Submitted: 14 September, 2023 Accepted: 05 December, 2023 Published: 30 April, 2024 DOI:10.22514/jomh.2024.052

ORIGINAL RESEARCH

JOMH Journal of Men's Health

Application of a high-dimensional gene co-expression network to identify metal ion transport-associated epithelial cells with diagnostic function for prostate cancer

Jian Sun[1,†], Xueqi Zhu[2,†], Kai Li[1], Ke Zhang[1], Fei Wang[1,*]

[1]Department of Urology, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Gusu School, Nanjing Medical University, 215000 Suzhou, Jiangsu, China
[2]Department of ICU, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Gusu School, Nanjing Medical University, 215000 Suzhou, Jiangsu, China

***Correspondence**
lu99ky@stu.njmu.edu.cn
(Fei Wang)

† These authors contributed equally.

Abstract

Prostate cancer is a prevalent malignancy and leading cause of male mortality worldwide, thus highlighting the need for precision medicine and a combined single-cell and bulk transcriptome-based diagnostic assessment model. First, we used single-cell RNA sequencing data and the high dimensional weighted gene co-expression network analysis (hdWGCNA) method to identify specific cell types in tumor tissues. We identified higher proportions of epithelial cells and increased intra-tissue heterogeneity in prostate cancer; most of these epithelial cells were closely associated with metal ion transport functions. Lasso regression identified diagnostic-related genes that were incorporated into a sample evaluation model using multiple machine learning algorithms. Thus, we established a diagnostic model that combined immune cells, diagnostic genes and deep learning methods. The Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) models exhibited the highest diagnostic efficacy for Lasso genes. Neural network models, combining immune microenvironment assessment and diagnostic gene expression, also showed good diagnostic efficacy. Our study highlights the potential of machine learning and convolutional neural networks for the diagnostic assessment of prostate cancer, thus providing support for personalized treatment decisions, and identifying areas for future research.

Keywords

Prostate cancer; Diagnosis; Machine learning

1. Introduction

Over recent years, the incidence of prostate cancer has shown an increasing trend and has gradually become an important disease that threatens the health of men [1]. Diagnosis serves as a crucial indicator for determining patient treatment and management strategies, and plays a pivotal role in improving personalized medicine and patient survival rates. However, traditional methods for prostate cancer diagnosis have certain limitations as they rely solely on clinical and pathological features, as well as conventional statistical models, without comprehensive consideration of multivariable and complex relationships [2, 3].

Over recent years, the rapid development of artificial intelligence techniques, including machine learning and deep learning, has provided new opportunities and challenges for the diagnosis of prostate cancer [4–6]. Machine learning enables learning from large-scale data, while deep learning, building upon this foundation, effectively handles multidimensional arrays, such as convolutional neural network algorithms [7]. These technologies have achieved significant advancements in various fields. However, to date, their application in the field of prostate cancer diagnosis remains unexplored [8].

Metal ions are key components of many biochemical processes within cells. Cells regulate the influx and efflux of metal ions through transport proteins, ion channels, and ion pumps to maintain normal ion levels and homeostasis [9]. The transport of metal ions plays important roles in cellular physiology and disease. Previous research has shown that cancer cells may overexpress specific metal ion transport proteins, leading to the accumulation of these metal ions within cells, thus influencing cell signaling and metabolism [10]. In addition, abnormal metal ion transport may disrupt DNA damage repair mechanisms, thereby promoting the formation and progression of tumors. However, the potential mechanisms of metal ion transport associated with the occurrence and development of prostate cancer have yet to be elucidated and require further research and exploration [11].

The aim of this study was to explore a machine learning and convolutional neural network-based approach for the diagnosis of prostate cancer and evaluate its feasibility and efficacy in clinical practice [12]. We utilized a single-cell RNA sequencing dataset of prostate cancer tissues to identify specific cell

types within tumor tissues [13]. In addition, we used the high dimensional weighted gene co-expression network analysis (hdWGCNA) method to calculate gene modules associated with the identified cells [14]. By integrating conventional transcriptome sequencing data and utilizing Lasso regression, we identified diagnostic-related genes and constructed a sample evaluation model using multiple machine learning algorithms [15]. Furthermore, by evaluating the immune microenvironment of the samples, a diagnostic model integrating immune cells and genes was developed by deep learning methods. The specific relationship between our diagnostic model and patient prognosis was determined by Cox regression [16–18]. By performing metascape gene enrichment, we were able to identify this class of cells as being associated with transmembrane transport, particularly the transport of metal ions. By comparing the performance differences between our model and traditional methods, we aimed to demonstrate the advantages of machine learning and convolutional neural networks in the diagnosis of prostate cancer and provide robust support for future personalized treatment decisions.

2. Materials and methods

2.1 Data sources

The scRNA-seq dataset GSE185344 [19] (including samples of 7 primary tumors undergoing prostatectomy) and GSE117403 [20] (including 16 samples of normal human prostate), were acquired from the Gene Expression Omnibus (GEO); these databases are publicly available and can be accessed through the GEO website (https://www.ncbi.nlm.nih.gov/geo/). To complement our analysis, we obtained bulk RNA sequencing data from prostate adenocarcinoma (PRAD) from the Cancer Genome Atlas Program (TCGA, https://portal.gdc.cancer.gov/) consisting of 55 normal samples and 496 tumor samples. In addition, we obtained array data that was used to build the Cox regression model.

2.2 Data processing

The scRNA-seq data underwent preprocessing using the Seurat package (v4.2.1) [21] in the R programming language to ensure the quality of subsequent analyses. Initially, we excluded cells exhibiting mitochondrial gene expression exceeding 40%, those with a low number of detected genes (<200), and those with an excessive number of detected genes (>2500). The Harmony (v0.1.1) package [22] was used to remove batch effects between merged samples. The top 2000 highly variable genes were selected for principal component analysis (PCA). Subsequently, the first 15 principal components (PCs) were chosen for further analysis. Then, the "FindClusters" function was employed to cluster and visualize the PCs using a Uniform Manifold Approximation and Projection (UMAP) approach. Next, the "FindAllMarkers" function was utilized to identify differentially expressed genes within each cluster. The "DoHeatmap" function was used to plot a heatmap for selected genes. Cell annotation was performed by the "singleR" (2.0.0v) package [23], which compares gene expression profiles in the reference dataset to obtain a predicted cell type label for each cell in the single-cell sequencing dataset based on the closest match found in the reference dataset.

2.3 Intercellular communication analysis and pseudotime analysis

Intercellular communication analysis was performed by the CellChat (v1.6.1) package [24], an R package that was developed to analyze cell-to-cell interactions in single-cell transcriptome data. This software clusters cells, infers interactions between different units, and constructs cell-to-cell communication networks. Using Gene Set Enrichment Analysis (GSEA), this software annotates functional modules for each clustering module. CellChat is a reliable tool for studying cell interactions and the mechanisms of complex diseases. We chose "Secreted Signaling pathway" to explain the potential mechanism. To perform pseudotime analysis, we used the monocle (v2.26.0) package [25] in R to construct a cell ordering trajectory, reduce the dimensions of the gene expression data, create a minimum spanning tree, order cells along the trajectory, and conduct differential gene expression analysis.

2.4 High-dimensional weighted co-expression network analysis (hdWGCNA)

For scRNA-seq data, we applied the hdWGCNA (v0.1.1.9010) package to cluster genes [26]. We filtered out genes expressed in less than 5% of cells and used the remaining genes to construct the hdWGNCA object. The co-expression network was constructed using a soft power value of 9 for subsequent analysis. Module eigengenes (MEs) provided a representation of the gene expression profile within each co-expression module. We associated each of the MEs with a Seurat cluster in a dotplot to identify the connection between them. The top 50 genes with the highest correlation in the module were selected for downstream analysis.

2.5 Gene enrichment

To investigate the biological functions and pathways associated with common genes, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses using Metascape (`https://metascape.org/gp/index.html#/main/step1`), a web-based portal that provides a comprehensive resource to annotate and analyze gene lists for experimental biologists [27]. The screening conditions were set to a minimum overlap of 3 and a minimum enrichment of 1.5. Filter pathways with p-values < 0.01 were used to plot bar charts and network interactions.

2.6 Lasso regression and machine learning

Lasso regression is used to perform variable selection by shrinking the regression coefficients associated with the least important variables to zero, thus resulting in a parsimonious model. This strategy can enhance our understanding of complex biological processes and disease mechanisms. To construct a preliminary prediction model, we used bulk sequencing data from the TCGA-PRAD and conducted two-step analysis using the glmnet package. First, we used

logistic regression analysis to identify genes with statistical significance ($p < 0.05$). Then, we applied Lasso regression to further determine the most valuable variables. The cv.glmnet function was then utilized to identify highly valuable genes.

2.7 Machine learning

To further refine our model, we used machine learning methods to improve the predictive accuracy. The mlr3verse package was utilized to construct various classifiers, including Linear Discriminant Analysis (LDA), Logistic Regression (Log Reg), Random Forest, Support Vector Machine (SVM), Naive Bayes, K-nearest neighbor (KNN), rPart, Random Forest, among others. We used a 5-fold cross-validation method to partition the dataset. Finally, receiver operating characteristic (ROC) curve analysis was applied to evaluate the diagnostic accuracy and optimal threshold value of these models. For each model, we also calculated the area under the ROC curve (AUC), sensitivity and specificity.

2.8 Geneset immune infiltration evaluation

To analyze immune-typing, we conducted ssGSEA analysis using the GSVA package (v1.46.0) [28] to determine the heterogeneity of Lasso genes identified between tumor and normal tissues at the bulk level. Subsequently, we utilized the IOBR package (v0.99.0) [29] to calculate the correlation between Lasso genes and immune cells. The IOBR package provides a framework that allows us to calculate the correlation between Lasso genes and immune cells [29]. By integrating results, it is possible to gain insights into the immune-typing of tumors and normal tissues and evaluate the role of Lasso genes in the immune response. We integrated the results by calculating the gene-set score using methods such as PCA, z-score, and ssGSEA; in addition, MCPcounter was applied to calculate detailed taxonomic scores for epidemic cell taxa.

2.9 Non-negative matrix factorization

To construct disease subtypes using the MEs module gene, we used the NMF package (v0.25) [30] in R to perform non-negative matrix factorization. This strategy can improve the resolution of clustering or trajectory analysis in single-cell transcriptomes, thereby accelerating studies of complex biological processes and disease mechanisms. Firstly, we reprocessed the gene expression data to ensure accurate results. Then, the "NMF" function was used to compute the NMF with specific settings (rank: 5; nrun: 100).

2.10 Deep learning

Finally, deep learning was utilized to develop a novel diagnostic prediction model that integrated the immune microenvironment and Lasso gene expression levels to predict the prognosis of patients at multiple levels. The Keras package (v2.11.0.9000) [31] was utilized to integrate patient information, Lasso gene expression levels, deconvolution MCPcounter scores, and immune cell correlations into multi-dimensional arrays, which were then used to train and test the deep learning model utilizing convolution algorithms.

3. Results

3.1 Annotation and identification of subgroups in scRNA-seq data

Following quality control, a total of 35,673 cells were included in this study from GSE185344 and GSE117403 (Fig. 1A). Harmony was utilized to remove batch effects and integrate the two cell datasets. PCA analysis (Fig. 1B) showed that the sample heterogeneity caused by different batches was essentially eliminated. Seurat was utilized for further dimensionality reduction and clustering, and singleR was utilized to identify specific cell types and proportions, including T cells, epithelial cells, monocytes, smooth muscle cells, endothelial cells, natural killer cells (NK cells), common-myeloid progenitors (CMP), chondrocytes, macrophages, and B cells (Fig. 1C,D). Analysis showed that the proportion of epithelial cells was significantly increased in prostate cancer samples. The identified cell types and their expression levels were used for differential analysis and heat map visualization (Fig. 1E).

Furthermore, to investigate the changes in epithelial cells, Seurat was used to isolate epithelial cells and perform secondary dimensionality reduction clustering. In the second clustering, 19 epithelial cell-related clusters were identified, and their proportions were calculated (Fig. 2A,B), as well as their proportions in healthy and cancerous tissues (Fig. 2C,D). Clusters 0, 1, 2, 4, 6, 7, 9, 10 and 18 were exclusive to prostate cancer tissue.

3.2 Intercellular communication and pseudotime analysis

Intercellular communication refers to the process of information transmission between cells and between cells and the environment and the integration of signals in the cell; this plays a key role in maintaining the normal function and physiological processes of multicellular organisms [32, 33]. Intracellular communication is able to regulate the growth, differentiation and function of cells to guarantee the correct development and normal functionality of tissues and organs [24]. Intercellular communication abnormalities are closely related to the occurrence and progression of a variety of important diseases [34]. Therefore, it is of great significance to understand the mechanism and regulation of intercellular communication to reveal the molecular mechanisms responsible for disease occurrence and develop new therapeutic methods. To identify the interactions between these tumor-specific epithelial cell clusters and other cells, we utilized the Intercellular Communication Analysis method to identify ligand-receptor pairs between these cell clusters. Fig. 3A shows the total number and strength of interactions between these epithelial cells and other cells. We found that tumor-specific epithelial cells had a strong effect on endothelial cells, monocytes, and B cells. Fig. 3B,C display the main ligand-receptor pairs when epithelial cells were used as the source and target, respectively. When epithelial cells were used as the target, there were few statistically significant ligand-receptor pairs (Midkine (*MDK*)-Nucleolin (*NCL*), α-Methyl-p-tyrosine (*AMPT*)-Insulin receptor (*INSR*), Pleiotrophin (*PTN*)-NCL). When these were used as the ligand, we found that they had higher levels of activity. The

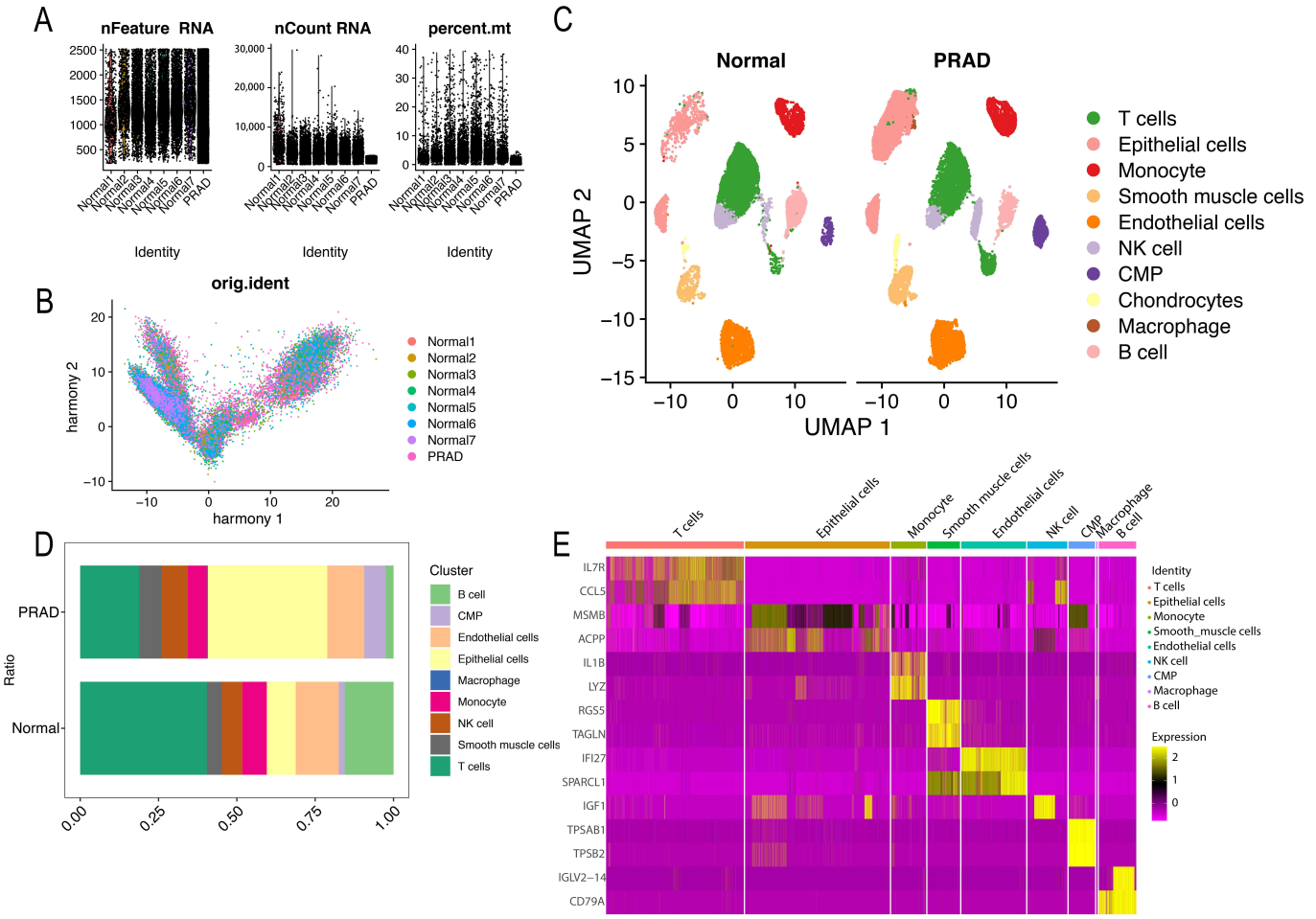

FIGURE 1. Overview of single-cell RNA sequencing analysis. (A) Flowchart of data quality control and analysis. (B) Principal component analysis (PCA) plot showing batch effects removal. (C) t-SNE plot showing cell clustering by cell type. (D) Cell proportions of different subpopulations of cells in PRAD tissue and normal tissue. (E) Differential gene expression analysis and heatmap visualization of specific cell types in prostate cancer samples compared to healthy tissues. NK cells: natural killer cells; CMP: common-myeloid progenitors; PRAD: prostate cancer; *IL70*: Interleukin 70; *CCL5*: C-C Motif Chemokine Ligand 5; *MSMB*: Microseminoprotein Beta; *ACPP*: Prostatic Acid Phosphatase; *IL1B*: Interleukin1B; *LYZ*: Lysozyme; *RGS5*: Regulator Of G Protein Signaling 5; *TAGLN*: Transgelin; *IFI27*: Interferon Alpha Inducible Protein 27; *SPARCL1*: SPARC Like 1; *IGF1*: Insulin Like Growth Factor 1; *TPSAB1*: Tryptase Alpha/Beta 1; *TPSB2*: Tryptase Beta 2; *IGLV2-14*: Immunoglobulin Lambda Variable 2-14; *CD79A*: CD79a Molecule.

relationship between macrophage migration inhibitory factor (*MIF*) and *MDK* as the ligand-receptor pair played a dominant role. Most ligand-receptor pairs were present in both tumor and healthy cells, but overall, the communication effect of tumor cells stronger (Fig. 3D).

Next, pseudotime analysis was used to construct the developmental trajectory of all cells. Six developmental branching points were identified, as shown in Fig. 4A, which displays the relationship between Seurat clusters and the trajectory. Some prostate cancer-specific epithelial cell clusters, such as cluster 0, showed a specific distribution along the trajectory. Pseudotime analysis was then used to analyze the developmental direction of the trajectory (Fig. 4B). The Beam algorithm was used to cluster pseudotime analysis results and generate differential genes before and after developmental branching points (Fig. 4C).

3.3 hdWGCNA identification of tumor-associated epithelial clusters and functional enrichment

Next, the hdWGCNA algorithm was used to analyze gene differential expression in the epithelial cell clusters. The appropriate soft power for a weighted average co-expression network was selected to establish the co-expression network (Fig. 5A,B) in order to reflect the inherent properties of the data. Eight module Eigengenes (MEs) were identified, except for the gray module (brown, blue, red, yellow, green, black, turquoise, pink), and their interrelations are shown in Fig. 5C. Notably, the pink module had a significant association with the turquoise module, while the green module was associated with the black module. The distribution of the 8 MEs in epithelial cells is demonstrated in Fig. 5D, and the expression of MEs in each Seurat cluster is quantified in Fig. 5E. The pink, turquoise, black, green and yellow modules had a higher proportion of

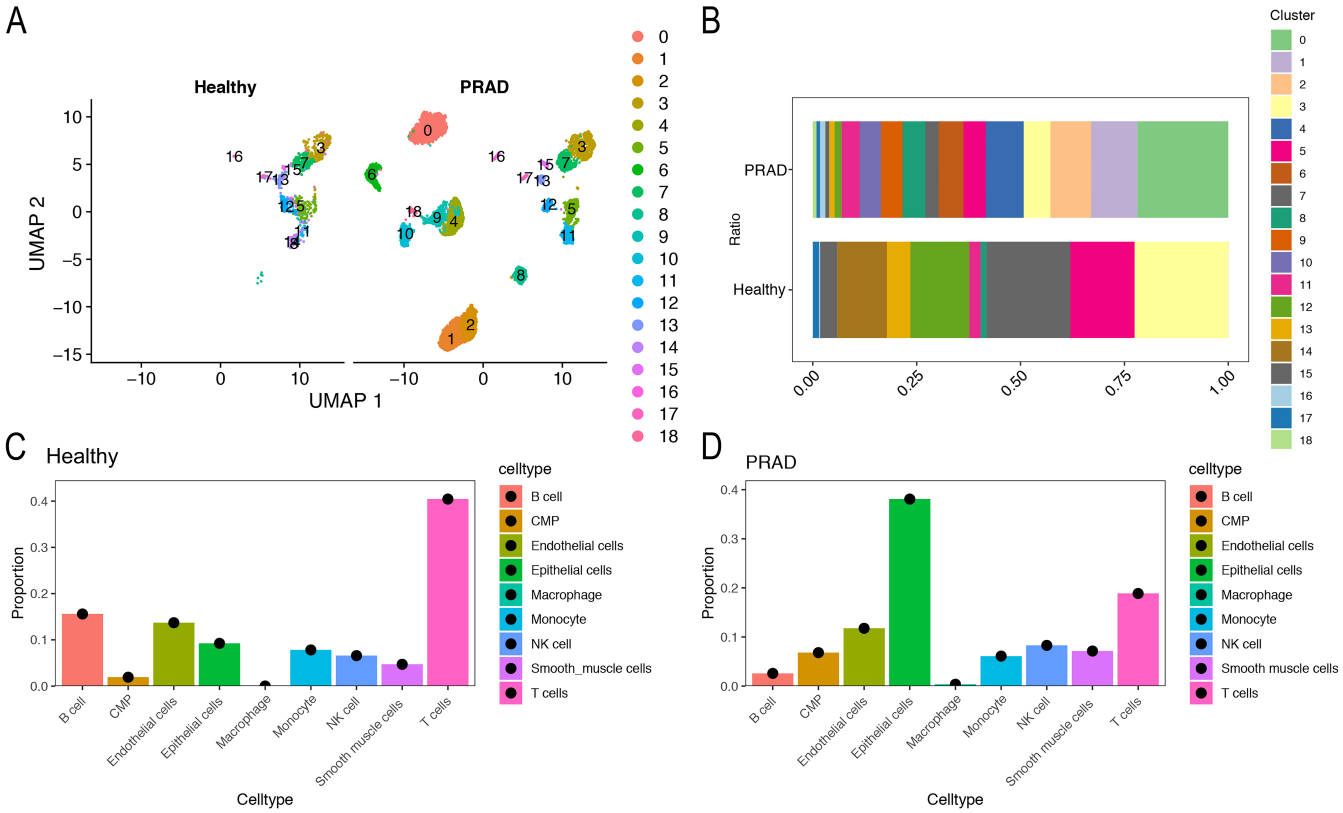

FIGURE 2. Epithelial cell-related clustering analysis. (A) t-SNE plot showing the clustering of epithelial cells. (B) Secondary dimensionality reduction clustering of epithelial cells showing 19 epithelial cell-related clusters. (C) Comparison of the proportion of epithelial cell-related clusters between healthy and cancerous tissues. (D) Proportions of epithelial cell-related clusters in healthy and PRAD tissues. NK cells: natural killer cells; CMP: common-myeloid progenitors; PRAD: prostate cancer.

gene expression and average expression level in the tumor-specific epithelial cell clusters. The top 50 genes with the highest correlation per module were selected (**Supplementary Table 1**).

To clarify the functionality of the genes identified in each module, we used the Metascape database for Kyoto Encyclopedia of Genes and Genomes (KEGG) as well as GO functional enrichment (Fig. 6A,B). Analysis showed that the genes identified by hdWGCNA were involved in monoatomic cation transport, copper homeostasis, transition metal ion transport and other processes of metal ion transport across membranes. In addition to this, we identified close links with fatty acid metabolism, cholesterol metabolism and cellular respiration. Other genes were involved in fatty acid metabolism, cholesterol metabolism, cellular respiration and other physiological processes.

3.4 The use of machine learning to identify potential diagnostic biomarkers for prostate cancer

Logistic regression and Least Absolute Shrinkage and Selection Operator (LASSO) regression algorithms were applied to screen for candidate genes, as demonstrated in Fig. 7A,B. LASSO regression identified 33 potential biomarkers (**Supplementary Table 2**). Machine learning based on the mlr3vers package was performed, and the results are presented in Fig. 7C. The Linear Discriminant Analysis (LDA), Logistic Regression (Log Reg), Random Forest, and Support Vector Machine (SVM) models exhibited high sensitivity in the training set. In the testing set, the LDA and SVM models demonstrated the highest diagnostic sensitivity and can be used effectively for the diagnosis of prostate cancer (Fig. 7D,E).

3.5 Immune infiltration assessment of prostate cancer samples

The results of single-sample Gene Set Enrichment Analysis (ssGSEA) assessment for prostate cancer samples are shown in Fig. 8A, with LASSO gene scores significantly higher in tumor tissues than in normal tissues. Next, Non-Negative Matrix Factorization (NMF) clustering was performed based on the expression profile of Lasso genes, and the NMF rank survey plot is presented in Fig. 8B, with 5 selected as the appropriate NMF clustering parameter. Ultimately, we identified 5 subtypes of prostate cancer tumor cells, as demonstrated in Fig. 8C, along with their relationships with immune cells. The correlation between immune cells and Lasso genes is illustrated in Fig. 8D. We found that the expression levels of Aryl hydrocarbon receptor (*AHR*) were positively correlated with fibroblasts ($r > 0.5$, $p < 0.001$), endothelial cells ($r > 0.5$, $p < 0.001$), neutrophils ($r > 0.5$, $p < 0.001$), myeloid dendritic cells ($r > 0.5$, $p < 0.001$), monocytic lineage ($r > 0.5$, $p < 0.001$), cytotoxic lymphocytes ($r > 0.5$, $p < 0.001$), and T cells ($r > 0.5$, $p < 0.001$). The expression levels

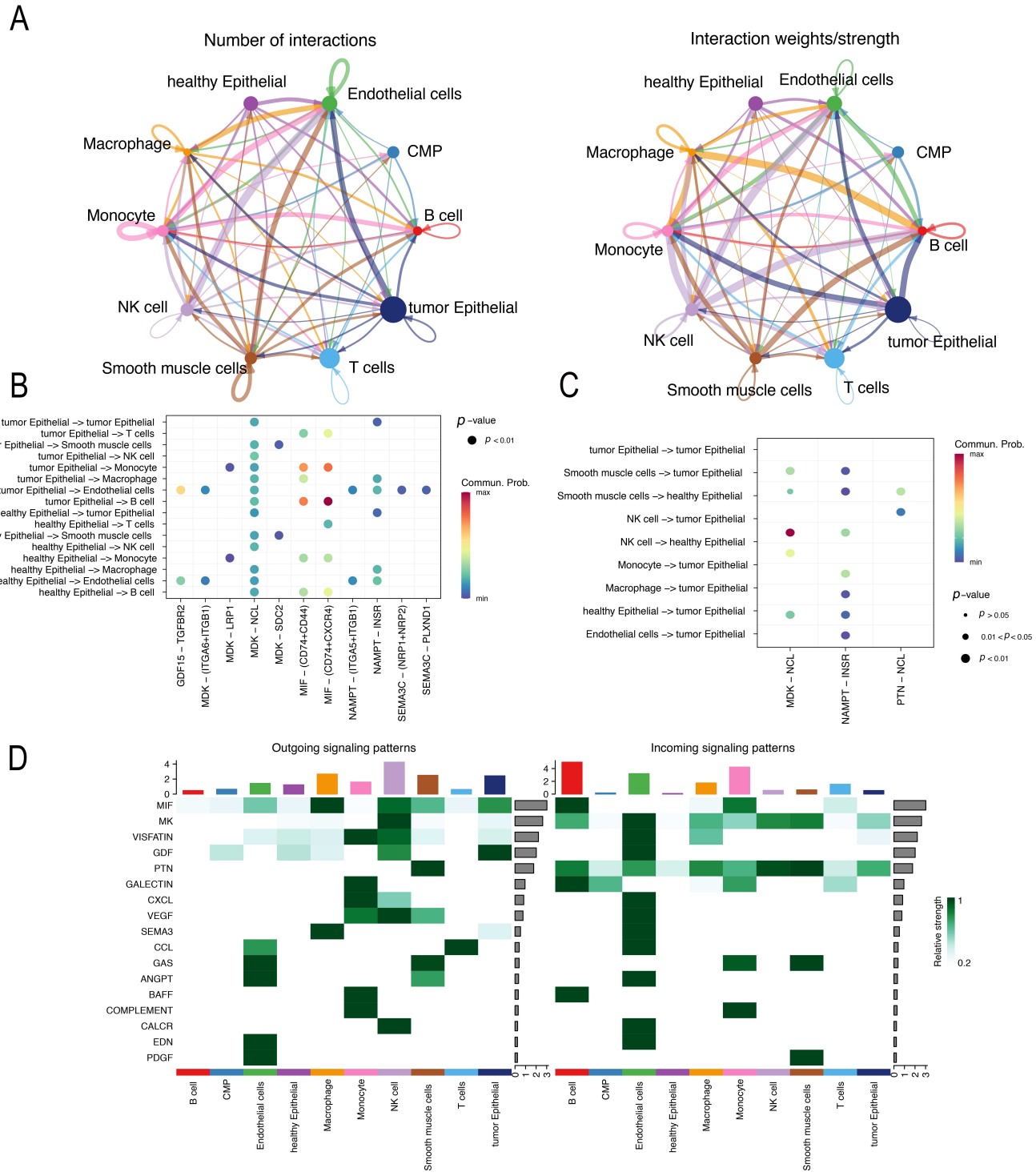

FIGURE 3. Intercellular communication analysis of tumor-specific epithelial cells. (A) Network diagram showing the total number and strength of interactions between epithelial cells and surrounding cells. (B,C) Heatmaps of ligand-receptor pairs when epithelial cells were used as the source (B) and target (C). (D) Comparison of the proportion of ligand-receptor pairs between healthy and cancerous tissues. Most ligand-receptor pairs were present in both tumor and healthy cells; however, the communication effect of tumor cells was stronger. NK cells: natural killer cells, CMP: common-myeloid progenitors; *MIF*: migration inhibitory factor; *MK*: Mevalonate kinase; *GDF*: Growth Differentiation Factor; *MDK*: Midkine; *PTN*: Pleiotrophin; *CXCL*: C-X-C Motif Chemokine Ligand; *VEGF*: Vascular Endothelial Growth Factor; *SEMA3*: Semaphorin 3; *CCL*: C-C Motif Chemokine Ligand, *GAS*: Growth Arrest-specific Transcripts; *ANGPT*: Angiopoietin 2; *BAFF*: B-Cell-Activating Factor, *CALCR*: Calcitonin Receptor; *EDN*: Endothelin; *PDGF*: Platelet Derived Growth Factor.

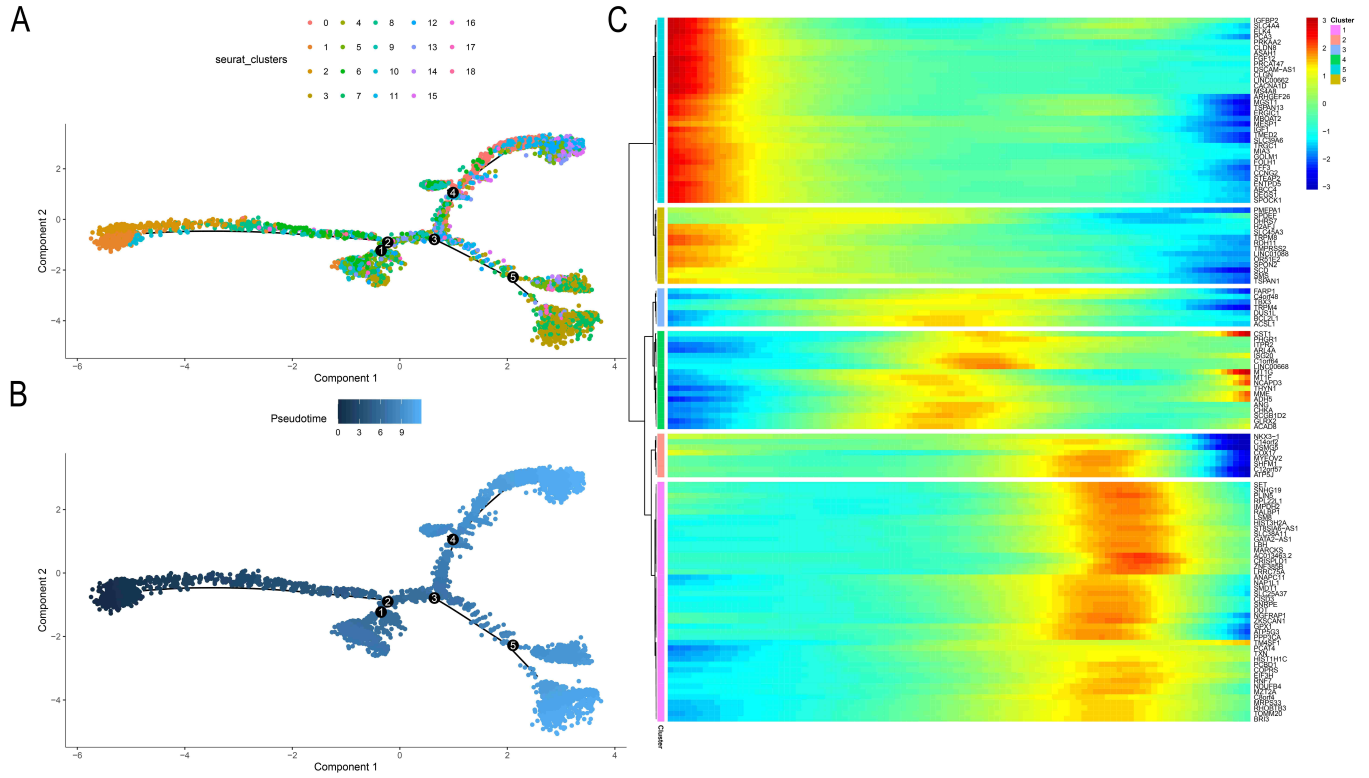

FIGURE 4. Pseudotime analysis of developmental trajectory. (A) Visualization of the developmental trajectory of all cells. (B) Plot showing the developmental direction of the trajectory, generated using pseudotime analysis. (C) Differential gene expression analysis of genes before and after developmental branching points, as generated by the the BEAM algorithm.

of serine/threonine kinase 38 like (*STK38L*) were positively correlated with fibroblasts ($r > 0.5$, $p < 0.001$), endothelial cells ($r > 0.5$, $p < 0.001$), neutrophils ($r > 0.5$, $p < 0.001$), monocytic lineage ($r > 0.5$, $p < 0.001$), NK cells ($r > 0.25$, $p < 0.001$), cytotoxic lymphocytes ($r > 0.25$, $p < 0.001$), and T cells ($r > 0.25$, $p < 0.001$). The expression levels of transcription elongation factor A like 1 (*TCEAL1*) were positively correlated with fibroblasts ($r > 0.5$, $p < 0.001$), endothelial cells ($r > 0.5$, $p < 0.001$), neutrophils ($r > 0.25$, $p < 0.001$), monocytic lineage ($r > 0.25$, $p < 0.001$), cytotoxic lymphocytes ($r > 0.25$, $p < 0.001$), CD8 T cells ($r > 0.25$, $p < 0.001$), and T cells ($r > 0.25$, $p < 0.001$). Fig. 8E shows a comparison of the content of 10 types of immune cells in each subtype, among which T cells, monocytic lineage, and CD8+ T cells exhibited the most significant statistical differences in proportion.

3.6 Construction of a convolutional neural network model with immune scores and gene expression

The Keras package was used for preliminary processing. The calculated ratio of immune score and gene expression was used as the learning data for the neural network model, after constructing a high-dimensional array for deep learning. The results of deep learning are presented in Fig. 8F. In the training set, the ROC curve had an AUC of 0.843; this compared to an AUC of 0.834 in the testing set, thus demonstrating the strong diagnostic performance of the deep learning model (Fig. 8G).

4. Discussion

Prostate cancer is a highly prevalent form of cancer, and the diagnosis and treatment of this condition represents a major challenge to the medical community. Single-cell RNA sequencing technology has emerged as an efficient tool for investigating the molecular mechanisms underlying the formation and development of cancer [35, 36]. The rapid development of single-cell sequencing technology has provided a new approach for investigating the prostate cancer [37]. This method previously demonstrated tumor heterogeneity in prostate cancer, and comprehensive analysis of the tumor microenvironment demonstrated the complex cell composition of prostate cancer, explored potential interactions between tumor cells and other cells, and provided a deeper understanding of the composition of the tumor microenvironment [19]. This has provided more reliable data for clinical treatment and provided a theoretical foundation for further research, and provided guidance for precise personalized treatment [38].

In the present study, we utilized single-cell RNA sequencing to investigate changes in specific cell types in samples of prostate cancer, including T cells, epithelial cells, monocytes, smooth muscle cells, endothelial cells, NK cells, CMP, chondrocytes, macrophages, and B cells. We used advanced analytical methods, including Harmony and Seurat, to eliminate batch effects, cluster cells, and identify unique cell types. Our analysis showed that the proportion of epithelial cells were significantly increased in samples of prostate cancer. Then, we performed secondary clustering analysis on epithelial cells and identified 19 epithelial cell-related clusters that were exclusive

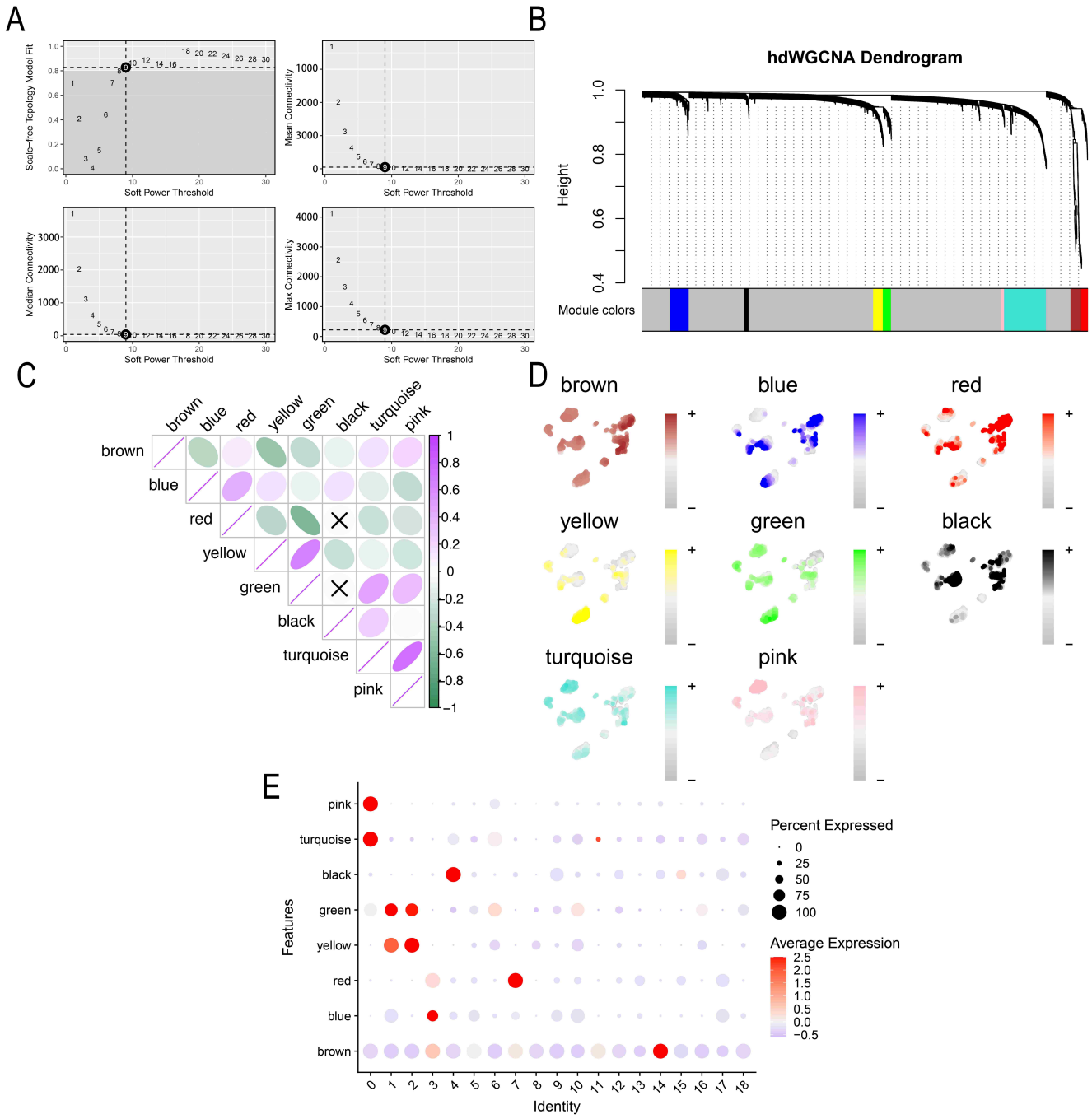

FIGURE 5. hdWGCNA analysis of epithelial cell clusters. (A) Soft threshold selection. (B) Construction of a co-expression network and the generation of a clustering tree. (C) Heatmap showing module-trait relationships between modules and epithelial cell clusters. (D) The proportion of eight modules in epithelial cells and the expression of 8 modules in each Seurat cluster. (E) Quantification of the expression of MEs in each Seurat cluster. hdWGCNA: high dimensional weighted gene co-expression network analysis.

to prostate cancer tissue. These findings concur with previous studies, which suggested that the histological changes that occur first during the malignant transformation of normal prostate tissue involved abnormal epithelial hyperplasia, and that the proliferation of tumor epithelial cells is usually associated with abnormal immune function and abnormal signal transduction changes [39, 40]. Previous research also showed that abnormal proliferation of the epithelial cells of the prostate can often lead to prostate cancer [37]. The higher the proportion of epithelial cells, the higher the likelihood of cancerous transformation and the more severe the condition. This may be related to the proliferative capacity of cancer cells, along with the depth of infiltration, and their tendency to metastasize [41–43]. Therefore, an increase in the proportion of epithelial cells is of great significance for the prognostic assessment of patients with prostate cancer. Base on this, our findings are consistent with previous single-cell RNA sequencing studies on the heterogeneity and complexity of prostate cancer, thus highlighting the importance of characterizing unique cell clusters when attempting to understand disease progression and

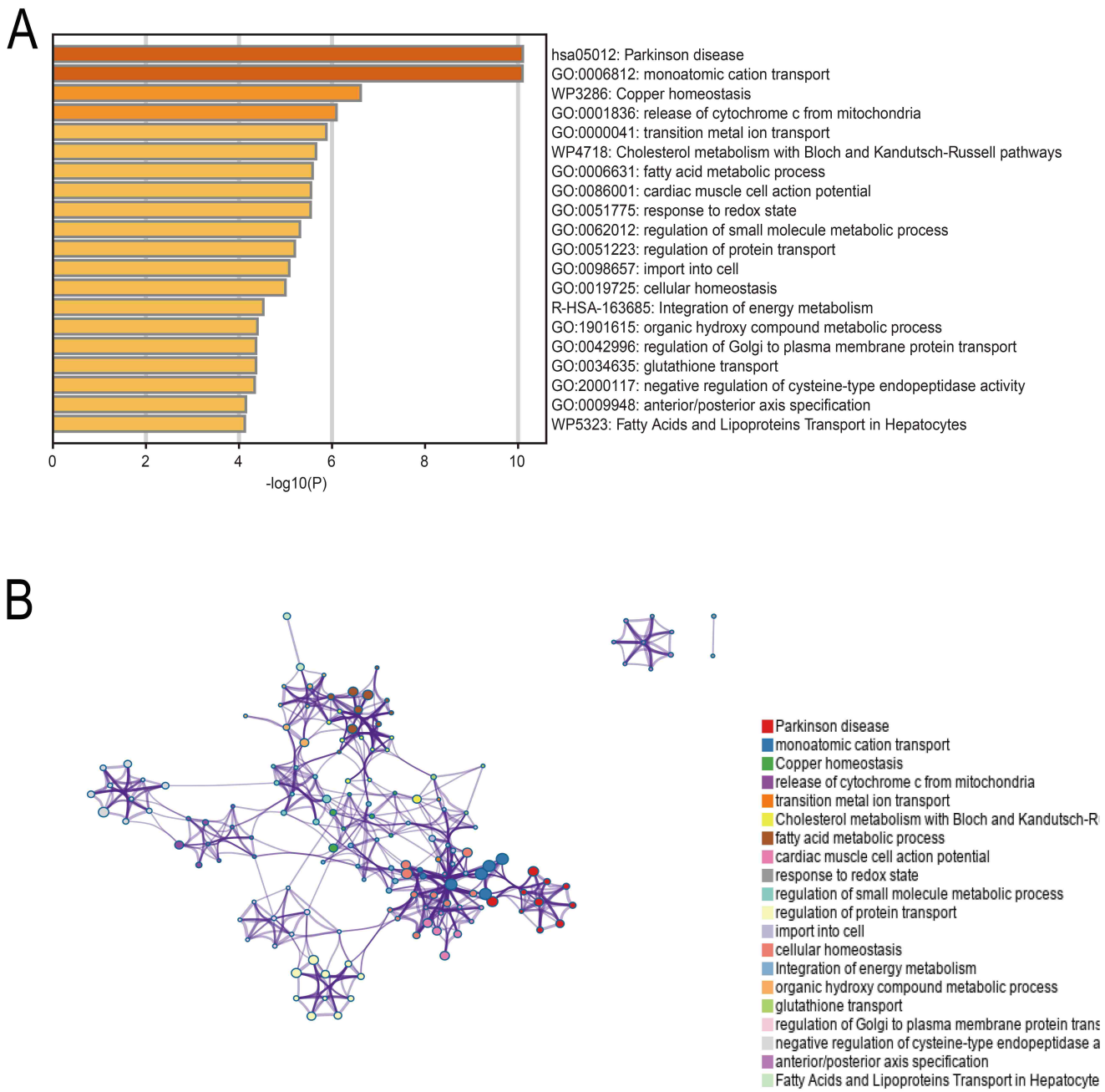

FIGURE 6. Functional enrichment analysis of genes identified by hdWGCNA. (A) Results of pathway enrichment analysis using the Metascape database. (B) Functional role network. GO: gene ontology; has: homo sapiens; WP: WikiPathways.

identify biomarkers for diagnosis and prognosis.

Matrix metalloproteinases (MMPs) are a series of structurally related zinc-dependent endopeptidases that can degrade almost all components of the extracellular matrix and basement membrane [44]. These peptidases play important roles in embryonic development, wound healing, angiogenesis, atherosclerosis, cancer, and tissue ulceration. Thus far, more than 30 MMPs have been found to be involved in and play important roles in the occurrence and biochemical recurrence of prostate cancer in patients [45]. *MMP7* is a member of the matrix metalloproteinase family, which can regulate the signal transduction pathways involved in cell growth and angiogenesis, promote cancer cell infiltration and angiogenesis, degrade the extracellular matrix and basement membrane, and promote bone resorption. *MMP7* also plays an important role in promoting the occurrence and invasive development of prostate cancer [43, 46]. Therefore, MMPs have potential clinical applications for the diagnosis and treatment of prostate cancer. We used intercellular communication analysis to identify ligand-receptor pairs between cells and found that these tumor-specific epithelial cells had a strong influence on other cells such as endothelial cells, monocytes and B cells. Tumor cells as a class of metabolically active abnormal cells, they are mostly involved in cellular communication processes in the form of secretion to achieve tumor progression. This is process is also necessary for the formation of cancerous lesions, especially for epithelial cells [47]. In a previous study, Wu *et al.* [48] showed that

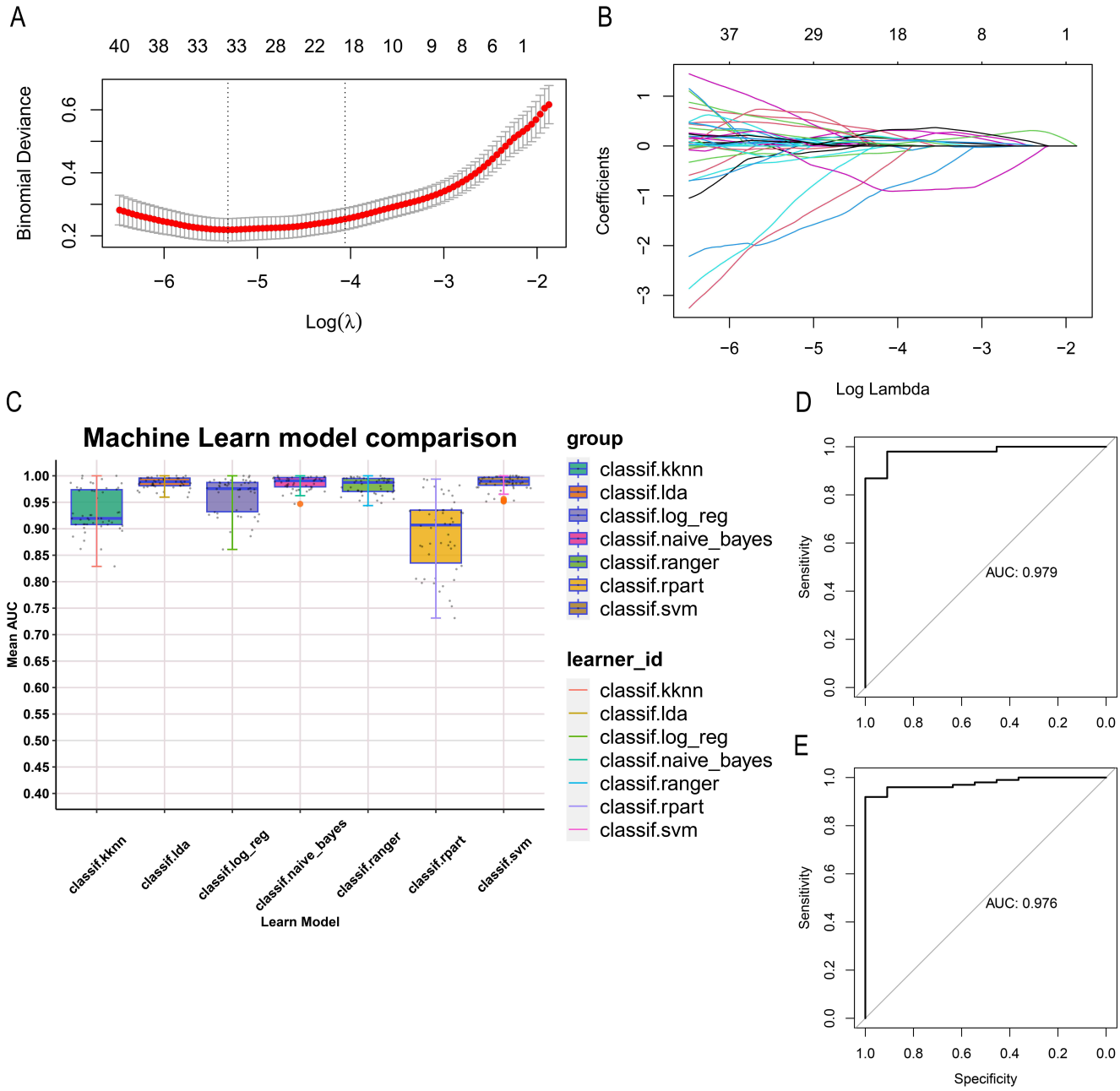

FIGURE 7. Machine learning-based screening of potential biomarkers for the diagnosis of prostate cancer. (A) Path diagram depicting the trajectory of hdWGCNA module genes based on LASSO coefficients. Each curve shows the change in value of each hub gene, in which the vertical axis represents the gene values, the lower horizontal axis represents log(λ), and the upper horizontal axis represents the number of non-zero hub genes included in the model at a particular time. (B) The LASSO regression cross-validation curve showing the optimal λ values selected through 10-fold cross-validation. (C) The machine learning models in training sets, using the mlr3vers package. ROC curves and AUC values for each model are shown. (D,E) The diagnostic sensitivity and specificity of the LDA (D) and SVM (E) models were evaluated in the testing sets. AUC: the area under the ROC curve.

the epithelial cells of prostate cancer exhibited tight cellular communication with tumor-associated fibroblasts. In our analysis, epithelial cells were shown to be more inclined to act as active secretors of signaling factors than as target cells, thus confirming the concept that epithelial cells are the starting point for tumorigenesis and development. In contrast, the main products secreted by tumorigenic epithelial cells in our analysis were *MIF*, Growth Differentiation Factor (*GDF*) and Midkine (*MDK*).

MIF is a protein produced by mononuclear cells and other immune cells, and plays a crucial role in regulating immune responses and inflammatory processes [49]. In previous studies, investigators observed that *MIF* usually shows a tendency to be highly expressed in patients with high levels of metastatic activity [50]. Other studies showed that *MIF* was overexpressed in trend-resistant prostate cancer and induced

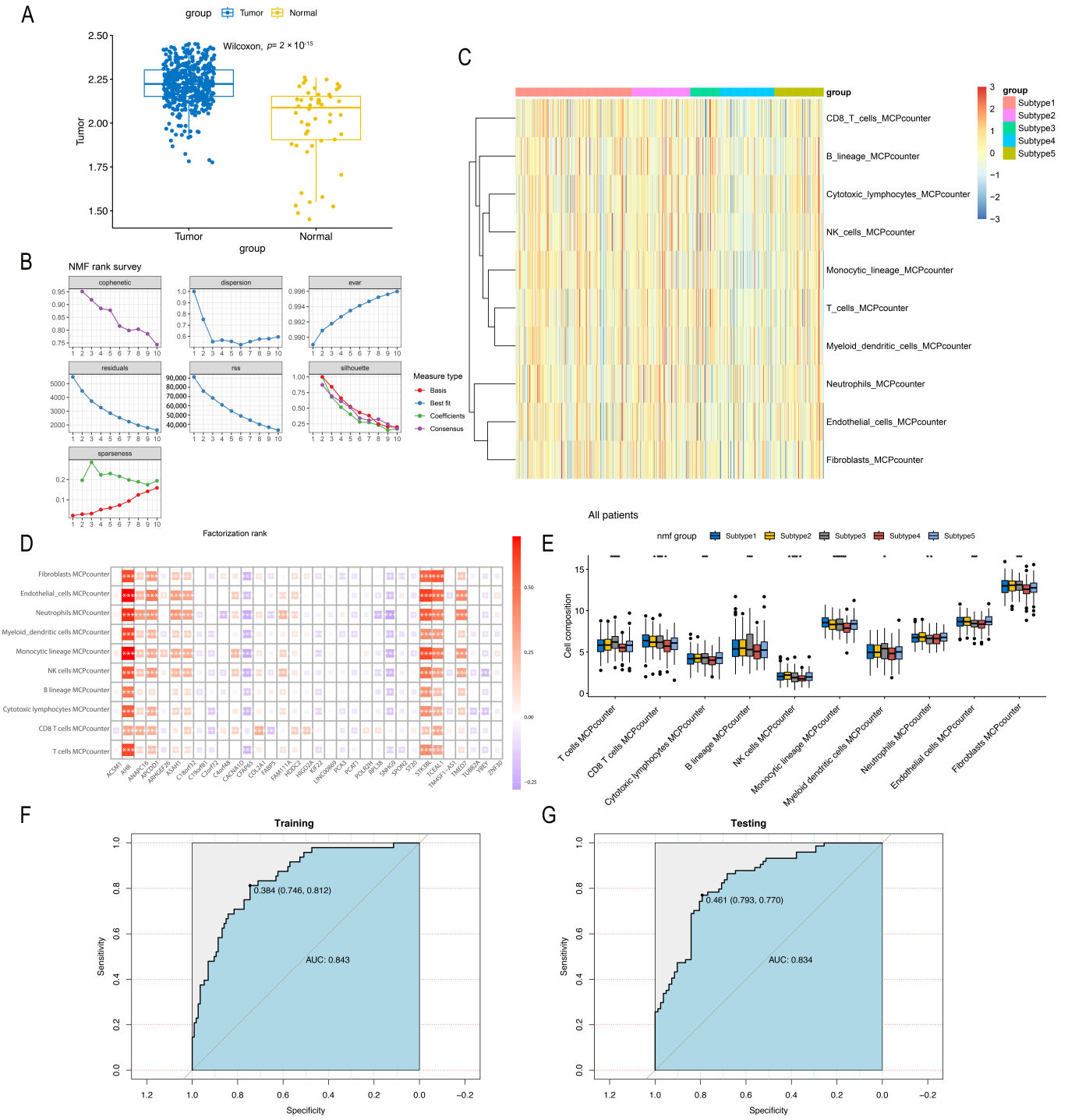

FIGURE 8. Subtype classification and immune cell correlation analysis in prostate cancer and deep learning. (A) GSVA assessment of prostate cancer samples comparing Lasso gene scores between tumor and normal tissues. (B) NMF rank survey plot. (C) Subtype classification of prostate cancer tumor cells based on the expression of Lasso genes and their relationship with immune cells. (D) Correlation network between immune cells and Lasso genes. (E) Comparison of the content of 10 types of immune cells in each subtype. (F,G) The diagnostic sensitivity and specificity of convolutional neural network models in the training (F) and testing (G) datasets. AUC: the area under the ROC curve.

the dysregulation of cell cycle gene expression by regulating *CXCR7* and further activating the *AKT* signaling pathway [51]. In contrast, the *GDF* gene family enhances tumor pathogenicity through cellular checkpoint escape mechanisms, induces epithelial-mesenchymal transition, regulates extracellular matrix synthesis and degradation, apoptosis, cell migration, and the activation of cell signaling pathways, thereby promoting tumor growth, spread, and metastasis [52, 53]. In the case of *MDK*, current studies have demonstrated an association between neuroendocrine differentiation and prostate cancer, and that this process is inextricably linked to immune regulation [54, 55]. Prostate cancer is a complex disease and multiple

etiological factors play roles in disease progression. Genetic factors, age, nutrition, and lifestyle habits have all been identified as potential risk factors for prostate cancer. Advanced age is a well-established risk factor, and the incidence of prostate cancer is known to increase with age [56]. In addition, a high intake of animal fats can increase the risk of prostate cancer by increasing testosterone levels in the body [57]. Previous studies have found that smoking and a lack of exercise are also associated with the development of prostate cancer [58]. Moreover, neuroendocrine prostate cancer (NEPC) is a special subtype of prostate cancer that is rare but highly malignant; most cases of NEPC arise from the neuroendocrine differentiation (NED) of prostate cancer; this process is regarded as an adaptive response or drug resistance mechanism of prostate cancer. NEPC is characterized by the loss of androgen receptor expression and the acquisition of neuroendocrine markers such as chromogranin A and synaptophysin [59, 60].

We also constructed developmental trajectories for all cells using latent time analysis and identified differential genes by pseudotime analysis. Currently, pseudotime analysis is not commonly used for prostate cancer. Pseudotime analysis was used previously by Lee *et al.* [61] to validate the differentiation trajectory and fate of Androgen Receptor (*AR*)+ cells. In contrast, in the present study, we estimated the developmental trajectory of the neoplastic epithelium, which we hope to use to identify the process that causes healthy epithelial cells to develop into neoplastic epithelial cells and the specific biomarkers expressed during the differentiation process. By performing pseudotime analysis, we were able to estimate the developmental trajectory of tumor epithelial cells and gain further understanding of the developmental process of tumor epithelial cells in prostate cancer.

In addition, we analyzed differential gene expression in epithelial cell clustering using the hdWGCNA algorithm and used machine learning techniques to screen for potential prostate cancer biomarkers. By performing enrichment analysis, we found that the transmembrane transport of ions, particularly metal ions, was closely associated with the expression of genes identified by hdWGCNA, such as copper. Copper particles are now widely believed to be involved in the metabolism of epithelial cells in prostate cancer, primarily through the uptake of copper ions by tumors cells and the formation of new complexes. In line with this, a range of dithiocarbamates have been developed to induce apoptosis in tumor epithelial cells [62]. In contrast, in normal prostate tissue, the levels of copper and zinc plasma begin to increase with age, especially in the central region of the prostate. Thus, changes in ion levels can also be used to predict the process of transformation of prostate tissue towards malignancy, at least to some extent [63].

In order to develop a machine learning diagnostic model, we chose to screen for genes with high predictive relevance based on Lasso and logistic regression. Lasso regression is a method used to select linear regression variables from a set of variables that correlate. Several transcriptomic studies have already used machine learning to build different diagnostic models, such as the preoperative scoring of prostate cancer, and the risk assessment of drug-resistant prostate cancer [64, 65]. The machine learning models we constructed were based on Lasso genes; our model incorporating SVM and LDA showed strong sensitivity in both the training set and the test set. Future research should train this model using data from a larger cohort. Of course, diagnostic models based solely on transcriptomic gene expression may also have their own shortcomings, mainly due to the large individual differences in transcriptomes and the difficulty of combining them with more common clinical information. Therefore, a diagnostic model that combines immune scoring with transcriptomic gene expression should be developed as a matter of urgency.

Next, we evaluated immune infiltration in prostate cancer samples and constructed a convolutional neural network model for deep learning analysis. Our findings provide new concepts for further understanding the mechanisms responsible for the occurrence and development of prostate cancer, formulating individualized treatment strategies, and improving the prognosis of patients. Specifically, we identified the presence of immunological features between tumor samples and normal samples following initial immunological analysis of the samples. Furthermore, by using the NMF method, disease subtypes were established within tumors rents and identified immune cell species that were highly associated with disease subtype and Lasso genes.

The immune microenvironment of prostate cancer has long been a hot topic of research, and it is thought that immune cells play a two-way role in the development of this disease. On the one hand, immune cells, such as T cells and natural killer cells, are important in identifying and eliminating cancer cells, and can thus serve as a defense mechanism against the development of prostate cancer. However, on the other hand, immune cells could also promote the growth and spread of cancer cells by creating an inflammatory environment that supports tumor growth and suppresses anti-cancer immune responses [66]. In this respect, scRNA-seq data can better represent the heterogeneity within tumors. A previous ScRNA-seq study from the University of California demonstrated significant heterogeneity of epithelial cells in primary prostate cancer, particularly early growth response (EGR)-epithelial cells, which often elicit common tumor microenvironment responses [37]. This finding provided important clues for us to better understand the developmental mechanisms and treatment of prostate cancer [67, 68].

Deep learning enables the perfect combination of immune microenvironment assessment results and transcriptome expression levels by using the ratio of immune microenvironment scores to Lasso gene expression levels to build a high-dimensional array, perform convolution, and generate a new diagnostic model. Deep learning, a more advantageous modeling approach than machine learning, is currently used in the study of prostate cancer, mostly for pathological image recognition [69, 70]. Moreover, in a study, researchers were able to predict drug-resistant patients by building a deep learning model of biology, known as P-NET [71].

5. Conclusions

In this study, we utilized single-cell RNA sequencing to explore the characteristics of prostate cancer cells at the sequence level, cell-to-cell interactions, and immune infiltration. We

identified tumor-specific epithelial cell clusters; these tumor-associated epithelial cells were identified as exhibiting strong metal ion transport activities and could be used to screen for potential diagnostic biomarkers using machine learning. Our findings provide a deeper understanding of the mechanisms responsible for the development of prostate cancer and identified potential biomarker candidates for clinical diagnosis and treatment.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

AUTHOR CONTRIBUTIONS

FW, XQZ and JS—conception and design, obtaining funding and drafting the manuscript; JS and XQZ—acquisition of the data and drafting the manuscript; KL and KZ—statistical analysis and technical support. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The authors state that the manuscript has been approved by the Ethics Committee of Nanjing Medical University (20210329). Written informed consent has been obtained from the participants involved.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research was funded by Suzhou Youth Project of Science and Education for Medicine (grant no. SKJY2021127).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPLEMENTARY MATERIAL

Supplementary material associated with this article can be found, in the online version, at https://oss.jomh.org/files/article/1784850177118355456/attachment/Supplementary%20material.xlsx.

REFERENCES

[1] Culp MB, Soerjomataram I, Efstathiou JA, Bray F, Jemal A. Recent global patterns in prostate cancer incidence and mortality rates. European Urology. 2020; 77: 38–52.

[2] Roberts MJ, Maurer T, Perera M, Eiber M, Hope TA, Ost P, *et al*. Using PSMA imaging for prognostication in localized and advanced prostate cancer. Nature Reviews Urology. 2023; 20: 23–47.

[3] Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. The 2014 international society of urological pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. American Journal of Surgical Pathology. 2016; 40: 244–252.

[4] Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, *et al*. Machine and deep learning methods for radiomics. Medical Physics. 2020; 47: e185–e202.

[5] Trujillo B, Wu A, Wetterskog D, Attard G. Blood-based liquid biopsies for prostate cancer: clinical opportunities and challenges. British Journal of Cancer. 2022; 127: 1394–1402.

[6] Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial intelligence in anesthesiology. Anesthesiology. 2020; 132: 379–394.

[7] Petegrosso R, Li Z, Kuang R. Machine learning and statistical methods for clustering single-cell RNA-sequencing data. Briefings in Bioinformatics. 2020; 21: 1209–1223.

[8] García Garzón JR, de Arcocha Torres M, Delgado-Bolton R, Ceci F, Alvarez Ruiz S, Orcajo Rincón J, *et al*. 68Ga-PSMA PET/CT in prostate cancer. Revista EspañOla De Medicina Nuclear E Imagen Molecular. 2018; 37: 130–138.

[9] Wang X, An P, Gu Z, Luo Y, Luo J. Mitochondrial metal ion transport in cell metabolism and disease. International Journal of Molecular Sciences. 2021; 22: 7525.

[10] Bozzi AT, Gaudet R. Molecular mechanism of Nramp-family transition metal transport. Journal of Molecular Biology. 2021; 433: 166991.

[11] To PK, Do MH, Cho J-H, Jung C. Growth modulatory role of zinc in prostate cancer and application to cancer therapeutics. International Journal of Molecular Sciences. 2020; 21: 2991.

[12] Ghafoor S, Burger IA, Vargas AH. Multimodality imaging of prostate cancer. Journal of Nuclear Medicine. 2019; 60: 1350–1358.

[13] Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nature Reviews Immunology. 2018; 18: 35–45.

[14] Chen X, Sun Y, Guan N, Qu J, Huang Z, Zhu Z, *et al*. Computational models for lncRNA function prediction and functional similarity calculation. Briefings in Functional Genomics. 2019; 18: 58–82.

[15] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521: 436–444.

[16] Bao J-h, Lu W-c, Duan H, Ye YQ, Li JB, Liao WT, *et al*. Identification of a novel cuproptosis-related gene signature and integrative analyses in patients with lower-grade gliomas. Frontiers in Immunology. 2022; 13: 933973.

[17] Adamidi ES, Mitsis K, Nikita KS. Artificial intelligence in clinical care amidst COVID-19 pandemic: a systematic review. Computational and Structural Biotechnology Journal. 2021; 19: 2833–2850.

[18] Benítez-Parejo N, Rodríguez del Águila MM, Pérez-Vicente S. Survival analysis and Cox regression. Allergologia Et Immunopathologia. 2011; 39: 362–373.

[19] Wong HY, Sheng Q, Hesterberg AB, Croessmann S, Rios BL, Giri K, *et al*. Single cell analysis of cribriform prostate cancer reveals cell intrinsic and tumor microenvironmental pathways of aggressive disease. Nature Communications. 2022; 13: 6036.

[20] Henry GH, Malewska A, Joseph DB, Malladi VS, Lee J, Torrealba J, *et al*. A cellular anatomy of the normal adult human prostate and prostatic urethra. Cell Reports. 2018; 25: 3530–3542.e5.

[21] Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, *et al*. Integrated analysis of multimodal single-cell data. Cell. 2021; 184: 3573–3587.e29.

[22] Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, *et al*. Fast, sensitive and accurate integration of single-cell data with Harmony. Nature Methods. 2019; 16: 1289–1296.

[23] Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, *et al*. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nature Immunology. 2019; 20: 163–172.

[24] Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, *et al*. Inference and analysis of cell-cell communication using CellChat. Nature Communications. 2021; 12: 1088.

[25] Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, *et al*. Reversed graph embedding resolves complex single-cell trajectories. Nature Methods. 2017; 14: 979–982.

[26] Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. Cell Reports Methods. 2023; 3: 100498.

[27] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, *et al*. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications. 2019; 10: 1523.

[28] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics. 2013; 14: 7.

[29] Zeng D, Ye Z, Shen R, Yu G, Wu J, Xiong Y, *et al*. IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. Frontiers in Immunology. 2021; 12: 687975.

[30] Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010; 11: 367.

[31] Mohammad N, Muad AM, Ahmad R, Yusof MYPM. Accuracy of advanced deep learning with tensorflow and keras for classifying teeth developmental stages in digital panoramic imaging. BMC Medical Imaging. 2022; 22: 66.

[32] Buddingh' BC, Elzinga J, van Hest JC. Intercellular communication between artificial cells by allosteric amplification of a molecular signal. Nature Communications. 2020; 11: 1652.

[33] Noël F, Massenet-Regad L, Carmi-Levy I, Cappuccio A, Grandclaudon M, Trichot C, *et al*. Dissection of intercellular communication using the transcriptome-based framework ICELLNET. Nature Communications. 2021; 12: 1089.

[34] Hou R, Denisenko E, Ong HT, Ramilowski JA, Forrest ARR. Predicting cell-to-cell communication networks using NATMI. Nature Communications. 2020; 11: 5011.

[35] Zhang Y, Wang D, Peng M, Tang L, Ouyang J, Xiong F, *et al*. Single-cell RNA sequencing in cancer research. Journal of Experimental & Clinical Cancer Research. 2021; 40: 81.

[36] Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Experimental & Molecular Medicine. 2018; 50: 1–14.

[37] Song H, Weinstein HN, Allegakoen P, Wadsworth MH 2nd, Xie J, Yang H, *et al*. Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. Nature Communications. 2022; 13: 141.

[38] Masetti M, Carriero R, Portale F, Marelli G, Morina N, Pandini M, *et al*. Lipid-loaded tumor-associated macrophages sustain tumor growth and invasiveness in prostate cancer. Journal of Experimental Medicine. 2022; 219: e20210564.

[39] Chen Y, Zhang P, Liao J, Cheng J, Zhang Q, Li T, *et al*. Single-cell transcriptomics reveals cell type diversity of human prostate. Journal of Genetics and Genomics. 2022; 49: 1002–1015.

[40] Han H, Lee HH, Choi K, Moon YJ, Heo JE, Ham WS, *et al*. Prostate epithelial genes define therapy-relevant prostate cancer molecular subtype. Prostate Cancer and Prostatic Diseases. 2021; 24: 1080–1092.

[41] Sekhoacha M, Riet K, Motloung P, Gumenku L, Adegoke A, Mashele S. Prostate cancer review: genetics, diagnosis, treatment options, and alternative approaches. Molecules. 2022; 27: 5730.

[42] Andl T, Ganapathy K, Bossan A, Chakrabarti R. MicroRNAs as guardians of the prostate: those who stand before cancer. What do we really know about the role of microRNAs in prostate biology? International Journal of Molecular Sciences. 2020; 21: 4796.

[43] Gonzalez-Avila G, Sommer B, García-Hernandez AA, Ramos C, Flores-Soto E. Nanotechnology and matrix metalloproteinases in cancer diagnosis and treatment. Frontiers in Molecular Biosciences. 2022; 9: 918789.

[44] Cabral-Pacheco GA, Garza-Veloz I, Castruita-De la Rosa C, Ramirez-Acuña JM, Perez-Romero BA, Guerrero-Rodriguez JF, *et al*. The roles of matrix metalloproteinases and their inhibitors in human diseases. International Journal of Molecular Sciences. 2020; 21: 9739.

[45] Geng X, Chen C, Huang Y, Hou J. The prognostic value and potential mechanism of matrix metalloproteinases among prostate cancer. International Journal of Medical Sciences. 2020; 17: 1550–1560.

[46] Niland S, Riscanevo AX, Eble JA. Matrix metalloproteinases shape the tumor microenvironment in cancer progression. International Journal of Molecular Sciences. 2021; 23: 146.

[47] Carruba G, Stefano R, Cocciadiferro L, Saladino F, Di Cristina A, Tokar E, *et al*. Intercellular communication and human prostate carcinogenesis. Annals of the New York Academy of Sciences. 2002; 963: 156–168.

[48] Wu Y, Clark KC, Niranjan B, Chüeh AC, Horvath LG, Taylor RA, *et al*. Integrative characterisation of secreted factors involved in intercellular communication between prostate epithelial or cancer cells and fibroblasts. Molecular Oncology. 2023; 17: 469–486.

[49] Erdogan S, Doganlar ZB, Doganlar O, Turkekul K, Serttas R. Inhibition of midkine suppresses prostate cancer CD133 + stem cell growth and migration. The American Journal of the Medical Sciences. 2017; 354: 299–309.

[50] Meyer-Siegler K, Hudson PB. Enhanced expression of macrophage migration inhibitory factor in prostatic adenocarcinoma metastases. Urology. 1996; 48: 448–452.

[51] Rafiei S, Gui B, Wu J, Liu XS, Kibel AS, Jia L. Targeting the MIF/CXCR7/AKT signaling pathway in castration-resistant prostate cancer. Molecular Cancer Research. 2019; 17: 263–276.

[52] Bokobza SM, Ye L, Kynaston H, Jiang WG. Growth and differentiation factor 9 (GDF-9) induces epithelial–mesenchymal transition in prostate cancer cells. Molecular and Cellular Biochemistry. 2011; 349: 33–40.

[53] Bokobza SM, Ye L, Kynaston HG, Jiang WG. GDF-9 promotes the growth of prostate cancer cells by protecting them from apoptosis. Journal of Cellular Physiology. 2010; 225: 529–536.

[54] Nordin A, Wang W, Welén K, Damber J. Midkine is associated with neuroendocrine differentiation in castration-resistant prostate cancer. The Prostate. 2013; 73: 657–667.

[55] Zhou Q, Yang C, Mou Z, Wu S, Dai X, Chen X, *et al*. Identification and validation of a poor clinical outcome subtype of primary prostate cancer with Midkine abundance. Cancer Science. 2022; 113: 3698–3709.

[56] Hilscher M, Røder A, Helgstrand JT, Klemann N, Brasso K, Vickers AJ, *et al*. Risk of prostate cancer and death after benign transurethral resection of the prostate—a 20-year population-based analysis. Cancer. 2022; 128: 3674–3680.

[57] Oczkowski M, Dziendzikowska K, Pasternak-Winiarska A, Włodarek D, Gromadzka-Ostrowska J. Dietary factors and prostate cancer development, progression, and reduction. Nutrients. 2021; 13: 496.

[58] Al-Fayez S, El-Metwally A. Cigarette smoking and prostate cancer: a systematic review and meta-analysis of prospective cohort studies. Tobacco Induced Diseases. 2023; 21: 19.

[59] Storck WK, May AM, Westbrook TC, Duan Z, Morrissey C, Yates JA, *et al*. The role of epigenetic change in therapy-induced neuroendocrine prostate cancer lineage plasticity. Frontiers in Endocrinology. 2022; 13: 926585.

[60] Merkens L, Sailer V, Lessel D, Janzen E, Greimeier S, Kirfel J, *et al*. Aggressive variants of prostate cancer: underlying mechanisms of neuroendocrine transdifferentiation. Journal of Experimental & Clinical Cancer Research. 2022; 41: 46.

[61] Lee D-H, Olson AW, Wang J, Kim WK, Mi J, Zeng H, *et al*. Androgen action in cell fate and communication during prostate development at single-cell resolution. Development. 2021; 148: dev196048.

[62] Safi R, Nelson ER, Chitneni SK, Franz KJ, George DJ, Zalutsky MR, *et al*. Copper signaling axis as a target for prostate cancer therapeutics. Cancer Research. 2014; 74: 5819–5831.

[63] Daragó A, Klimczak M, Stragierowicz J, Jobczyk M, Kilanowicz A. Age-related changes in zinc, copper and selenium levels in the human prostate. Nutrients. 2021; 13: 1403.

[64] Yu Y, Liu S, Ren B, Nelson J, Jarrard D, Brooks JD, *et al*. Fusion gene detection in prostate cancer samples enhances the prediction of prostate cancer clinical outcomes from radical prostatectomy through machine learning in a multi-institutional analysis. The American Journal of Pathology. 2023; 193: 392–403.

[65] Blatti C, de la Fuente J, Gao H, Marín-Goñi I, Chen Z, Zhao SD, *et al*. Bayesian machine learning enables identification of transcriptional network disruptions associated with drug-resistant prostate cancer. Cancer Research. 2023; 83: 1361–1380.

[66] Wang Y, Wu S, Su H, Zhao X. Role of tumor-associated immune cells in prostate cancer: angel or devil? Asian Journal of Andrology. 2019; 21: 433.

[67] Meng J, Zhou Y, Lu X, Bian Z, Chen Y, Zhou J, *et al*. Immune response drives outcomes in prostate cancer: implications for immunotherapy. Molecular Oncology. 2021; 15: 1358–1375.

[68] Runcie KD, Dallos MC. Prostate cancer immunotherapy—finally in from the cold? Current Oncology Reports. 2021; 23: 88.

[69] Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell

N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Medicine. 2021; 13: 152.

[70] Naik N, Tokas T, Shetty DK, Hameed BMZ, Shastri S, Shah MJ, *et al*. Role of deep learning in prostate cancer management: past, present and future based on a comprehensive literature review. Journal of Clinical Medicine. 2022; 11: 3575.

[71] Elmarakeby HA, Hwang J, Arafeh R, Crowdis J, Gang S, Liu D, *et al*. Biologically informed deep neural network for prostate cancer discovery. Nature. 2021; 598: 348–352.

How to cite this article: Jian Sun, Xueqi Zhu, Kai Li, Ke Zhang, Fei Wang. Application of a high-dimensional gene co-expression network to identify metal ion transport-associated epithelial cells with diagnostic function for prostate cancer. Journal of Men's Health. 2024; 20(4): 39-53. doi: 10.22514/jomh.2024.052.