Research article

Identification of a 24-gene panel and a novel marker of *PODXL2* essential for the pathological diagnosis of early prostate cancer

Xiaoshi Ma [a,b,1], Lipeng Chen [c,1], Tao Chen [d,1], Kun Chen [e,1], Huirong Zhang [c], Kaipeng Huang [f], Han Zheng [g], Hongtao Jin [e,*], Zhiqiang Cheng [h,*], Kefeng Xiao [a,b,**], Jinan Guo [a,b,**]

[a] *Department of Urology, Shenzhen People's Hospital (The Second Clinical Medical College of Jinan University), Shenzhen 518020, Guangdong, China*
[b] *Clinical Research Center for Geriatrics, Shenzhen People's Hospital (The Second Clinical Medical College of Jinan University), Shenzhen 518020, Guangdong, China*
[c] *Clinical Medical Research Center, Shenzhen People's Hospital (The Second Clinical Medical College of Jinan University), Shenzhen 518020, Guangdong, China*
[d] *School of Medicine, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China*
[e] *Department of Radiotherapy, Nanjing Medical University The Fourth School of Clinical Medicine, Nanjing 210000, Jiangsu, China*
[f] *Department of Pathology, Shenzhen People's Hospital (The Second Clinical Medical College of Jinan University), Shenzhen 518020, Guangdong, China*
[g] *Department of Urology, The First Affiliated Hospital of Wannan Medical College (Yijishan Hospital of Wannan Medical College), Wuhu 241000, Anhui, China*
[h] *Department of Pathology, Third People's Hospital of Shenzhen (The Second Affiliated Hospital of Southern University of Science and Technology), Shenzhen 518100, Guangdong, China*

ARTICLE INFO

ABSTRACT

Precise diagnosis of early prostate cancer (PCa) is critical for preventing tumor progression. However, the diagnostic outcomes of currently used markers are far from satisfactory due to the low sensitivity or specificity. Here, we identified a diagnostic subpopulation in PCa tissue with the integrating analysis of single-cell and bulk RNA-seq. The representative markers of this subpopulation were extracted to perform intersection analysis with early-PCa-related gene module generated from weighted correlation network analysis (WGCNA). A total of 24 overlapping genes were obtained, the diagnostic roles of which were validated by distinguishing normal and tumorous prostate samples from the public dataset. A least absolute shrinkage and selection operator (LASSO) model was constructed based on these genes and the obtained 24-gene panel showed high sensitivity and specificity for PCa diagnosis, with better identifying capability of PCa than the commercially used gene panel of Oncotype DX. The top two risk factors, *TRPM4* and *PODXL2*, were verified to be highly expressed in early PCa tissues by multiplex immunostaining, and *PODXL2* was more sensitive and specific compared to *TRPM4* and the pathologically used marker *AMACR* for early PCa diagnosis, suggesting a novel and promising pathology marker.

1. Introduction

Prostate cancer (PCa) is one of the most common malignancies in the male genitourinary system, with approximately 1.5 million new cases a year [1]. As the advanced PCa remains to be a leading factor in tumor causing deaths, early diagnosis has become increasingly critical [2].

Over the years, nuclear magnetic resonance imaging (NMRI), serum prostate-specific antigen (PSA) testing and digital rectal examination (DRE) has been the main approaches for early PCa detection [3]. However, all of these approaches have their own limitations, such as low sensitivity or specificity and a high false positive rate, resulting in plenty of overdiagnosis and misdiagnosis [3]. Prostatic needle biopsy is a

relatively more accurate and currently preferred approach for early PCa identification. Sometimes, morphological changes are difficult to identify by only histological observation due to the very limited malignant glands in small specimens, causing problems in confirming or discarding the possibility of PCa [4].

To accurately examine the morphological features of PCa, such as the loss of basal cells and hyperplasia of luminal epithelia, abundant markers have been used for immunohistochemical identification. TP63, as a nuclear protein from the tumor-suppressor gene *TP53* family, specifically stains basal cells and is therefore widely used in the pathological diagnosis of PCa [5]. However, PCa biopsies are mostly not sufficient for the evaluation of the presence or absence of basal cells, and to worsen this situation, several lines of studies have shown that nearly 0.3–1.1% of PCa tissues are positive for basal cells [6,7]. AMACR is a mitochondrial and peroxisomal enzyme involved in the β-oxidation of fatty acids and is significantly overexpressed in malignant luminal cells [8]. In recent decades, it has usually been used as an adjunct to basal cell markers to identify early PCa [9]. Despite its usefulness in PCa diagnosis, *AMACR* is also occasionally positive in benign prostates and conversely negative in up to 25 % of typical PCa tissues [5,6]. It is urgent to identify more specific and sensitive markers to prevent missing the optimal time for PCa treatment.

Better understanding of the mutational and transcriptomic landscapes in PCa has helped us to identify various potential markers for early PCa detection, including *NKX 3.1* [10], *GOLM1* [11] and *FOLH1* [12]. However, the clinical applications of these candidate markers in PCa diagnosis are very limited due to their low sensitivity or specificity. PCa tissues consist of multiple types of cells, and it is practically impossible to precisely examine the expression characteristics of tumor cells by bulk RNA-seq. With the development of high-throughput sequencing in recent years, single-cell RNA-seq has been applied in various types of malignancies to reveal the expression profiles of tumor cells and identify distinct subpopulations, providing many novel perspectives on tumorigenesis, progression and drug resistance [13–16]. Additionally, in our previous study, we identified a unique subpopulation in PCa essential for diagnosis and stratification and verified the expression features of its representative marker *HPN* [17].

To identify more sensitive and specific markers for early PCa diagnosis, in the present study, we combined the single-cell and bulk RNA-seq profiles to identify a diagnostic subpopulation similar to our previous findings [17], and performed weighted correlation network analysis (WGCNA) to identify the diagnostic module. To narrow down the candidate genes, we conducted intersection analysis and used the overlapping genes to construct a least absolute shrinkage and selection operator (LASSO) model for early PCa diagnosis. Furthermore, top two risk factors from this gene model were identified and examined if capable of distinguishing early PCa from normal prostates by immunostaining in tissue microarray. Our findings provided a diagnostic gene panel consisted of genes essential for PCa early development and a novel molecule with high sensitivity and specificity in pathology diagnosis of early PCa.

2. Materials and methods

2.1. Data collection and processing

The single-cell RNA-seq profiles of 10 PCa patients were obtained from the website www.prostatecellatlas.org. Tumorous and matching normal samples from patients with Gleason scores of 6 or 7 were extracted. The mRNA expression profiles of these patients have underwent quality control to remove environmental RNA contamination and doublets via the SoupX R [18] and Scrublet packages [19], respectively. High-quality cells with identified genes ranged from 200 to 6000, and mitochondrial gene proportions less than 15 % were preserved. Mitochondrial and ribosomal genes were further removed, as well as those expressed in fewer than 3 cells. The gene expression matrices for the remaining cells were normalized and integrated according to the standard procedure of the Seurat R package [20].

The bulk RNA-seq profiles of PCa patients were downloaded from the TCGA and GTEx databases. The TCGA dataset contains 495 PCa samples and 51 normal prostate samples, while the GTEx dataset includes 100 normal prostate samples. The count matrices of these two datasets were integrated, and we used the SVA package to remove batch effects (Fig. S1) [21]. Tumor samples with Gleason scores greater than 7 were removed from further analysis.

2.2. Clustering and cell-type annotation

The cell cycle effects on clustering were removed in ScaleData. The top 2000 variable genes were extracted, and dimensionality reduction was then performed with principal component analysis (PCA). Cell clusters were visualized by a t-distributed stochastic neighbor embedding (tSNE) plot. Marker genes of each cluster were identified by FindMarkers and then used to align to the known markers of major types of cells in the CellMarker database for cell-type annotation.

Subclustering of epithelial cells was performed using the same method described above. Marker genes of each subpopulation identified by FindMarkers were further filtered by $\log_2FC > 0.5$, *p* value < 0.05 and pct.1 > 0.3. Since there are no universal standards for the annotation of the epithelial subpopulation, the subpopulations were annotated as representative marker genes.

2.3. Gene set variation analysis (GSVA)

To determine the potential pathways in which each subpopulation was enriched, we examined the expression levels of representative genes from 50 hallmark pathways in the molecular signature database using the GSVA package (version 1.48.0) with default settings as described in the GSVA github repository [22]. The obtained hallmark gene set scores for each subpopulation were visualized by heatmap.

2.4. Expression correlation to public data

Public bulk RNA-seq profiles of PCa patients were integrated by TCGA and GTEx datasets as described above. The top 20 markers of each subpopulation were extracted to represent their expression signatures using the AddModuleScore function and then used to examine the expression levels in normal and tumorous samples from the integrated dataset. Conversely, the expression levels of the PCa signature, obtained by the upregulated genes in tumorous samples (Gleason score = 6, 7) from the integrated dataset, were examined in each subpopulation and visualized by feature plot and violin plot. Furthermore, the correlations of each subpopulation to normal and tumorous prostate tissues (Gleason score = 6, 7) were examined by a hypergeometric distribution test and visualized by heatmap.

2.5. Identification of the PCa diagnostic subpopulation

BayesPrism was used to perform deconvolution of bulk RNA-seq profiles from our integrated dataset based on the single-cell RNA-seq data [23], and the cell proportion of each epithelial subpopulation was estimated by their gene expression levels in bulk samples. The diagnostic capability of each subpopulation was evaluated by receiver operating characteristic (ROC) curves created by the pROC R package based on the cell proportion and clinical information of the tumor event [24]. Furthermore, the cell proportion of each subpopulation in patients from single-cell data was analyzed and visualized by pie graph. The subpopulation with the highest area under the curve (AUC) value and positive rate in each patient was considered the PCa diagnostic subpopulation.
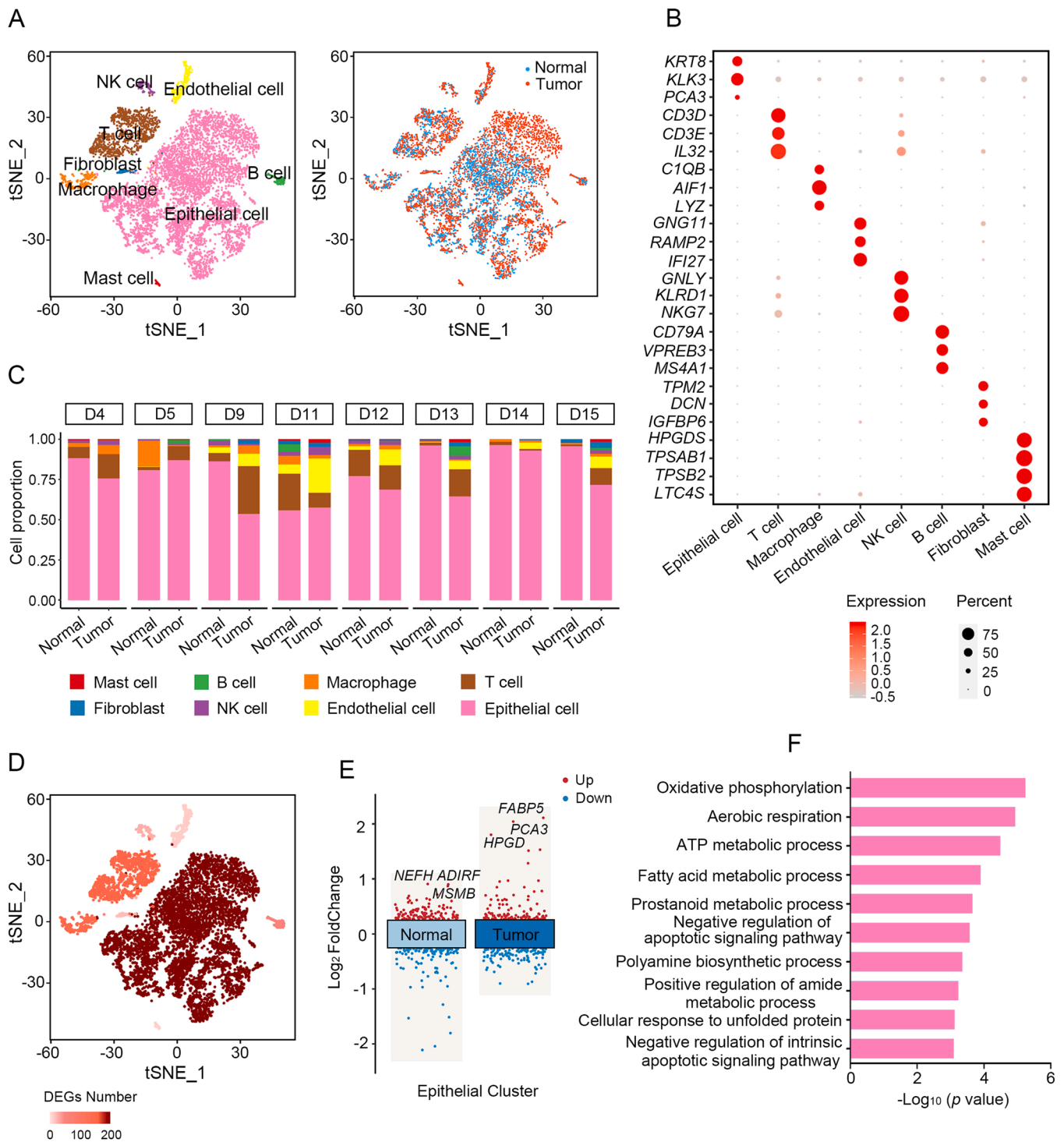
Fig. 1. Cell-type landscape of tumorous and matched normal prostate tissues revealed by single-cell RNA-seq. (A) Cellular compositions of normal and tumorous prostate tissues. (B) The expression levels of representative markers of each cluster. (C) Cell-type proportions of normal and tumorous prostate tissues in each patient. (D) Number of DEGs between normal and tumorous prostate tissues within each cluster. (E) Upregulated (red) and downregulated (blue) DEGs in epithelial cluster between normal and tumorous prostate tissues. (F) The biological progresses that DEGs between normal and tumorous epithelia mainly enriched in.

2.6. Weighted correlation network analysis (WGCNA)

The WGCNA R package was used to construct the gene co-expression network [25]. After removal of outlier samples and genes, a total of 10, 000 genes with the top median absolute deviations were selected for subsequent analysis. To explore the correlation between gene modules and early PCa development, normal and tumorous samples with Gleason scores of 6 or 7 were converted into numerical values of 0 and 1, respectively. The pathology gradings of tumorous samples were represented by their Gleason scores, while the normal samples were assigned the value of 0. The gene module with the most significant correlation to both tumor event and Gleason score was considered the PCa diagnostic module.
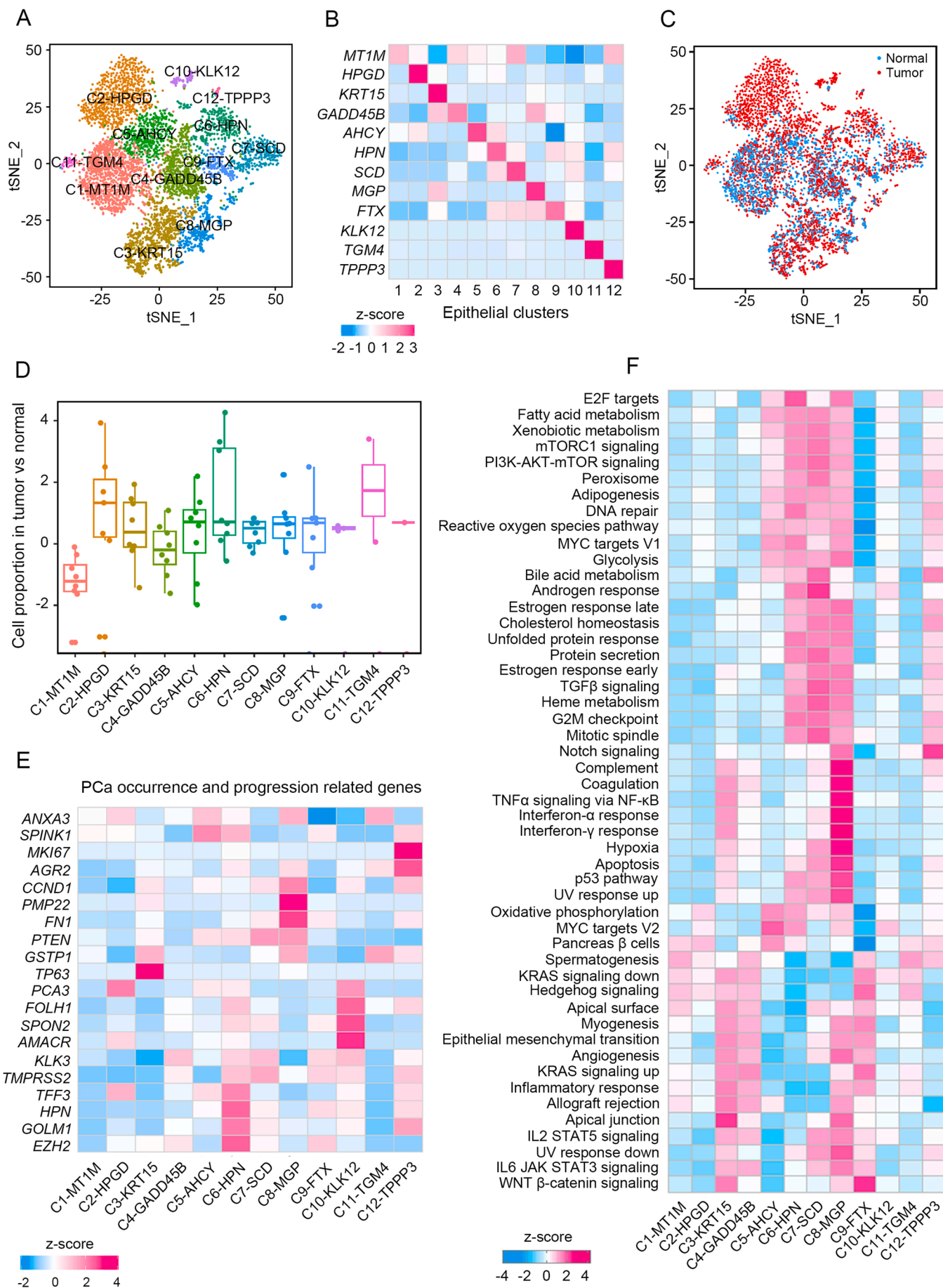
(caption on next page)

Fig. 2. Epithelial cell reclustering and the characterization of each subpopulation. (A) Epithelial cells reclustered into 12 subpopulations. (B) The expression levels of representative markers of each epithelial subpopulation. (C) Normal and tumorous tissue distribution of each subpopulation. (D) Cell proportion of each subpopulation in tumorous vs normal tissue. (E) The expression levels of PCa progression-related genes in each subpopulation. (F) The pathway enrichment analysis of each subpopulation evaluated by GSVA.

2.7. Least absolute shrinkage and selection operator (LASSO)

A total of 24 overlapping genes between the marker genes of the diagnostic subpopulation and the genes in the diagnostic module were identified. Based on these genes, LASSO regression analysis was performed to establish a diagnostic model of early PCa [26]. Samples from the integrated public dataset were randomly divided into the training set and test set at a ratio of 7:3. The diagnostic model was established using the training set to distinguish normal and tumorous prostates. The predictive capability was evaluated by ROC analysis and further validated using the test set. The top risk genes were identified by the regression coefficient of each gene from the 24-gene panel, the higher of which indicated more critical roles in the LASSO model.

2.8. Multiplex Immunostaining

The human PCa tissue microarray (HProA150PG02, Shanghai Outdo Biotech Co. Ltd., China), consisting of 55 normal tissues and 95 tumorous tissues with Gleason scores ranging from 6 to 9, was obtained to examine the expression patterns of PODXL2, TRPM4 and AMACR. The multiplex immunostaining assay was performed as described before with the primary antibodies of anti-PODXL2 (diluted at 1:200, Sigma Aldrich, cat. no. HPA042265), anti-AMACR (MXB Biotechnologies, cat. no. RMA-0546) and anti-TRPM4 (diluted at 1:200, Proteintech, cat. no. 21985–1-AP) [17]. The corresponding staining order was Alexa Fluor™ 488-labeled tyramide (Thermo Fisher Scientific, cat. no. B40953), Alexa Fluor™ 555-labeled tyramide (Thermo Fisher Scientific, cat. no. B40955) and Alexa Fluor™ 647-labeled tyramide (Thermo Fisher Scientific, cat. no. B40958).

The staining score was evaluated by two qualified pathologists as follows: positive signal rate (negative: 0 points, 1–25 % of cells: 1 point, 26–50 % of cells: 2 points, 51–75 % of cells: 3 points, and 76–100 % of cells: 4 points) and positive signal intensity (none: 0 points, weak staining: 1 point, moderate staining: 2 points, strong staining: 3 points). Integration of the positive signal rate and intensity scores by multiplication was the final staining score of each antibody. The staining scores were presented as the means ± standard deviation and analyzed by one-way analysis of variance (ANOVA). Differences between groups were evaluated using the Student-Newman-Keuls multiple comparison test in SPSS version 20.0 and considered statistically significant at $p < 0.05$. Samples of squamous carcinoma, signet-ring cell carcinoma and mucinous carcinoma from the tissue microarray were filtered out during the staining score evaluation.

3. Results

3.1. Cellular composition of normal and tumorous prostates

The single-cell RNA-seq profiles of 16 samples of normal and tumorous prostates from 8 patients with Gleason scores of 6 or 7 were preserved for dissection of the cellular composition. A total of 10,176 cells passed quality control and were divided into 8 clusters (Fig. 1A). Based on the expression levels of the representative cell-type markers (Table S1), the clusters were annotated as B cells, endothelial cells, epithelial cells, fibroblasts, macrophages, mast cells, NK cells and T cells (Fig. 1A, B). None of these cells were disease- or patient-specific, demonstrating that the cells were clustered predominantly by cell-type characteristics rather than batch effects (Fig. 1A, C). The cellular composition between tumorous and matching normal prostates within the same individual varied significantly, as did that among patients,

suggesting high heterogeneity of PCa (Fig. 1C).

Differently expressed genes (DEGs) between normal and tumorous prostates within each cluster were examined and showed epithelial cells with the highest transcriptomic differences (Fig. 1D). Tumorous epithelial cells highly expressed genes related to PCa progression, such as *FABP5*, *PCA3* and *HPGD* (Fig. 1E). Subsequently, we performed functional enrichment analysis to explore the biological significance and found that tumorous epithelial cells highly expressed genes mainly involved in oxidative phosphorylation, biological macromolecule metabolism and negative regulation of apoptosis (Fig. 1F). Considering PCa is an epithelia-derived malignancy and the epithelial cells exhibited prominent transcriptomic changes related to tumor occurrence and progression, we therefore focused on epithelial cells for further analysis of identifying PCa diagnosis-related cells.

3.2. High heterogeneity of PCa tissue

The epithelial cells were extracted and performed data normalization using ScaleData function in the Seurat R package. After removal of cell cycle effects, we conducted PCA and clustered them into 12 subpopulations with the resolution parameter of 0.5 and dims of 1:20. Marker genes were identified and used for the annotation of each subpopulation (Fig. 2A, B, Table S2). Although none of these cells were disease-specific, several subpopulations were approximately enriched in tumorous prostates, including C2-HPGD, C6-HPN, C10-KLK12 and C11-TGM4, suggesting positive correlations with early PCa development (Fig. 2C). To eliminate the impact of different captured cell counts, we contrasted the cell proportions of each subpopulation in normal and tumorous samples and further validated the tumorous enrichment of these subpopulations in PCa (Fig. 2D). In addition, except for C1-MTIM and C4-GADD45B, all subpopulations were mainly distributed in tumorous prostates (Fig. 2D).

To determine the malignant subpopulations, we examined the expression levels of reported PCa-related genes and performed GSVA for each subpopulation. C1-MT1M showed no significantly high expression of any PCa-related genes, suggesting no participation in PCa occurrence of progression (Fig. 2D, E). C2-HPGD highly expressed the urine marker of PCa, *PCA3*, and genes enriched in oxidative phosphorylation and MYC targets, exhibiting malignant potentials (Fig. 2E, F). C3-KRT15 cells highly expressed basal markers such as *TP63* and *KRT15*, possibly being basal cells [4] (Fig. 2E). Additionally, the biological processes of C4-GADD45B markers enriched in were very similar to those of C3-KRT15 (Fig. 2E, F). C5-AHCY highly expressed genes, including *ANXA3* and *SPINK1*, were mainly enriched in MYC targets and PCa-related aberrant metabolism (Fig. 2E, F). In contrast, C6-HPN highly or moderately expressed abundant PCa diagnostic markers, including *SPINK1*, *PCA3*, *FOLH1*, *AMACR*, *HPN*, *GOLM1* and *EZH2*, suggesting a diagnostic role for PCa [4,12] (Fig. 2E). C7-SCD and C8-MGP exhibited similar expression characteristics to C6-HPN, despite the low expression levels of diagnostic markers (Fig. 2F). C9-FTX moderately expressed PCa early development-related genes, such as *KLK3*, *TMPRSS2*, *HPN* and *EZH2*, and was mainly involved in epithelial mesenchymal transition (EMT) and angiogenesis, suggesting an infiltration- or migration-promoting role [4,12] (Fig. 2E, F). Similar to C6-HPN, C10-KLK12 highly expressed *PCA3*, *FOLH1* and *AMACR*, the widely used PCa diagnostic molecules [4,12] (Fig. 2E). High expression of *GSTP1* suggested that C11-TGM4 could be benign luminal cells, as it is an epigenetic biomarker for PCa, DNA methylation of which is nearly universally present in all PCa cells [27] (Fig. 2E). C12-TPPP3 exhibited the most proliferative potential due to the high expression levels of
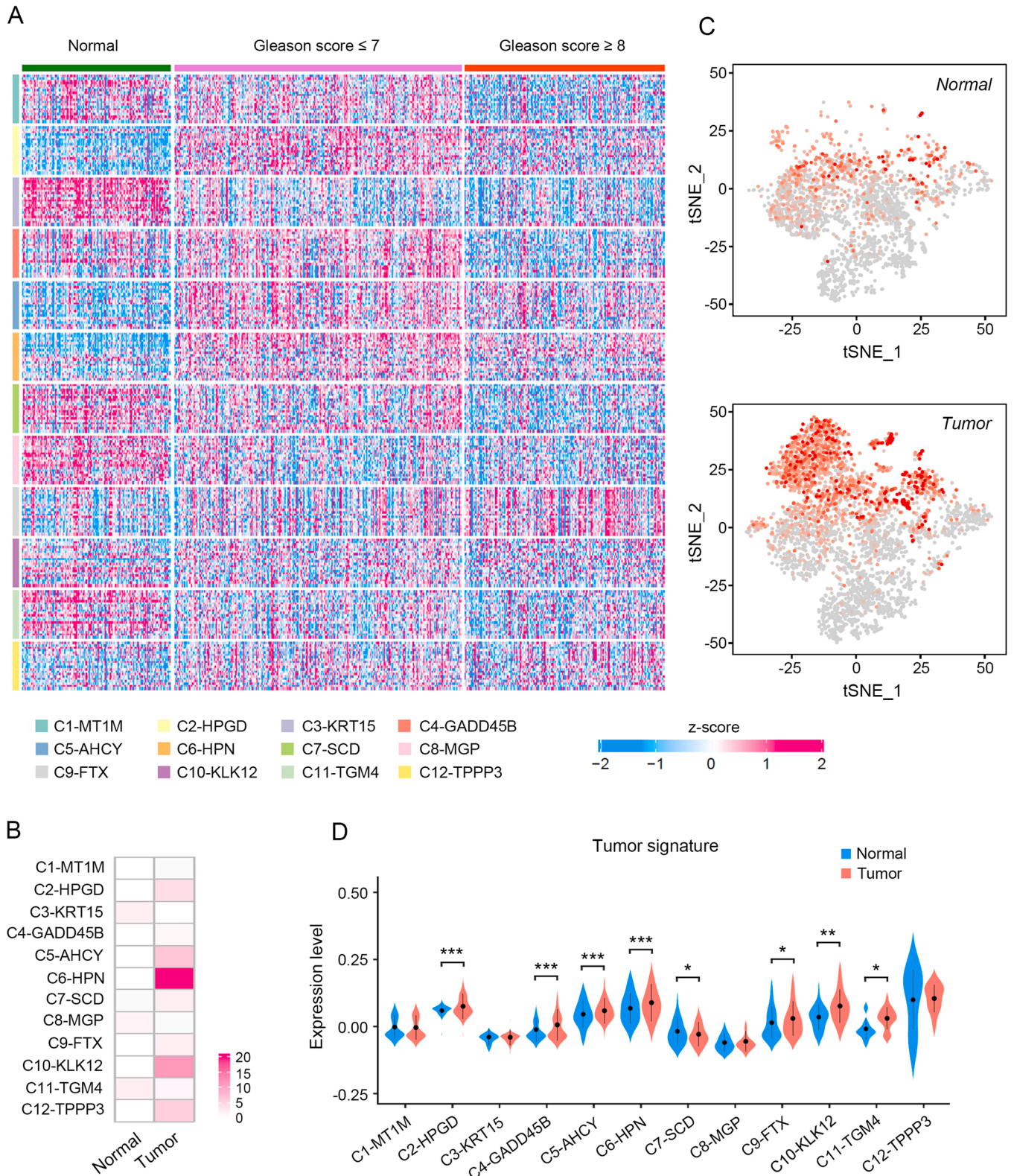
Fig. 3. Identification of malignant cells by clinical correlation analysis of each subpopulation in PCa progression. (A) The expression levels of marker genes from each subpopulation examined by the integrated dataset of TCGA and GTEx. (B) The correlations between the highly expressed genes in each subpopulation and the DEGs of normal and tumorous prostate samples from public datasets. (C) Feature plot and (D) violin plot showing the expression levels of PCa highly expressed genes in each subpopulation of normal and tumorous prostates, respectively. $* p < 0.05$, $* * p < 0.01$, $* * * p < 0.001$.

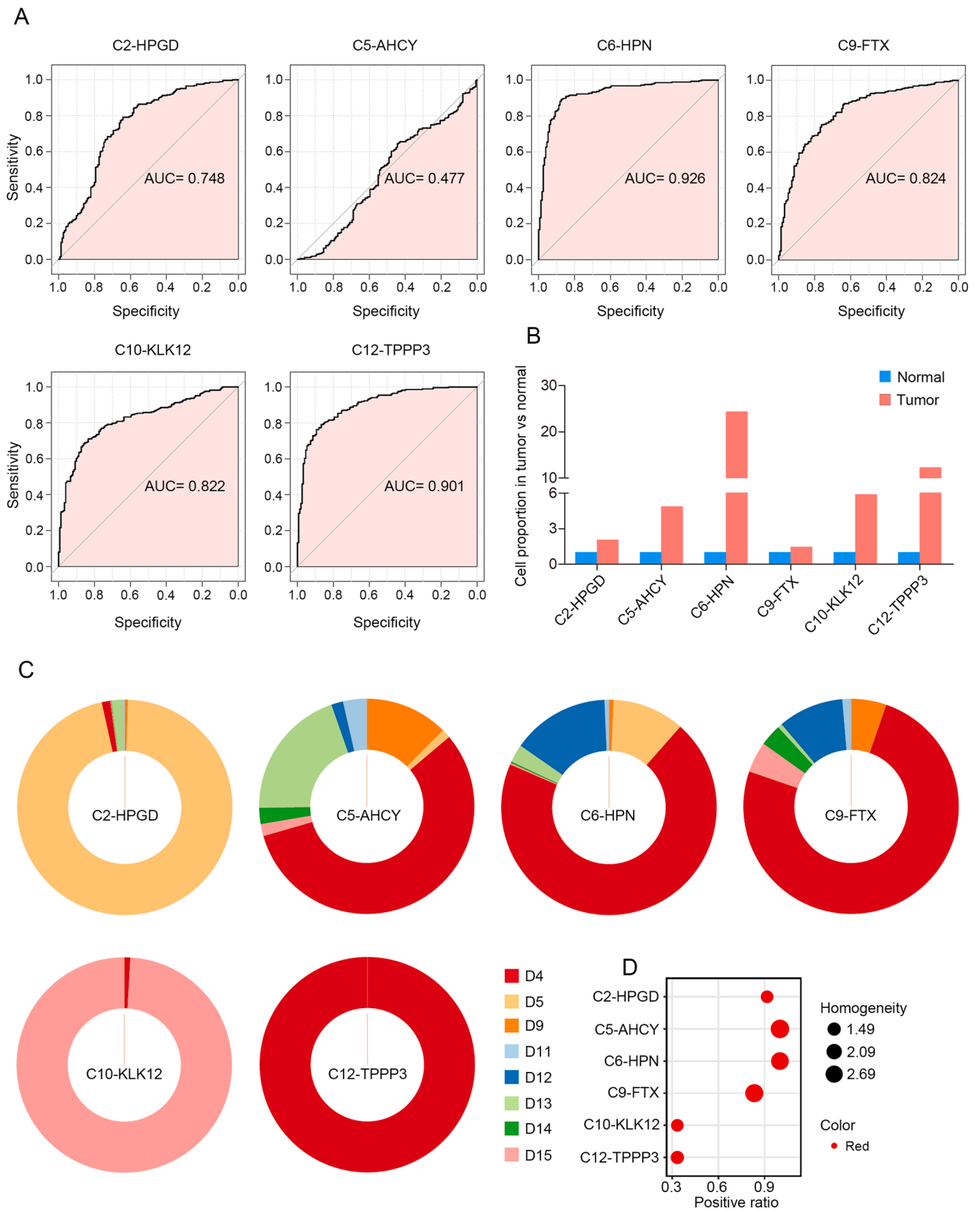**Fig. 4.** The diagnostic capability and sample distribution of each malignant subpopulation. (A) The diagnostic capability of each candidate malignant subpopulation in early PCa evaluated by ROC analysis. (B) Cell proportion of each malignant subpopulation in tumorous vs normal tissue. (C) The sample distribution of each malignant subpopulation. (D) The homogeneity of the sample distribution of each malignant subpopulation.
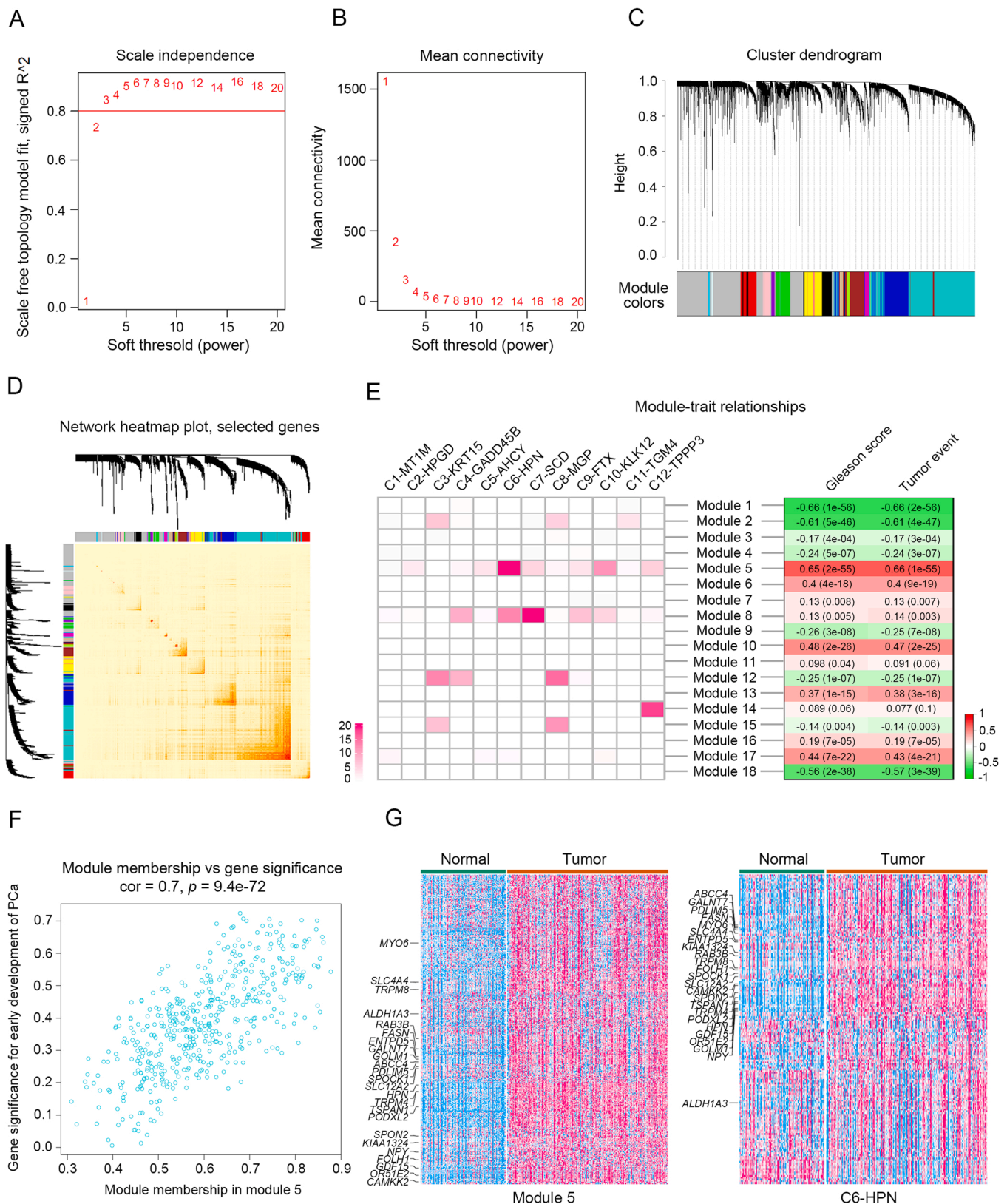
Fig. 5. Screening of 24 genes with critical potential in early PCa diagnosis by WGCNA. (A) The correlation of the power values of the soft threshold with the scale-free topology model fit. (B) The relationship between power values and mean connectivity. (C) The clustered gene modules with different colors. Each color represents one kind of gene module. Gray color represents genes that could not be classified into any gene module. (D) Topological correlations between genes within different modules. (E) The correlations of different gene modules with the gene expression signature of each subpopulation and the clinical information, including Gleason score and tumor event. (F) The correlation of genes in module 5 with the upregulated genes in early PCa. (G) The expression levels of genes in module 5 and markers of C6-HPN in normal and tumorous prostates. The overlapping genes are shown on the left.
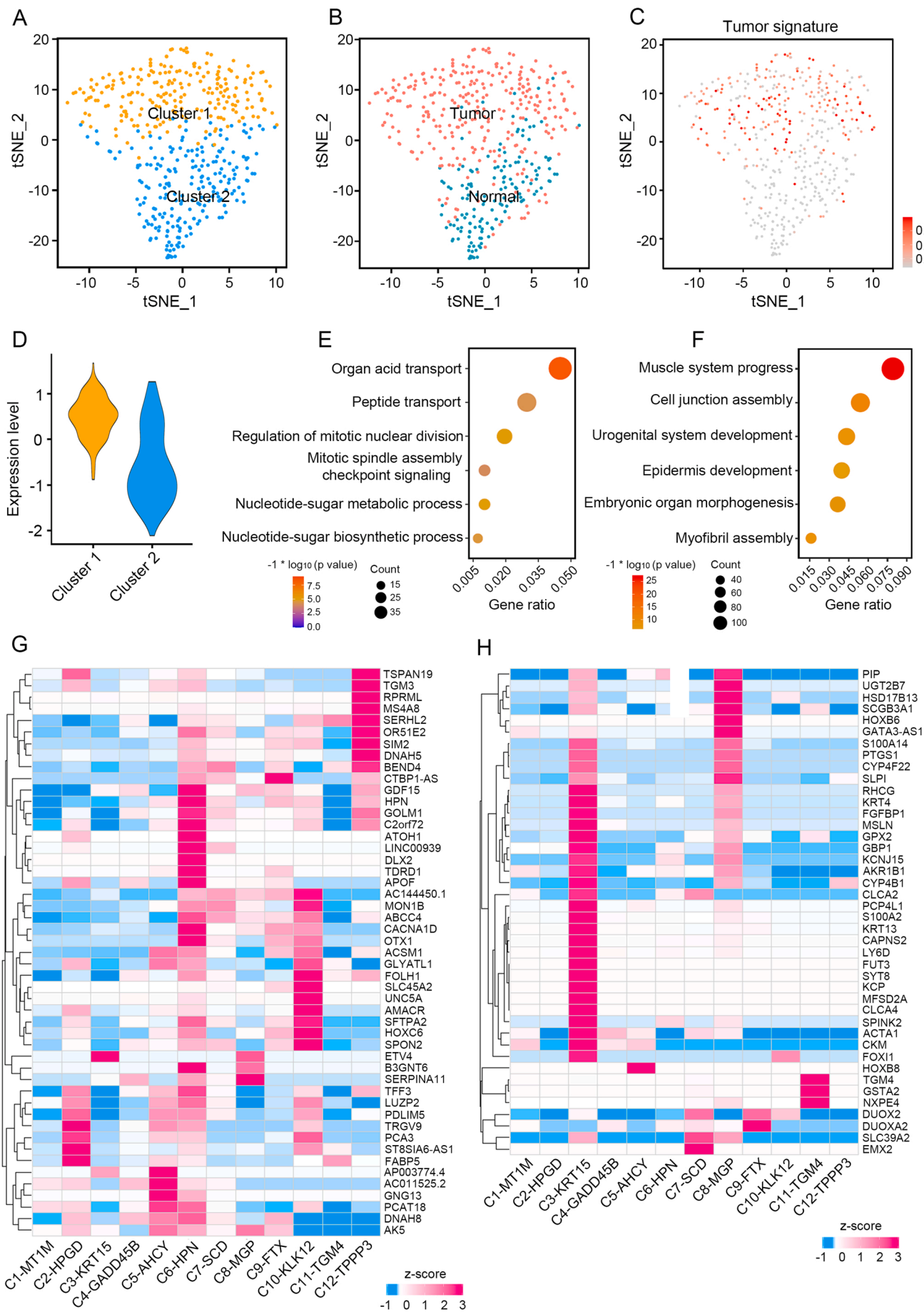
(caption on next page)

Fig. 6. Verification of the diagnostic roles of the 24 candidate genes with bulk and single-cell RNA-seq profiles. (A) Samples from the integrated public dataset clustered according to the expression levels of the 24 candidate genes. (B) Normal and tumorous tissue distribution of the samples shown by tSNE plot. (C) Feature plot and (D) violin plot showing the expression levels of highly expressed PCa genes in clusters 1 and 2, respectively. (E) The biological processes that cluster 1 highly expressed genes were enriched in. (F) The biological processes that cluster 2 highly expressed genes were enriched in. (G) The expression levels of cluster 1 highly expressed genes in each epithelial subpopulation from single-cell RNA-seq profiles. (h) The expression levels of cluster 2 highly expressed genes in each epithelial subpopulation from single-cell RNA-seq profiles.

MKI67 and genes involved in the G2M checkpoint and mitotic spindle [4] (Fig. 2E, F).

Taken together, epithelial cells from PCa consisted of 12 subpopulations with various roles in PCa progression. C1-MT1M, C3-KRT15, C4-GADD45B and C11-TGM4 were likely the normal cells with low expression of PCa progression-related genes and high expression of basal markers or PCa-suppressing genes. Other subpopulations were more likely to be malignant components in PCa. In particular, C6-HPN and C10-KLK12 could be involved in PCa diagnosis considering the highly expressed diagnostic markers.

3.3. Identification of malignant subpopulation

The expression levels of the top subpopulational markers were examined in normal and tumorous prostate samples from the TCGA and GTEx integrated public datasets to determine the clinical correlation to PCa progression. The top markers of C2-HPGD, C5-AHCY, C6-HPN, C9-FTX, C10-KLK12 and C12-TPPP3 were highly expressed in tumorous prostates (Fig. 3A). While the gene expression characteristics of C1-MT1M, C3-KRT15, C4-GADD45B, C7-SCD, C8-MGP and C11-TGM4 were more similar to normal tissues (Fig. 3A). Furthermore, we performed multimodal intersection analysis (MIA) to examine the correlation between the expression signature of each subpopulation and PCa tissues. Consistently, abundant upregulated genes in PCa tissues were also detected in C2-HPGD, C5-AHCY, C6-HPN, C9-FTX, C10-KLK12 and C12-TPPP3 (Fig. 3B).

To further validate the clinical correlation of each subpopulation, we examined the expression levels of PCa-upregulated genes in each subpopulation (Table S3) and found that these genes were mostly expressed in the potentially malignant subpopulations we identified earlier (Fig. 3C). Despite moderate expression of PCa-upregulated genes in cells from normal samples was detected, the positive cell counts were small and the expression levels were relatively low (Fig. 3C, D). Therefore, via the overall considerations of cell proportion by tumorous derivation, the potential biological function and the clinical correlation of each subpopulation, C2-HPGD, C5-AHCY, C6-HPN, C9-FTX, C10-KLK12 and C12-TPPP3 were regarded as malignant subpopulations.

3.4. Screening for the diagnostic subpopulation

The deconvolution of bulk-seq data from the integrated public dataset was performed by BayesPrism, and the cell proportion of each malignant subpopulation in the normal and tumorous samples was calculated. We further performed ROC analysis based on the cell proportion of each malignant subpopulation and the tumor event to examine the capability of predicting PCa occurrence. In contrast, C6-HPN and C12-TPPP3 showed extremely high sensitivity and specificity for predicting PCa, with AUC values higher than 0.9 (Fig. 4A). Furthermore, the cell proportion by tumorous derivation of each subpopulation was examined, and similar results were obtained: the cell proportions of C6-HPN and C12-TPPP3 in PCa samples were significantly higher than those in normal prostates, and the differences were huge (Fig. 4B). Particularly, C6-HPN had a slightly higher AUC value and larger differences between the cell proportions in PCa and normal prostates compared to C12-TPPP3 (Fig. 4A, B).

We further analyzed the cell proportions by patient derivation and found that C2-HPGD was mainly derived from one patient labeled D5 (Fig. 4C). Similarly, C10-KLK12 and C12-TPPP3 mostly originated from

patients D15 and D4, respectively (Fig. 4C). However, C5-AHCY, C6-HPN and C9-FTX were distributed in almost all of the patients, suggesting that they are more universal than other subpopulations (Fig. 4C). The universality of subpopulational distribution in PCa patients was evaluated by the homogeneity and positive ratio, showing that C5-AHCY and C6-HPN were homogeneously distributed in all patients (Fig. 4D). Taken together, C6-HPN was the most universal subpopulation with great predictive capability for PCa, making it a potentially diagnostic subpopulation for PCa occurrence.

3.5. Identification of genes essential for early PCa diagnosis

WGCNA was performed using the integrated public data to identify the tumor-event-correlated gene module. A total of 10,000 genes with the top median absolute deviations were preserved for subsequent analysis. Integrating the correlation analysis of the power value of the soft threshold with the scale-free topology model fit and with the mean connectivity, a soft threshold of 6 was selected (Fig. 5A, B). With this soft threshold, a co-expression network consisting of 18 gene modules was constructed (Fig. 5C, D, Table S4).

To identify the diagnostic genes for early PCa, we examined the correlations of each gene module with the epithelial subpopulations and clinical phenotypes, and found that gene modules 5 and 8 exhibited high correlations to C6-HPN (Fig. 5E). Meanwhile, gene module 5 was also positively and significantly related to Gleason score and tumor event, which was further validated by correlation analysis of genes from module 5 and those highly expressed in early PCa (Fig. 5E, F). The expression levels of gene module 5 and C6-HPN markers were examined in normal and tumorous prostate samples in the TCGA and GTEx datasets. As expected, most of these genes were highly expressed in PCa, and a total of 24 overlapping genes between gene module 5 and C6-HPN markers were filtered out as candidate genes essential for PCa diagnosis (Fig. 5G).

3.6. Molecular diagnosis based on the candidate genes

Normal and tumorous prostate samples were extracted from TCGA and GTEx integrated datasets and divided into 2 clusters according to the expression levels of 24 candidate genes (Fig. 6A). We examined the origin of these samples and found that cluster 1 and cluster 2 closely overlapped with tumorous and normal samples, respectively (Fig. 6B). To further validate that the 24 genes could divide these samples into normal and tumorous prostates, we examined the expression levels of PCa signature genes in these 2 clusters and found that they were highly and specifically expressed in cluster 1, demonstrating that cluster 1 mostly consisted of tumorous samples (Fig. 6C, D). In addition, functional enrichment analysis showed that cluster 1 was mainly involved in the regulation of mitotic nuclear division and mitotic spindle assembly checkpoint signaling, essential for cell proliferation (Fig. 6E). Cluster 2 was mainly related to urogenital system development and epidermis development, suggesting an enrichment of normal tissues in cluster 2 (Fig. 6F).

We next analyzed the expression levels of the marker genes of cluster 1 and cluster 2 in our single-cell RNA-seq data. The cluster 1 top markers were highly expressed in the malignant subpopulations, including C2-HPGD, C5-AHCY, C6-HPN, C9-FTX, C10-KLK12 and C12-TPPP3 (Fig. 6G). Moreover, C6-HPN highly expressed approximately all of the cluster 1 top markers, showing a significant correlation between
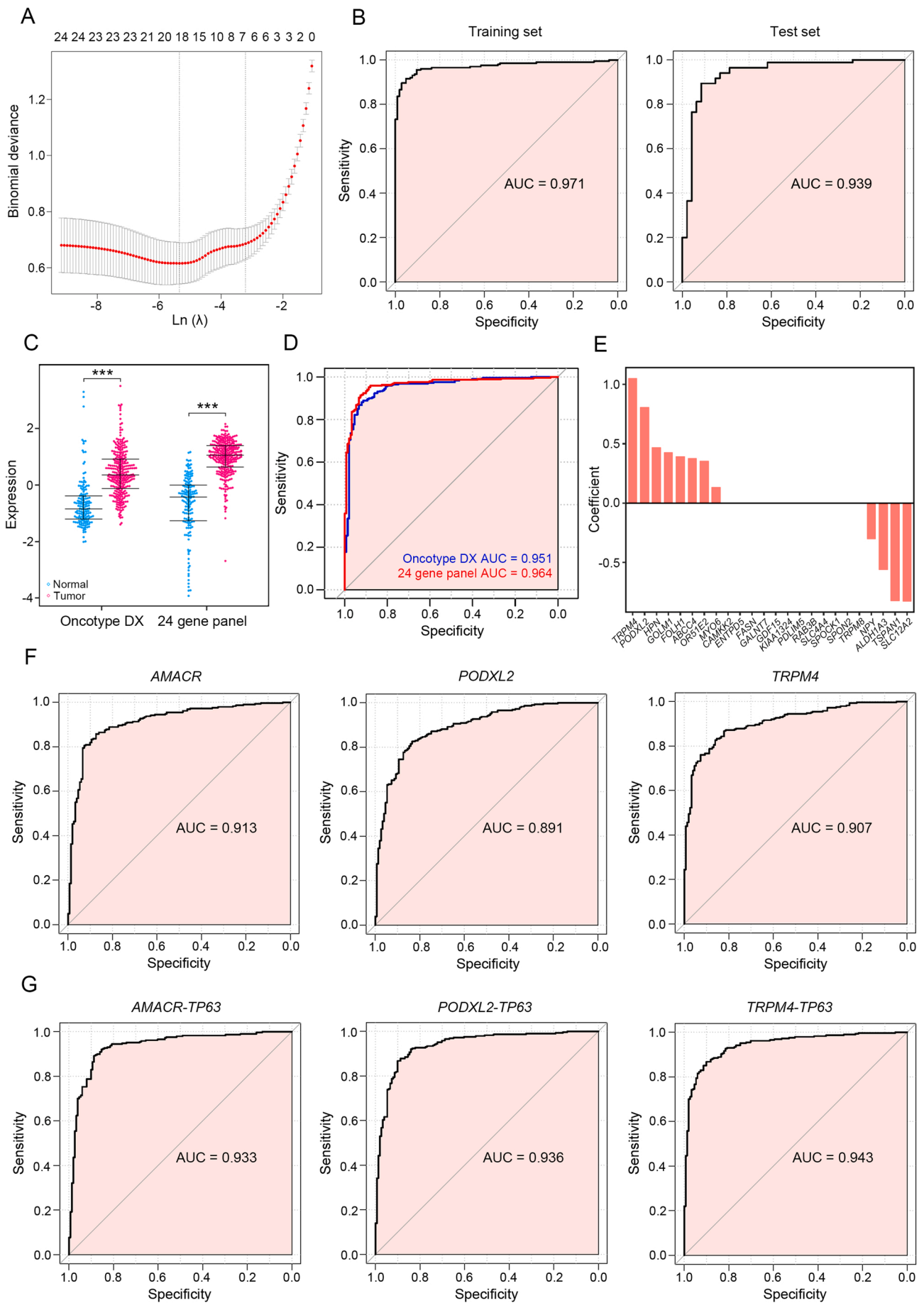
(caption on next page)

Fig. 7. Top risk genes from the 24-gene panel identified by the LASSO model. (A) Binomial deviance under different lambda values. (B) The diagnostic capability of the 24-gene panel in the training and test sets was analyzed by ROC. (C) The expression levels of genes from Oncotype DX and the 24-gene panel in normal and tumorous prostates from the integrated dataset. (D) The diagnostic capability of Oncotype DX and the 24-gene panel evaluated by ROC analysis using the integrated dataset. (E) Regression coefficients of the 24 candidate genes in the LASSO model. (F) The diagnostic capability of AMACR, TRPM4 and PODXL2 analyzed by ROC using the integrated dataset. (G) The diagnostic capability of AMACR, TRPM4 and PODXL2 combined with TP63 individually analyzed by ROC using the integrated dataset.

cluster 1 and C6-HPN (Fig. 6G). However, the cluster 2 top markers were significantly enriched in C3-KRT15, which was earlier identified as basal cells (Fig. 6H). Taken together, the 24 candidate genes obtained from the diagnostic subpopulation and gene module were verified to be capable of identifying early PCa and had enormous potentials for PCa diagnosis.

3.7. Establishment of a diagnostic model for PCa

To establish a risk gene model for early PCa occurrence, we performed LASSO regression on the 24 candidate genes. The optimal lambda of 0.0042 was selected with the binomial deviance tended to be 0 (Fig. 7A). ROC analysis was performed with this gene model in the training set and showed great sensitivity and specificity for PCa diagnosis with an AUC value of 0.971, which was further validated in the test set with an AUC value of 0.939 (Fig. 7B). Furthermore, we examined the expression levels of the 24-gene panel in normal and tumorous samples from the integrated dataset and made a comparison with the commercially used gene panel of Oncotype DX. Both of them were highly expressed in PCa tissues, and significant differences were observed between normal and tumorous samples (Fig. 7C). In comparison, ROC analysis of these two gene panels showed a higher AUC value of our candidate gene panel (AUC = 0.964) than Oncotype DX (AUC = 0.951), suggesting that it is more sensitive and specific in early PCa diagnosis (Fig. 7D).

Furthermore, we examined the regression coefficients of these 24 genes from the LASSO model and identified *TRPM4* and *PODXL2* as the top two critical factors (Fig. 7E). Since *AMACR* is a widely used pathology marker for PCa diagnosis and our candidate gene panel did not contain this marker, we performed ROC analysis and compared the AUC value of *AMACR* to *TRPM4* and *PODXL2* in predicting PCa. *AMACR* showed a slightly higher AUC value than either *TRPM4* or *PODXL2*, suggesting that they were not adequate to replace *AMACR* in PCa diagnosis (Fig. 7F). However, *AMACR* exhibited a slightly lower AUC value than *TRPM4* and *PODXL2* when combined with basal marker *TP63* (Fig. 7G). In summary, *TRPM4* and *PODXL2*, as the top two critical genes for PCa diagnosis from the candidate 24-gene panel, had similar capability for early PCa detection to *AMACR* and even better capability when combined with *TP63*.

3.8. Verification of TRPM4 and PODXL2 in PCa diagnosis

To further validate the diagnostic roles of *TRPM4* and *PODXL2* in PCa, we examined their expression levels in a tissue microarray consisting of 55 normal prostate tissues and 95 PCa tissues with different Gleason scores. Meanwhile, we examined the expression level of *AMACR* in the same section to make a comparison. All of these markers exhibited relatively low expression levels in normal prostates, while in tumorous tissues, *AMACR* and *PODXL2* showed higher expression levels than *TRPM4* (Fig. 8A, B). Moreover, we found that not all tumor cells presented overlapping signals, particularly the *AMACR*-positive cells and those with high expression levels of *TRPM4* and *PODXL2*, which was consistent with our earlier findings that they represented different subpopulations (Fig. 8A). To determine if *TRPM4* and *PODXL2* were qualified markers for early PCa detection, we examined the expression differences of normal and tumorous prostates with various Gleason scores individually. All of these markers exhibited significantly high expression levels in PCa tissues with Gleason scores of 6 and 7 (Fig. 8C).

However, compared with *TRPM4*, the expression differences in early PCa and the corresponding normal tissues were more significant in *PODXL2* and *AMACR* (Fig. 8C).

Based on the expression levels of each candidate marker in the tissue microarray, we performed ROC analysis to examine the sensitivity and specificity for PCa diagnosis. *PODXL2* exhibited the best diagnostic capability for early PCa with an AUC value of 0.941, which was much higher than that of both *AMACR* and *TRPM4* (Fig. 8D). In contrast, *TRPM4* presented a relatively weak capability for PCa diagnosis with an AUC value of 0.782 (Fig. 8D). Overall, *PODXL2* was highly expressed in early PCa and exhibited significant sensitivity and specificity compared with the widely used pathology marker *AMACR*, showing a promising potential for future PCa diagnosis.

4. Discussion

Early PCa identification by prostatic needle biopsy is critical for preventing malignant progression, which currently remains challenging when malignant glands are few or benign mimickers are present. Despite the extensive use of *AMACR* and *TP63* to mark the tumor and basal cells, the overdiagnosis and misdiagnosis rates remain to be high in clinically diagnosed cases. Numerous markers for PCa diagnosis have been identified with the development of sequencing technologies [4,12]. However, the clinical applications for the pathological diagnosis of PCa are still very limited [4]. The reasons for this dilemma could be the low sensitivity and specificity of markers identified from conventional bulk RNA-seq. The gene expression profiles obtained from bulk-seq are usually normalized by various types of cells rather than tumor cells, let alone the heterogeneity of tumor cells [28,29]. In comparison, single-cell RNA-seq measures gene expression at the individual cell level, allowing examination of tumor cells alone or their subpopulations [30,31]. This relatively new technology has been used in a variety of types of tumors to identify distinct subpopulations critical for tumor progression and drug resistance in PCa [32–34]. However, to our knowledge, no diagnostic subpopulation of PCa has ever been reported.

In this study, we integrated single-cell and bulk RNA-seq data from published and public datasets to reveal the intratumoral heterogeneity of PCa and identified a distinct subpopulation, C6-HPN, with high expression levels of genes related to early PCa development, such as *GOLM1*, *HPN*, *TFF3*, *TMPRSS2* and *KLK3*. GOLM1 is a Golgi transmembrane protein highly expressed in PCa and critical for TGFβ1/Smad2 signaling-induced epithelial-mesenchymal transition (EMT) [35,36]. HPN is a type II transmembrane serine protease overexpressed in early PCa, targeting which could significantly suppress PCa invasion and metastasis [37,38]. TFF3 has a non-CpG island promoter with significant hypomethylation in PCa compared to normal prostates, and the overexpression of mRNA levels can be found in most PCa samples, exhibiting a diagnostic potential for early PCa [39]. TMPRSS2-ERG gene fusion has long been identified as a specific genomic aberration in PCa and has recently been developed as a promising urine marker for early PCa detection [40,41]. KLK3 is the coding gene of PSA and has been widely used for early PCa screening, although its low specificity has caused abundant overdiagnosis and overtreatment [42]. Furthermore, we performed deconvolution to obtain the cell proportion in normal and tumorous prostates and calculated the AUC values for identifying PCa. In particular, C6-HPN exhibited the highest AUC value of 0.926, suggesting great sensitivity and specificity to diagnose early PCa, and was further considered the diagnostic subpopulation.
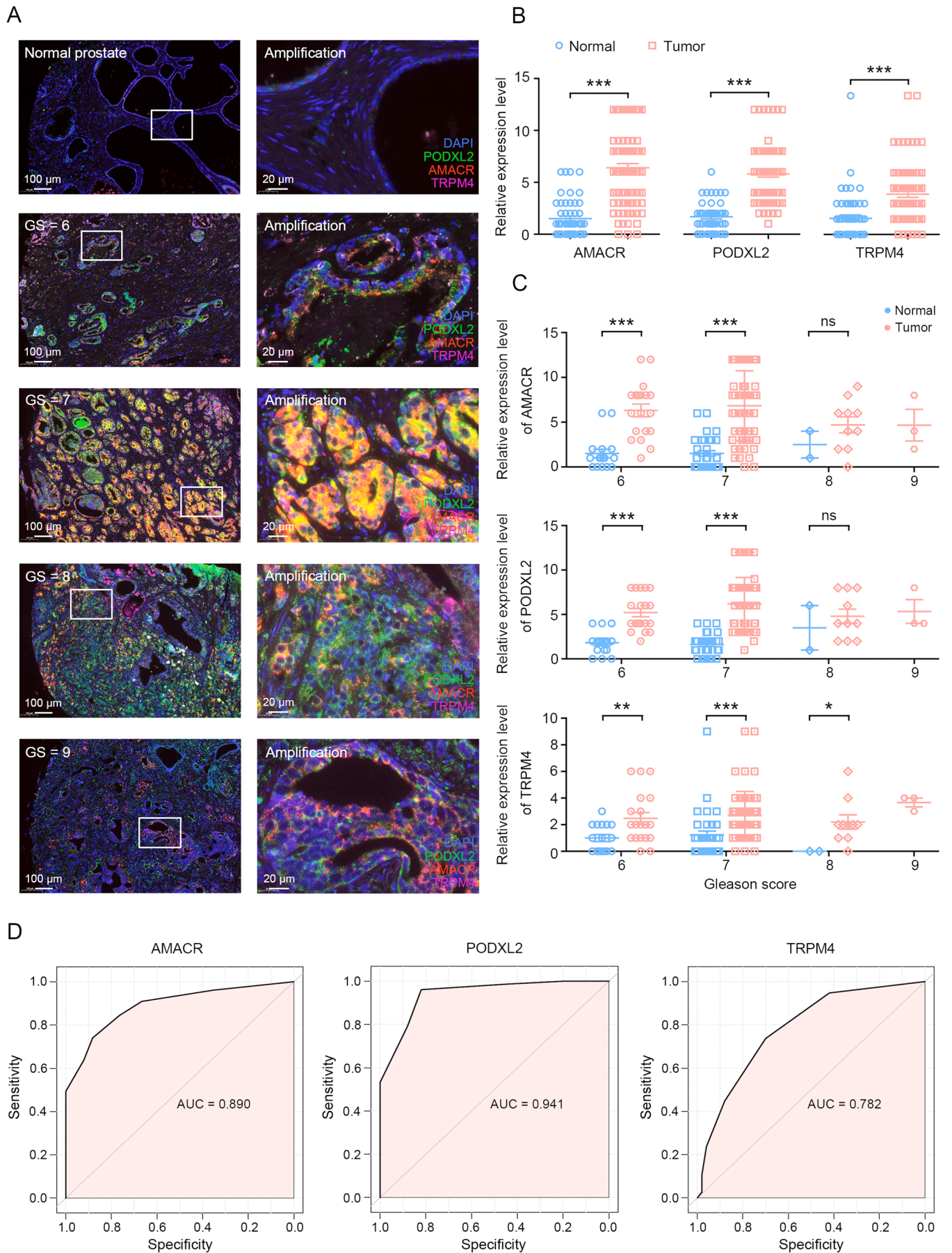
(caption on next page)

Fig. 8. The expression levels of candidate genes for PCa diagnosis examined by multiplex immunostaining. (A) Multiplex immunostaining of AMACR, TRPM4 and PODXL2 in a PCa tissue microarray. Blue fluorescence indicates nuclei stained with DAPI. Green fluorescence indicates PODXL2-positive signals, red fluorescence indicates AMACR-positive signals, and purple fluorescence indicates TRPM4-positive signals. (B) The expression levels of AMACR, TRPM4 and PODXL2 in all normal and tumorous prostate tissues from the tissue microarray. (C) The expression levels of AMACR, TRPM4 and PODXL2 in normal and PCa tissues with different Gleason scores from the tissue microarray. Values are presented as the mean ± SD, ns indicates no significance, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (D) The diagnostic capability of *AMACR*, *TRPM4* and *PODXL2* analyzed by ROC based on the expression levels in normal and tumorous prostate tissues from a tissue microarray.

Most of the markers from the diagnostic subpopulation showed enormous potentials in early PCa detection. To determine the most critical markers for PCa diagnosis, we performed WGCNA to identify a PCa occurrence-related gene module and filtered out the overlapping genes with the C6-HPN markers to construct a LASSO model. The obtained 24-gene panel exhibited significant expression differences between normal and tumorous prostates. To further explore the potential applications in PCa diagnosis, we compared the sensitivity and specificity of this gene panel with the commercially used gene panel of Oncotype DX using the integrated dataset. Oncotype DX is a widely used gene panel for the prediction of a variety of malignancies, including PCa, which contains a total of 12 tumor progression-related genes, specifically *TPX2*, *AZGP1*, *KLK2*, *SRD5A2*, *FAM13C*, *FLNC*, *GSN*, *TPM2*, *GSTM2*, *BGN*, *COL1A1* and *SFRP4* [43,44]. Therefore, more significant roles in early PCa detection exhibited by the 24-gene panel suggested that it could be a more sensitive and specific tissue-based gene panel for future PCa diagnosis.

From the candidate gene panel, *TRPM4* and *PODXL2* were identified as the top two risk factors based on the regression coefficient. TRPM4 is a calcium-activated monovalent-selective cation channel overexpressed in various types of malignancies, including breast cancer, endometrial cancer, cervical cancer, colorectal cancer and prostate cancer [45–49]. Particularly in PCa, several lines of studies have shown that *TRPM4* is significantly involved in regulating EMT and associated with biochemical recurrence following surgery [50,51]. However, the diagnostic role of *TRPM4* in early PCa has yet to be reported. Here, we found that *TRPM4* was highly expressed in tumorous prostates compared to normal tissues and exhibited great diagnostic capability for early PCa, as determined by a public dataset and immunostaining of over 100 samples. PODXL2 is a transmembrane glycoprotein of the CD34 family significantly correlated with poor prognosis of breast cancer and could promote tumor progression via the RAC1/AKT pathway [52]. No further correlations of *PODXL2* with other malignancies have been reported, including PCa. In this study, we examined the expression characteristics in normal and tumorous prostates and found that *PODXL2* was specifically and highly expressed in early PCa with Gleason scores of 6 and 7. The diagnostic capability of early PCa evaluated by ROC analysis showed that *PODXL2* is more sensitive and specific than *TRPM4* and is even a more qualified marker than the widely used marker *AMACR*.

5. Conclusion

In summary, we identified a diagnostic subpopulation for early PCa with the integrating analysis of single-cell and bulk RNA-seq profiles. Furthermore, a diagnostic gene module generated from WGCNA was used to perform intersection analysis with the marker genes of the diagnostic subpopulation. A total of 24 overlapping genes were obtained, based on which a LASSO model was constructed and showed better diagnostic capability for early PCa than the commercially used gene panel of Oncotype DX. The top two risk factors from the 24-gene panel, *TRPM4* and *PODXL2*, were identified based on the regression coefficient and showed similar capability of PCa diagnosis verified by the integrated public dataset. The expression characteristics of these two genes were further validated by multiplex immunostaining on a PCa tissue microarray. Both *TRPM4* and *PODXL2* were highly expressed in PCa tissues, but in comparison, *PODXL2* showed more sensitivity and specificity for PCa diagnosis than *TRPM4* and was even a better marker than the pathologically used marker *AMACR*. Taken together, our

findings have provided a novel and promising diagnostic marker for early PCa detection, while the roles in tumor occurrence remain to be further investigated.

CRediT authorship contribution statement

Xiaoshi Ma and Lipeng Chen analyzed the scRNA-seq data. Tao Chen performed the multiplex immunostaining and the statistical analysis of immunostaining scores in various groups. Kun Chen constructed the LASSO model and performed the ROC analysis. Huirong Zhang revised the manuscript with additional analysis. Kaipeng Huang and Han Zheng performed WGCNA with bulk RNA-seq. Hongtao Jin and Zhiqiang Cheng evaluated the immunostaining scores. Hongtao Jin, Zhiqiang Cheng, Kefeng Xiao and Jinan Guo designed the project. Jinan Guo supervised this project. Xiaoshi Ma integrated the data and wrote the manuscript. All authors reviewed the manuscript.

As Huirong Zhang did most of the revision work of this manuscript, including the additional analysis of PODXL2 in identifying PCa with PSA in gray-zone and the cellular derivation of each patient. We decided to add her as an author of this article. This decision was approved by all of the authors.

Declaration of Competing Interest

The authors declare that no competing interest exists.

Data availability

All the data supporting the conclusions of this study are available within the article.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.10.044.

References

[1] Sung H, Ferlay J, Siegel RL. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.

[2] Merriel SWD, Funston G, Hamilton W. Prostate cancer in primary care. Adv Ther 2018;35:1285–94.

[3] Mottet N, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, et al. EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. Part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol 2021;79:243–62.

[4] Rodrigues Â, Freitas R, Nogueira-Silva P, Jerónimo C, Henrique R. Biopsy sampling and histopathological markers for diagnosis of prostate cancer. Expert Rev Anticancer Ther 2014;14:1323–36.

[5] Paner GP, Luthringer DJ, Amin MB. Best practice in diagnostic immunohistochemistry: prostate carcinoma and its mimics in needle core biopsies. Arch Pathol Lab Med 2008;132:1388–96.

[6] Hameed O, Humphrey PA. Immunohistochemistry in diagnostic surgical pathology of the prostate. Semin Diagn Pathol 2005;22:88–104.

[7] Ali TZ, Epstein JI. False positive labeling of prostate cancer with high molecular weight cytokeratin: p63 a more specific immunomarker for basal cells. Am J Surg Pathol 2008;32:1890–5.

[8] Molinié V, Hervé JM, Lugagne PM, Lebret T, Botto H. Diagnostic utility of a p63/ alpha-methyl coenzyme A racemase (p504s) cocktail in ambiguous lesions of the prostate upon needle biopsy. BJU Int 2006;97:1109–15.

[9] Boran C, Kandirali E, Yilmaz F, Serin E, Akyol M. Reliability of the 34βE12, keratin 5/6, p63, bcl-2, and AMACR in the diagnosis of prostate carcinoma. Urol Oncol 2011;29:614–23.

[10] Gurel B, Ali TZ, Montgomery EA, Begum S, Hicks J, Goggins M, et al. NKX3.1 as a marker of prostatic origin in metastatic tumors. Am J Surg Pathol 2010;34: 1097–105.

[11] Li W, Wang X, Li B, Lu J, Chen G. Diagnostic significance of overexpression of Golgi membrane protein 1 in prostate cancer. Urology 2012;80 (952.e951-957).

[12] Kristiansen G. Diagnostic and prognostic molecular biomarkers for prostate cancer. Histopathology 2012;60:125–41.

[13] Chen S, Zhu G. Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. Nat Cell Biol 2021; 23:87–98.

[14] Zheng H, Pomyen Y, Hernandez MO, Li C, Livak F, Tang W, et al. Single-cell analysis reveals cancer stem cell heterogeneity in hepatocellular carcinoma. Hepatology 2018;68:127–40.

[15] Aissa AF, Islam A. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. Nat Commun 2021;12: 1628.

[16] Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ, et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell 2021;39:1150–62. e1159.

[17] Ma X, Guo J, Liu K, Chen L, Liu D, Dong S, et al. Identification of a distinct luminal subgroup diagnosing and stratifying early stage prostate cancer by tissue-based single-cell RNA sequencing. Mol Cancer 2020;19:147.

[18] Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. Gigascience 2020;9.

[19] Wolock SL, Lopez R, Klein AM. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. Cell Syst 2019;8:281–91. e289.

[20] Hao Y, Hao S, Andersen-Nissen E, Mauck 3rd WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell 2021;184:3573–87. e3529.

[21] Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012;28:882–3.

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

[23] Chu T, Wang Z, Pe'er D. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. Nat Cancer 2022;3:505–17.

[24] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinforma 2011;12:77.

[25] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinforma 2008;9:559.

[26] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):22.

[27] Yoon HY, Kim YW, Kang HW, Kim WT, Yun SJ, Lee SC, et al. DNA methylation of GSTP1 in human prostate tissues: pyrosequencing analysis. Korean J Urol 2012;53: 200–5.

[28] Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science 2011;331:1565–70.

[29] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5:621–8.

[30] Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD. Eleven grand challenges in single-cell data science. Genome Biol 2020;21:31.

[31] Eberwine J, Sul JY, Bartfai T, Kim J. The promise of single-cell sequencing. Nat Methods 2014;11:25–7.

[32] Song H, Weinstein HNW. Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. Nat Commun 2022;13:141.

[33] Cheng Q, Butler W, Zhou Y, Zhang H, Tang L, Perkinson K, et al. Pre-existing castration-resistant prostate cancer-like cells in primary prostate cancer promote resistance to hormonal therapy. Eur Urol 2022;81:446–55.

[34] Karthaus WR, Hofree M. Regenerative potential of prostate luminal cells revealed by single-cell analysis. Science 2020;368:497–505.

[35] Qin X, Liu L, Li Y, Luo H, Chen H, Weng X. GOLM1 promotes epithelial-mesenchymal transition by activating TGFβ1/Smad2 signaling in prostate cancer. Technol Cancer Res Treat 2023;22:15330338231153618.

[36] Varambally S, Laxman B, Mehra R, Cao Q, Dhanasekaran SM, Tomlins SA, et al. Golgi protein GOLM1 is a tissue and urine biomarker of prostate cancer. Neoplasia 2008;10:1285–94.

[37] Kelly KA, Setlur SR, Ross R, Anbazhagan R, Waterman P, Rubin MA, et al. Detection of early prostate cancer using a hepsin-targeted imaging agent. Cancer Res 2008;68:2286–91.

[38] Hudson BD, Kulp KS, Loots GG. Prostate cancer invasion and metastasis: insights from mining genomic data. Brief Funct Genom 2013;12:397–410.

[39] Nørgaard M, Haldrup C, Storebjerg TM, Vestergaard EM, Wild PJ, Høyer S, et al. Comprehensive evaluation of TFF3 promoter hypomethylation and molecular biomarker potential for prostate cancer diagnosis and prognosis. Int J Mol Sci 2017;18.

[40] Fujita K, Nonomura N. Urinary biomarkers of prostate cancer. Int J Urol 2018;25: 770–9.

[41] Salagierski M, Schalken JA. Molecular diagnosis of prostate cancer: PCA3 and TMPRSS2:ERG gene fusion. J Urol 2012;187:795–801.

[42] Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, et al. EAU guidelines on prostate cancer. Part II: treatment of advanced, relapsing, and castration-resistant prostate cancer. Eur Urol 2014;65:467–79.

[43] Basourakos SP, Tzeng M, Lewicki PJ, Patel K, Al Hussein Al Awamlh B, Venkat S, et al. Tissue-based biomarkers for the risk stratification of men with clinically localized prostate cancer. Front Oncol 2021;11:676716.

[44] Cullen J, Rosner IL, Brand TC, Zhang N, Tsiatis AC, Moncur J, et al. A biopsy-based 17-gene genomic prostate score predicts recurrence after radical prostatectomy and adverse surgical pathology in a racially diverse population of men with clinically low- and intermediate-risk prostate cancer. Eur Urol 2015;68:123–31.

[45] Kappel S, Stokłosa P, Hauert B, Ross-Kaschitza D, Borgström A, Baur R, et al. TRPM4 is highly expressed in human colorectal tumor buds and contributes to proliferation, cell cycle, and invasion of colorectal cancer cells. Mol Oncol 2019;13: 2393–405.

[46] Armisén R, Marcelain K, Simon F, Tapia JC, Toro J, Quest AF, et al. TRPM4 enhances cell proliferation through up-regulation of the β-catenin signaling pathway. J Cell Physiol 2011;226:103–9.

[47] Hong X, Yu JJ. MicroRNA-150 suppresses epithelial-mesenchymal transition, invasion, and metastasis in prostate cancer through the TRPM4-mediated β-catenin signaling pathway. Am J Physiol Cell Physiol 2019;316. C463-c480.

[48] Rivas J, Díaz N, Silva I, Morales D, Lavanderos B, Álvarez A, et al. KCTD5, a novel TRPM4-regulatory protein required for cell migration as a new predictor for breast cancer prognosis. FASEB J 2020;34:7847–65.

[49] Narayan G, Bourdon V, Chaganti S, Arias-Pulido H, Nandula SV, Rao PH, et al. Gene dosage alterations revealed by cDNA microarray analysis in cervical cancer: identification of candidate amplified and overexpressed genes. Genes Chromosomes Cancer 2007;46:373–84.

[50] Sagredo AI, Sagredo EA. TRPM4 channel is involved in regulating epithelial to mesenchymal transition, migration, and invasion of prostate cancer cell lines. J Cell Physiol 2019;234:2037–50.

[51] Berg KD, Soldini D, Jung M, Dietrich D, Stephan C, Jung K, et al. TRPM4 protein expression in prostate cancer: a novel tissue biomarker associated with risk of biochemical recurrence following radical prostatectomy. Virchows Arch 2016;468: 345–55.

[52] Lin Y-Y, Wang C-Y, Phan N, Chiao C-C, Li C-Y, Sun Z, et al. PODXL2 maintains cellular stemness and promotes breast cancer development through the Rac1/Akt pathway. Int J Med Sci 2020;17:1639–51.