seurat_Cluster_3
Seurat_Cluster_4
Seurat_Cluster_5
Seurat_Cluster_6
Seurat_Cluster_7
Seurat_Cluster_8
Seurat_Cluster_10
Seurat_Cluster_11
Seurat_Cluster_12
Seurat_Cluster_13
Seurat_Cluster_14
Seurat_Cluster_15
Seurat_Cluster_16

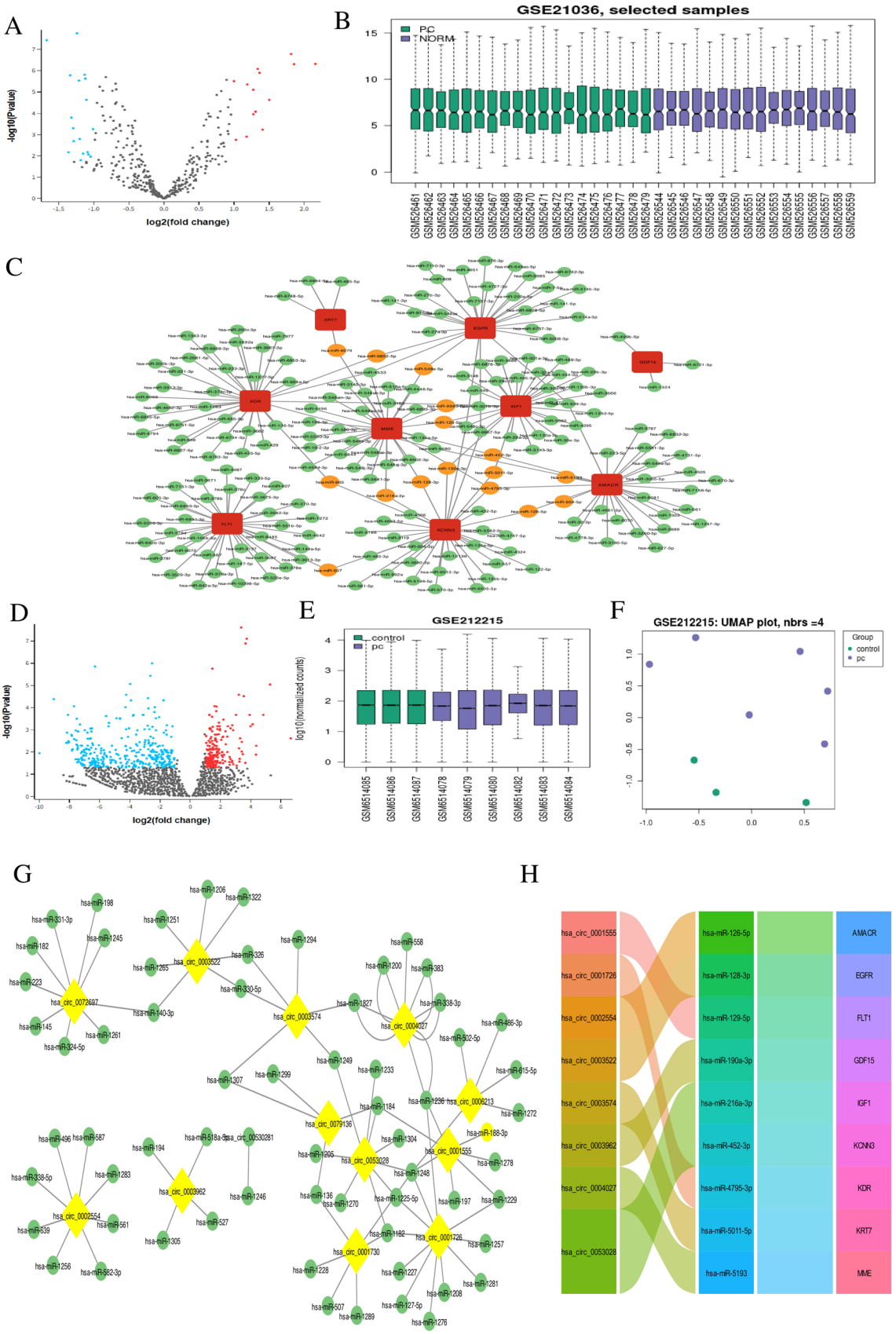


Fig. 10 P-value, q-value, and overlap gene with candidate microRNAs and circRNA extracted in miRNA and circRNA databases

investigation into the experimental verification of suspected disease genes and pathways involved in the development of prostate cancer.

Acknowledgements This work is supported by Tehran University.

Author contributions D.H.K: data assembly, manuscript writing, bioinformatics, supervision, editing, and performed experiments, Sobhan Bavandi: data assembly, Ali Shakeri Abroudi: bioinformatics analyses, and Melika Djamali: bioinformatics analyses. All authors have read and agreed to the published version of the manuscript.

Funding None.

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request or visit https://zenodo.org/records/12578498.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Shokri Varniab Z, et al. The levels and trends of cancer incidence in the elderly population at national and sub-national scales in Iran from 1990 to 2016. Cancer Rep 2024;7:1937.
2. Kucera R, et al. Prostate cancer management: long-term beliefs, epidemic developments in the early twenty-first century and 3PM dimensional solutions. EPMA J. 2020;11(3):399–418.
3. Spratt DE, et al. Treating the patient and not just the cancer: therapeutic burden in prostate cancer. Prostate Cancer Prostatic Dis. 2021;24(3):647–61.
4. Pagliarulo V. *Androgen deprivation therapy for prostate cancer.* Mol Diagn Imag Prost Cancer: Clin Appl Treat Strateg. 1–30. 2018.
5. Harris AE, et al. Exploring anti-androgen therapies in hormone dependent prostate cancer and new therapeutic routes for castration resistant prostate cancer. Front Endocrinol. 2022;13:1006101.
6. Bai B, et al. Molecular basis of prostate Cancer and Natural products as potential chemotherapeutic and Chemopreventive agents. Front Pharmacol. 2021;12:738235.
7. Testa U, Castelli G, Pelosi E. Cellular and molecular mechanisms underlying prostate Cancer development: therapeutic implications. Medicines. 2019;6(3):82.
8. Liu A, et al. The heterogeneity and clonal evolution analysis of the advanced prostate cancer with castration resistance. J Transl Med. 2023;21(1):641.
9. Granata I, Barboro P. Identification of molecular markers associated with prostate Cancer subtypes: an integrative bioinformatics approach. Biomolecules. 2024;14(1):87.
10. Pan J, et al. The prevalence and prognosis of next-generation therapeutic targets in metastatic castration-resistant prostate cancer. Mol Oncol. 2022;16(22):4011–22.
11. Pinet S, et al. Clinical management of molecular alterations identified by high throughput sequencing in patients with advanced solid tumors in treatment failure: real-world data from a French hospital. Front Oncol. 2023;13:1104659.
12. Yoo J, et al. Epigenetic roles of KDM3B and KDM3C in tumorigenesis and their therapeutic implications. Cell Death Dis. 2024;15(6):451.
13. Meng J, et al. Immune response drives outcomes in prostate cancer: implications for immunotherapy. Mol Oncol. 2021;15(5):1358–75.
14. Khosravi GR, et al. Immunologic tumor microenvironment modulators for turning cold tumors hot. Cancer Commun (Lond). 2024;44(5):521–53.
15. Hu A, et al. Harnessing innate immune pathways for therapeutic advancement in cancer. Signal Transduct Target Therapy. 2024;9(1):68.
16. Cha HR, Lee JH, Ponnazhagan S. Revisiting immunotherapy: a focus on prostate Cancer. Cancer Res. 2020;80(8):1615–23.
17. Sridaran D, et al. Prostate cancer immunotherapy: improving clinical outcomes with a multi-pronged approach. Cell Rep Med. 2023;4(10):101199.

18. Yang Q, et al. The role of tumor-associated macrophages (TAMs) in tumor progression and relevant advance in targeted therapy. Acta Pharm Sin B. 2020;10(11):2156–70.
19. Fan Y, et al. MLXIPL associated with tumor-infiltrating CD8 + T cells is involved in poor prostate cancer prognosis. Front Immunol. 2024;15:1364329.
20. Sobhani N, et al. CTLA-4 in regulatory T cells for Cancer immunotherapy. Cancers 2021;13(6):1440.
21. Shaopeng Z, et al. Regulation of regulatory T cells and tumor-associated macrophages in gastric cancer tumor microenvironment. Cancer Med. 2024;13(2):e6959.
22. Zhao H, et al. Inflammation and tumor progression: signaling pathways and targeted intervention. Signal Transduct Target Therapy. 2021;6(1):263.
23. Cao P, et al. TGF-β enhances immunosuppression of myeloid-derived suppressor cells to Induce transplant immune tolerance through affecting Arg-1 expression. Front Immunol. 2022;13:919674.
24. Morrison BJ, Steel JC, Morris JC. Reduction of MHC-I expression limits T-lymphocyte-mediated killing of Cancer-initiating cells. BMC Cancer. 2018;18(1):469.
25. Chang J, Ye JC. Bidirectional generation of structure and properties through a single molecular foundation model. Nat Commun. 2024;15(1):2323.
26. Feng T, et al. Four novel prognostic genes related to prostate Cancer identified using co-expression structure network analysis. Front Genet. 2021;12:584164.
27. Maleki F, et al. Gene set analysis: challenges, opportunities, and future research. Front Genet. 2020;11:654.
28. Liu S, et al. Three differential expression analysis methods for RNA sequencing: limma, EdgeR, DESeq2. JoVE. J Visualized Experiments. 2021;175:62528.
29. Varambally S, et al. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. Cancer Cell. 2005;8(5):393–406.
30. Malhotra S, et al. A tri-marker proliferation index predicts biochemical recurrence after surgery for prostate cancer. PLoS ONE. 2011;6(5):e20293.
31. LaTulippe E, et al. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. Cancer Res. 2002;62(15):4499–506.
32. Cai C, et al. ERG induces androgen receptor-mediated regulation of SOX9 in prostate cancer. J Clin Invest. 2013;123(3):1109–22.
33. Tamura K, et al. SHISA2 enhances the aggressive phenotype in prostate cancer through the regulation of WNT5A expression. Oncol Lett. 2017;14(6):6650–8.
34. Ross-Adams H, et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: a discovery and validation cohort study. EBioMedicine. 2015;2(9):1133–44.
35. Chandran UR, et al. Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. BMC Cancer. 2007;7:64.
36. Grasso CS, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature. 2012;487(7406):239–43.
37. Kumar A, et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat Med. 2016;22(4):369–78.
38. Amirian M, et al. VASA protein and gene expression analysis of human non-obstructive azoospermia and normal by immunohistochemistry, immunocytochemistry, and bioinformatics analysis. Sci Rep. 2022;12(1):17259.
39. Azizi H, Karoii DH, Skutella T. Whole exome sequencing and in silico analysis of human sertoli in patients with non-obstructive azoospermia. Int J Mol Sci. 2022;23(20):12570.
40. Azizi H, Hashemi Karoii D, Skutella T. Clinical management, differential diagnosis, follow-up and biomarkers of infertile men with nonobstructive azoospermia. Translational Androl Urol. 2024;13(2):359–62.
41. Danial Hashemi K, Hossein A. Undifferentiated and differentiated spermatogonial stem cells. In: Leisheng Z, editor. Advances in pluripotent stem cells. Rijeka: IntechOpen; 2023. p. 10.
42. Hashemi Karoii D, Azizi H. A review of protein-protein interaction and signaling pathway of Vimentin in cell regulation, morphology and cell differentiation in normal cells. J Recept Signal Transduction. 2022;42(5):512–20.
43. Hashemi Karoii D, Azizi H. OCT4 protein and gene expression analysis in the differentiation of spermatogonia stem cells into neurons by immunohistochemistry, immunocytochemistry, and bioinformatics analysis. Stem Cell Reviews Rep. 2023;19(6):1828.
44. Hashemi Karoii D, Azizi H. Functions and mechanism of noncoding RNA in regulation and differentiation of male mammalian reproduction. Cell Biochem Funct. 2023;41(7):767–78.
45. Hashemi Karoii D, Azizi H, Skutella T. Altered G-protein transduction protein gene expression in the testis of infertile patients with nonobstructive azoospermia. DNA and Cell Biology. 2023;42(10):617.
46. Hashemi Karoii D, Azizi H, Skutella T. Microarray and in silico analysis of DNA repair genes between human testis of patients with nonobstructive azoospermia and normal cells. Cell Biochem Funct. 2022;40(8):865–79.
47. Karoii DH, Azizi H, Amirian M. Signaling pathways and protein–protein interaction of vimentin in invasive and migration cells: a review. Cell Reprogramming. 2022;24(4):165–74.
48. Niazi Tabar A, et al. Testicular localization and potential function of vimentin positive cells during spermatogonial differentiation stages. Animals. 2022;12(3):268.
49. Karoii DH, Azizi H, Skutella T. Whole transcriptome analysis to identify non-coding RNA regulators and hub genes in sperm of non-obstructive azoospermia by microarray, single-cell RNA sequencing, weighted gene co-expression network analysis, and mRNA-miRNA-lncRNA interaction analysis. BMC Genomics. 2024;25(1):583.
50. Davoodi Nik B, et al. Differential expression of ion channel coding genes in the endometrium of women experiencing recurrent implantation failures. Sci Rep. 2024;14(1):19822.
51. Osterman TJ, Terry M, Miller RS. Improving cancer data interoperability: the promise of the Minimal Common Oncology Data Elements (mCODE) initiative. JCO Clin Cancer Informa. 2020;4:993–1001.

52. Takayama KI, et al. Subtype-specific collaborative transcription factor networks are promoted by OCT4 in the progression of prostate cancer. Nat Commun. 2021;12(1):3766.
53. Obinata D, et al. OCT1-target neural gene PFN2 promotes tumor growth in androgen receptor-negative prostate cancer. Sci Rep. 2022;12(1):6094.
54. Guan M, Jiao Y, Zhou L. Immune infiltration analysis with the CIBERSORT method in lung cancer. Dis Markers. 2022;2022(1):3186427.
55. Pei G, Chen L, Zhang W. WGCNA application to proteomic and metabolomic data analysis, in Methods in enzymology. Amsterdam: Elsevier; 2017.
56. Pereira WJ, et al. Asc-Seurat: analytical single-cell Seurat-based web application. BMC Bioinformatics. 2021;22:1–14.
57. Song H, et al. Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. Nat Commun. 2022;13(1):141.
58. Andreatta M, Carmona SJ. UCell: robust and scalable single-cell gene signature scoring. Comput Struct Biotechnol J. 2021;19:3796–8.
59. Thibodeau SN, et al. Identification of candidate genes for prostate cancer-risk SNPs utilizing a normal prostate tissue eQTL data set. Nat Commun. 2015;6:8653.
60. Chen Y, et al. Metabolic response to androgen deprivation therapy of prostate Cancer. Cancers 2024;16(11):1991.
61. Li LY, et al. DNA repair pathways in Cancer therapy and resistance. Front Pharmacol. 2020;11:629266.
62. Liu X, et al. Low expression of KCNN3 may affect drug resistance in ovarian cancer. Mol Med Rep. 2018;18(2):1377–86.
63. Moreira-Pais A, et al. Tracking prostate carcinogenesis over time through urine proteome profiling in an animal model: an exploratory approach. Int J Mol Sci. 2022;23(14):7560.
64. Varkaris A, et al. The role of HGF/c-Met signaling in prostate cancer progression and c-Met inhibitors in clinical trials. Expert Opin Investig Drugs. 2011;20(12):1677–84.
65. Huo W, Qi F, Wang K. Long non–coding RNA BCYRN1 promotes prostate cancer progression via elevation of HDAC11. Oncol Rep. 2020;44(3):1233–45.
66. Kim SL, et al. Lipocalin 2 negatively regulates cell proliferation and epithelial to mesenchymal transition through changing metabolic gene expression in colorectal cancer. Cancer Sci. 2017;108(11):2176–86.