

RESEARCH

Open Access



Identification of a distinct luminal subgroup diagnosing and stratifying early stage prostate cancer by tissue-based single-cell RNA sequencing

Xiaoshi Ma^{1†}, Jinan Guo^{1,2†}, Kaisheng Liu^{1†}, Lipeng Chen¹, Dale Liu¹, Shaowei Dong¹, Jinquan Xia¹, Qiaoyun Long¹, Yongjian Yue¹, Pan Zhao^{1,2}, Fengyan Hu¹, Zhangang Xiao³, Xinghua Pan^{4,5,6}, Kefeng Xiao¹, Zhiqiang Cheng^{1,2}, Zunfu Ke^{7*}, Zhe-Sheng Chen^{8*}  and Chang Zou^{1,2,3*}

Abstract

Background: The highly intra-tumoral heterogeneity and complex cell origination of prostate cancer greatly limits the utility of traditional bulk RNA sequencing in finding better biomarker for disease diagnosis and stratification. Tissue specimens based single-cell RNA sequencing holds great promise for identification of novel biomarkers. However, this technique has yet been used in the study of prostate cancer heterogeneity.

Methods: Cell types and the corresponding marker genes were identified by single-cell RNA sequencing. Malignant states of different clusters were evaluated by copy number variation analysis and differentially expressed genes of pseudo-bulks sequencing. Diagnosis and stratification of prostate cancer was estimated by receiver operating characteristic curves of marker genes. Expression characteristics of marker genes were verified by immunostaining.

Results: Fifteen cell groups including three luminal clusters with different expression profiles were identified in prostate cancer tissues. The luminal cluster with the highest copy number variation level and marker genes enriched in prostate cancer-related metabolic processes was considered the malignant cluster. This cluster contained a distinct subgroup with high expression level of prostate cancer biomarkers and a strong distinguishing ability of normal and cancerous prostates across different pathology grading. In addition, we identified another marker gene, Hepsin (*HPN*), with a 0.930 area under the curve score distinguishing normal tissue from prostate cancer lesion. This finding was further validated by immunostaining of HPN in prostate cancer tissue array.

(Continued on next page)

* Correspondence: kezunfu@mail.sysu.edu.cn; chenz@stjohns.edu; zou.chang@szhospital.com

[†]Xiaoshi Ma, Jinan Guo and Kaisheng Liu contributed equally to this work.

⁷Department of Pathology, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, Guangdong, China

⁸Department of Pharmaceutical Sciences, College of Pharmacy and Health Sciences, St. John's University, Queens, New York, USA

¹Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, Guangdong, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusion: Our findings provide a valuable resource for interpreting tumor heterogeneity in prostate cancer, and a novel candidate marker for prostate cancer management.

Keywords: Prostate cancer, Single-cell RNA sequencing, Diagnosis and stratification biomarker, HPN

Introduction

Prostate cancer (PCa) is the most common malignant cancer apart from lung cancer in males worldwide, with an estimated 1.3 million new cases and 359,000 associated deaths in 2018 [1]. Comprehensive treatments, including radical prostatectomy, radiotherapy and hormone therapy, have been used for many years and contribute to the increase of the total survival rate of PCa patients [2]. However, advanced PCa such as metastatic PCa and castration-resistant prostate cancer (CRPC) remains largely incurable due to the lack of adequate therapies [3]. Since the early stage localized PCa is very treatable with the combined therapy of surgery and hormonal drugs, it is necessary to identify more specific and sensitive markers for early detection or stratification of PCa so as to not delay optimal and appropriate treatment.

PCa diagnosis, stratification and treatment selection is usually based on multiple factors such as serum prostate specific antigen (PSA) level, tumor size, and pathology grading [4]. However, PSA levels can also be detected in benign prostatic hyperplasia (BPH), prostatitis, and even after digital rectal examination (DRE), resulting in a high rate of overdiagnosis and overtreatment of PCa [5]. DRE is conducted based on the shape, symmetry, firmness and nodularity of prostate tumor, and is more suitable for higher stage and advanced PCa [6]. Imaging tools, including transrectal ultrasound (TRUS), computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET), play important roles in clinical diagnosis and stratification of PCa [7]. Despite this, they are still not as sensitive or accurate as histological observation with the detection of specific prostatic marker genes.

With the advances in genomic and transcriptome sequencing technologies, more and more molecular aberrations and potential marker genes for PCa initiation and progression have been revealed. Androgen plays a pivotal role in the growth of hormone sensitive PCa. Androgen receptor (*AR*) gene mutation, amplification and splice variants are frequent *AR* alterations in CRPC, driving PCa growth in a ligand-independent way [8]. Transmembrane Serine Protease 2-Erythroblast Transformation Specific Related Gene (*TMPRSS2-ERG*) gene fusion is a very common genomic alteration in PCa, and can be detected in almost half of Caucasian PCa patients [9]. Genomic losses of tumor suppressor genes such as Cyclin Dependent Kinase Inhibitor 1B (*CDKN1B*),

Phosphatase and Tensin Homolog (*PTEN*), Retinoblastoma Transcriptional Corepressor 1 (*RBI*), Tumor Protein P53 (*TP53*) and genomic gains of oncogenes like Cyclin D1 (*CCND1*), Fibroblast Growth Factor Receptor 1 (*FGFR1*), Myelocytomatosis (*MYC*) are frequently identified in primary PCa [10, 11]. Loss-of-function mutations in DNA double strand breaks (DBS) repair-related genes such as Ataxia Telangiectasia Mutated (*ATM*), Breast Cancer 1/2 (*BRCA1/2*), Cyclin Dependent Kinase 12 (*CDK12*) are detected more prevalently in metastatic PCa than localized PCa, suggesting critical roles of these genes in the transition towards metastatic state [12]. Despite these findings, specific biomarkers still need to be identified for early diagnosis and precise stratification of PCa.

PCa is a highly heterogeneous tumor composed of various types of cells such as epithelial cells, fibroblasts, muscle cells, and immune cells [13]. Conventional bulk RNA sequencing (RNA-seq) averages the transcriptional profiles of all cells within a sample. This sequencing strategy might result in missing some important differences in gene expression in cancer cells. On the other hand, bulk RNA-seq was not capable of identifying and analyzing the rare but critical populations indicative of tumor progression, such as cancer stem cells. Therefore, it might not be an ideal tool in identifying precise biomarker of highly heterogeneous tumors especially for PCa. Single-cell RNA sequencing (scRNA-seq) is a powerful tool for characterizing the transcriptional profiles in thousands of individual cells. It allows for unbiased assessment of heterogeneous cell populations at the single-cell level, thus revealing the differences between individual types within one cell population. A number of studies have explored tumor heterogeneity, evolutionary lineages and detection of rare subpopulations with the use of scRNA-seq [14, 15]. Peng et al. revealed the tumor heterogeneity of pancreatic ductal adenocarcinoma and discovered a subgroup in ductal cells with unique proliferative features [16]. Young et al. showed that childhood Wilms tumor cells were aberrant fetal cells and adult renal carcinoma cells were derived from a rare kind cells called PT1 [17]. However, due to the low viability of PCa cells after tissue lysis, scRNA-seq of PCa has yet to be reported, resulting in a void in the study of PCa heterogeneity.

Here, we performed scRNA-seq in 7904 cells from two patients with primary PCa to investigate the heterogeneity

of PCa at a single-cell level. We found that PCa cells in the tissue consisted mostly of three different types of luminal cells, the marker genes of which were mainly involved in metabolic process, regulation of cell adhesion and vasculogenesis. Copy number variation (CNV) and pseudotime trajectory analysis suggested that type 1 luminal cells had the most malignant characteristics within the 3 types of luminal cells. In addition, five cell subgroups in type 1 luminal cells were further identified by featured gene expression profiles. The fifth cell subgroup, characterized by high expression of PCa marker genes, might be critical for PCa diagnosis and stratification. Furthermore, we found a specific marker for subgroup 5, Hepsin (*HPN*), with a much higher Area Under the Curve (AUC) score than those of widely used markers in the clinic, such as Alpha-Methylacyl-CoA Racemase (*AMACR*), Folate Hydrolase 1 (*FOLH1*), Kallikrein Related Peptidase 3 (*KLK3*) and Prostate Cancer Associated 3 (*PCA3*). These findings were further validated by immunostaining in PCa tissue array, which showed that the protein expression level of *HPN* in PCa with a Gleason score > 6 was significantly higher than those with a Gleason score = 6. Our findings revealed the heterogeneity of PCa at a single-cell level and a distinct luminal subpopulation with high expression level of *HPN*, a potential marker for PCa diagnosis and stratification.

Methods

Samples collection

Two PCa patients were diagnosed with primary PCa with Gleason scores of 6 (3 + 3) and 7 (3 + 4) by morphology observation and immunostaining analysis of PCa markers, including *AMACR* and Tumor Protein P63 (*TP63*). Samples were obtained from these patients by radical prostatectomy with the guidance of an experienced pathologist, and immediately transported to research facility on ice. No treatment was conducted on both patients before surgery. After washed in pre-cooled phosphate buffer saline (PBS) to remove blood contaminations, samples were dissected to pieces (< 10 mm³) for pathological examination and single-cell suspension preparation. Patient consent was obtained for the study and the sample collection was under ethical approval.

Single-cell suspension preparation

Most dissected samples were further cut into small pieces (< 1 mm³) and digested in a lyase cocktail collagenase/hyaluronidase/dispase solution (Fisher Scientific, cat. no. NC-9694308 and 354,235) at 37 °C with a shaking speed of 40 rpm for 60 min. Sample dissociation solutions were filtered by a 40-µm cell strainer prior to the removal of red blood cells with a RBC lysis buffer (Thermo Fisher Scientific, cat. no. 1966634). The single-cell suspension was stained with 0.4% trypan blue

(Thermo Fisher Scientific, cat. no. T10282) to examine the concentration of live cells. Live cells were further concentrated using a Dead Cell Removal Kit (Miltenyi Biotec, cat. no. 130-090-101) if the cell viability was lower than 80%.

Single-cell RNA sequencing and data processing

The single-cell suspension was adjusted to the required concentration of 700–1200 cells/µl and loaded onto the 10 × Genomics single-cell-A chip for a target capture of 3000–5000 cells/chip. The cDNA library was prepared according to the standard manufacturer's protocol from 10 × Genomics Single Cell 3' v2 Reagent Kit and then sequenced on a HiSeq 2500 instrument by SequMed Bio Technology Inc., Guangzhou, China.

Raw sequencing data was processed by Cell Ranger (10 × Genomics, version 3.0.2) pipeline and aligned to the human reference genome (GRCh38). Gene-Barcode matrices containing the barcoded cells and gene expression counts were imported into the Seurat R toolkit (version 3.1.2) [18]. Cells with small library size (< 200) or high mitochondrial transcript ratio (> 0.4), and genes expressed in less than 3 cells were all excluded. For the remaining cells, gene expression matrices were normalized using `NormalizeData` function in Seurat package to eliminate the effects of sequencing depth or library size, and were scaled with the `ScaleData` function to gain linear conversion [18]. Gene expression matrices from different samples were then integrated and the batch effects were removed by canonical correlation analysis and mutual nearest neighbors-anchors using Seurat package [18].

Identification of cell types and marker genes

Highly variable genes (top 2000) were extracted to perform the principal component analysis (PCA) and top 20 of significant principle components were used for cluster analysis. Clusters were visualized using the Uniform Manifold Approximation and Projection (UMAP). Cell type identities were characterized based on the expression of known markers in the Cell Marker database [19]. In addition, sub-clustering of luminal cells was performed with the same method as described above.

Marker genes for each cluster and subgroup were identified by contrasting gene expression of cells from certain cluster or subgroup to that of others using the Seurat `FindMarkers` function, and filtered by a detectable expression in more than 50% of all cells from that cluster or subgroup. Additionally, the expression fold change of marker genes in certain cluster or subgroup to others and the difference of detectable expression in that cluster or subgroup with others were both required to be in top 10 of all detected genes.

CNV analysis

Initial CNVs were estimated by the expression levels of genes within each chromosome region using inferCNV R package [20]. The relative expression values of analyzed genes were limited to $[-1, 1]$. We considered other cells apart from luminal cells to be non-malignant cells and used their average estimated CNV as background to identify which cluster of luminal cells was malignant.

Differentially expressed genes (DEGs) analysis of pseudo-bulks sequencing

Pseudo-bulks were generated by randomly combining every 10 cells in one cell type, the average gene expression level of which was regarded as its gene expression profile. The DEGs were identified using edgeR package and filtered by $|\log_2\text{FoldChange}| \geq 1$ and $\text{FDR} < 0.05$. Gene Ontology (GO) enrichment analysis was performed by examining the DEG involved in biological process and conducting the Fisher exact test.

Pseudotime trajectory analysis

DEG of epithelial cells, including basal cell, luminal cell and neuroendocrine cell, were identified by Seurat. Genes expressed in at least 10 cells were kept and imported into Monocle2 R package. Top 1000 DEG of epithelial cells were used to perform the dimension reduction and construct the pseudotime trajectory. We identified the genes that varied according to pseudotime by using the “differentialGeneTest” function in Monocle2 and used them to perform GO enrichment.

Clinical relevance

To estimate the individual role of each subgroup of type 1 luminal cells in PCa development, we extracted a 50-gene set for each type of luminal cells in this study and RNA-seq data of PCa with different Gleason scores from TCGA, then analyzed their correlation by creating a hierarchical clustering heatmap with ComplexHeatmap R package [21]. The predictive sensitivity and specificity of selected marker genes of luminal cells for PCa progression was assessed with receiver operating characteristic (ROC) curves created by SPSS 20.0. To determine the correlation of candidate marker genes with the recurrence-free rates of PCa patients from TCGA, Kaplan-Meier curves were constructed according to the gene expression profiles after normalizing the expression of each gene to an average expression of 1 in PCa samples.

Immunostaining

We used Tyramide Signal Amplification system for chromogenic immunostaining. Prostate tissue sections were incubated in 3% H_2O_2 for 10 min at room temperature to block the endogenous peroxidase, after which they were incubated with primary antibody

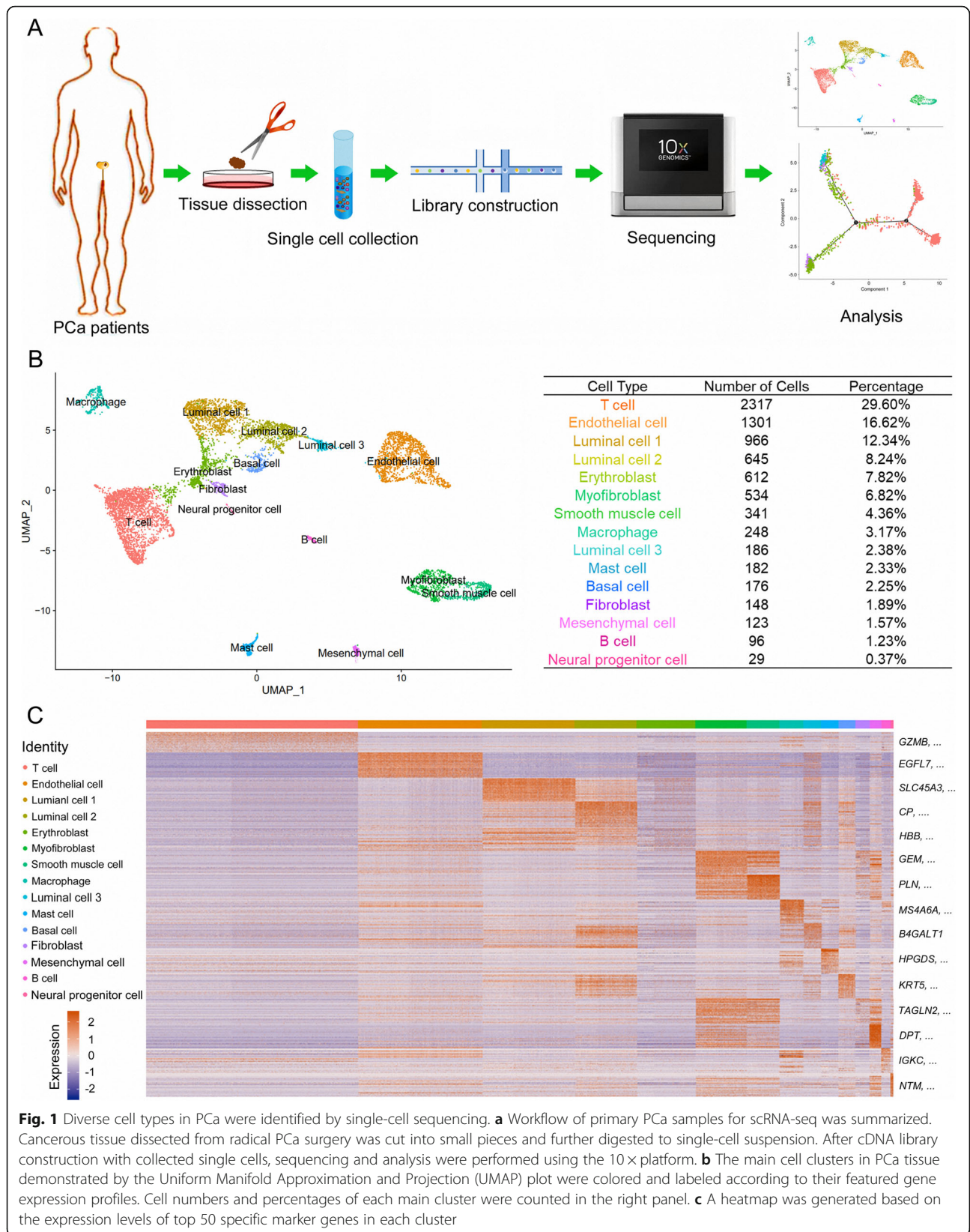
SLC45A3 (diluted at 1:200, Abcam, cat. no. ab137065) for 2 h at room temperature. The sections were then incubated with HRP-linked secondary antibody (Abcam, cat. no. ab7090) and stained with Alexa Fluor™ 488-labeled tyramide (Thermo Fisher Scientific, cat. no. B40953). For multiplex with primary antibodies, the sections were placed in citrate buffer and heated in a microwave for 15 min to release the antibodies. The sections were then rinsed in PBS and subsequently incubated with primary antibody CP (diluted at 1:200, Abcam, cat. no. ab48614), HRP-linked secondary antibody, Alexa Fluor™ 555-labeled tyramide (Thermo Fisher Scientific, cat. no. B40955). After the second round of heating and washing as described above, the sections were incubated with primary antibody B4GALT1 (diluted at 1:200, Abcam, cat. no. ab121326), HRP-linked secondary antibody, Alexa Fluor™ 647-labeled tyramide (Thermo Fisher Scientific, cat. no. B40958) and counterstained with DAPI.

PCa tissue array (Shanghai Outdo Biotech Co. Ltd., Shanghai, China) containing 55 normal prostate tissues and 95 PCa tissues with Gleason scores from 6 to 9 was used to perform immunostaining for the detection of HPN expression in normal prostate and cancerous prostate with different Gleason scores. PCa tissue chip was incubated in 3% H_2O_2 for 10 min at room temperature to block endogenous peroxidase and immersed in citrate buffer at 95 °C for 40 min for antigen retrieval, after which it was incubated with primary antibody HPN (diluted at 1:200, Abcam, cat. no. ab73133) for 2 h at room temperature. Thereafter, the chip was incubated with HRP-linked secondary antibody (Abcam, cat. no. ab7090), stained with 3,3' diaminobenzidine (DAB) and counterstained with hematoxylin. HPN immunostaining scoring of all cases was performed under the guidance of a qualified pathologist using H-score [22].

Results

scRNA-seq profiling and cell typing in PCa

Two patients were diagnosed as PCa by the overexpression of AMACR and loss expression of TP63 in prostate tissue (Supplementary Fig. 1). Cancerous samples from two patients were collected, dissected, and digested into single cells with which scRNA-seq was performed with 10 × Genomics Chromium (Fig. 1a). A total of 7904 qualified cells were applied for further analysis. The cellular composition was explored with unbiased clustering across all cells by PCA and visualized by UMAP (Fig. 1b). Identities of 15 main clusters were defined according to cluster-specific marker genes identified by DEG analysis: T cell, endothelial cell, type 1 luminal cell, type 2 luminal cell, erythroblast, myofibroblast, smooth muscle cell, macrophage, type 3 luminal cell, mast cell, basal cell, fibroblast, mesenchymal cell, B cell and neuroendocrine cell (Fig. 1c).



In our data, 3 types of luminal cells were identified according to their high expression level of luminal markers including Keratin 8 (*KRT8*) and Keratin 18 (*KRT18*) in PCa samples [23]. To characterize these luminal clusters, we examined the expression pattern of respective marker genes individually by VlnPlot and FeaturePlot. Results showed that Kallikrein Related Peptidase 4 (*KLK4*) and Solute Carrier Family 45 Member 3 (*SLC45A3*) were specifically expressed in type 1 luminal cell; thus, they could be used as specific markers for the detection of these cells (Fig. 2a, b). Ceruloplasmin (*CP*) was uniquely expressed in type 2 luminal cells, while Cellular Retinoic Acid Binding Protein 2 (*CRABP2*) was highly expressed in mesenchymal cell, as well as type 2 and type 3 luminal cells, suggesting a more suitable marker of *CP* for type 2 luminal cells (Fig. 2a, b). Type 3 luminal cells exhibited much higher expression levels of Beta-1,4-Galactosyltransferase 1 (*B4GALT1*) and Interferon Alpha Inducible Protein 6 (*IFI6*) compared with other clusters, indicating that *B4GALT1* and *IFI6* may identify these cells (Fig. 2a, b). To investigate the cytological localizations of each type of luminal cells in PCa tissue, we performed immunostaining using anti-*SLC45A3*, anti-*CP*, anti-*B4GALT1* antibodies and counterstained the tissue sections with DAPI (Fig. 2c). *SLC45A3* was expressed in most luminal cells of the prostate tissue (Fig. 2c). In contrast, *CP* was detected in a small part of luminal cells with a low expression level of *SLC45A3* (Fig. 2c). *B4GALT1* was located at similar positions to *CP* positive areas but not entirely overlapped, suggesting different roles for each type of luminal cells in PCa development (Fig. 2c).

Identification of malignant luminal cells in PCa

To evaluate the malignancy of identified luminal clusters, analysis of CNV levels in each cell type were performed according to average expression patterns across intervals of the genome [24]. Most non-luminal cells showed extremely low CNV levels except for erythroblast and fibroblast cells, which presented moderate CNV levels (Supplementary Fig. 2A). In luminal cells, type 1 luminal cell exhibited the highest CNV level, whereas type 2 and type 3 luminal cells showed lower CNV levels, indicating that type 1 luminal cells might be malignant cells in PCa (Supplementary Fig. 2B).

We further examined the gene expression patterns in these cells using edgeR package. Type 1 luminal cell showed much higher expression level of PCa markers including *FOLH1*, *KLK3* and Neuropeptide Y (*NPY*) (Supplementary Fig. 3A, D) [25, 26]. In contrast, type 2 luminal cells exhibited higher expression of PCa suppression-related genes, Latexin (*LXN*) and Secretory Leukocyte Peptidase Inhibitor (*SLPI*), the loss expression of which were usually detected in PCa and associated with negative prognosis of PCa patients (Supplementary

Fig. 3B, E) [27, 28]. However, Anterior Gradient 2 (*AGR2*), a urine marker for PCa [29], was also highly expressed in type 2 luminal cells, indicating that type 2 luminal cells were not entirely normal cells. Type 3 luminal cells exhibited high expression of C-C Motif Chemokine Ligand 3 (*CCL3*), C-X-C Motif Chemokine Ligand 1 (*CXCL1*) and Dickkopf WNT Signaling Pathway Inhibitor 1 (*DKKI*), which have been reported to have an increased expression level in PCa and relate to the occurrence of bone metastases and invasiveness in PCa (Supplementary Fig. 3C, F) [30–32]. In conclusion, type 1 and 3 luminal clusters were consisted of malignant cells with different roles in PCa initiation and progression, whereas type 2 luminal cells were normal prostate epithelial cells undergoing carcinogenesis.

To further determine the potential roles of each luminal cluster in PCa initiation and progression, functional enrichment was performed. Genes upregulated in type 1 luminal cells were mainly enriched for metabolic processes involved in cancer progression, such as fatty acid metabolic process, steroid biosynthetic process and peptide hormone secretion, indicating that type 1 luminal cells were indeed malignant cells (Fig. 3a, d). Most genes with high expression levels in type 2 luminal cells were enriched in positive regulation of peptidase activity and proteolysis, and negative regulation of cell-cell adhesion, suggesting that type 2 luminal cells were related to cancer growth and migration (Fig. 3b, e). Interestingly, other highly expressed genes in type 2 luminal cells were enriched in normal prostate developmental processes including epidermis development, epidermal cell differentiation and response to steroid hormones (Fig. 3b, e). Taken together, type 2 luminal cells were identified as cells of relatively low malignancy. To further verify this perspective, we performed a GO analysis of each luminal DEG acquired from pairwise comparison and received similar results (Supplementary Fig. 4). Upregulated genes expressed in type 3 luminal cells were mainly involved in tumor migration related processes like vasculogenesis, positive regulation of epithelial to mesenchymal transition and endothelium development, indicating that type 3 luminal cells were also malignant cells (Fig. 3c, f).

The pseudotime trajectory of PCa epithelial cells during tumor progression

PCa is usually characterized by luminal cell expansion and loss of basal cells [33, 34], therefore it is believed that luminal cells are the origins of neoplastic cells in PCa. However, Goldstein et al. demonstrated that origins might not originate from luminal cells because basal cells from primary benign prostate tissue can induce prostatic tumorigenesis in immune-deficient mice [35]. To investigate the origins of neoplastic cells in the tumorigenesis of PCa, pseudotime trajectory analysis

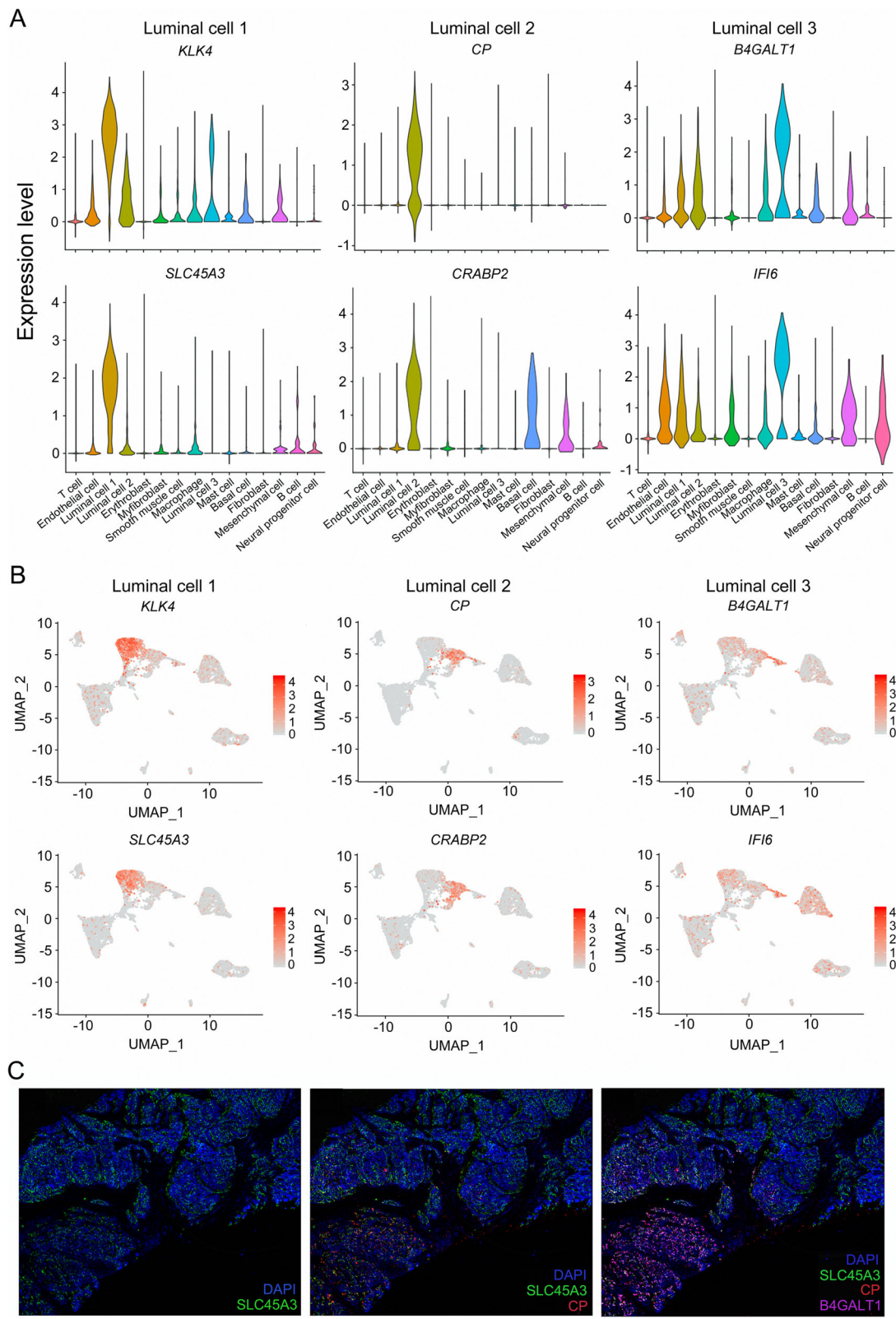


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 The expression levels of specific marker genes of diverse luminal clusters examined by scRNA-seq analysis and immunostaining in PCa tissue. **a** Violin plots displaying the expression levels of each luminal representative markers in each cluster. **b** Expression levels of representative markers for each luminal cluster plotted onto the UMAP. Color key from gray to red indicates relative expression levels from low to high. **c** Immunostaining showing the cytological localization of each luminal cluster cells in representative PCa tissues. Blue fluorescence represents nucleus stained with DAPI; green fluorescence represents type 1 luminal cells stained with anti-SLC45A3; red fluorescence represents type 2 luminal cells stained with anti-CP; purple fluorescence represents type 3 luminal cells stained with anti-B4GALT1

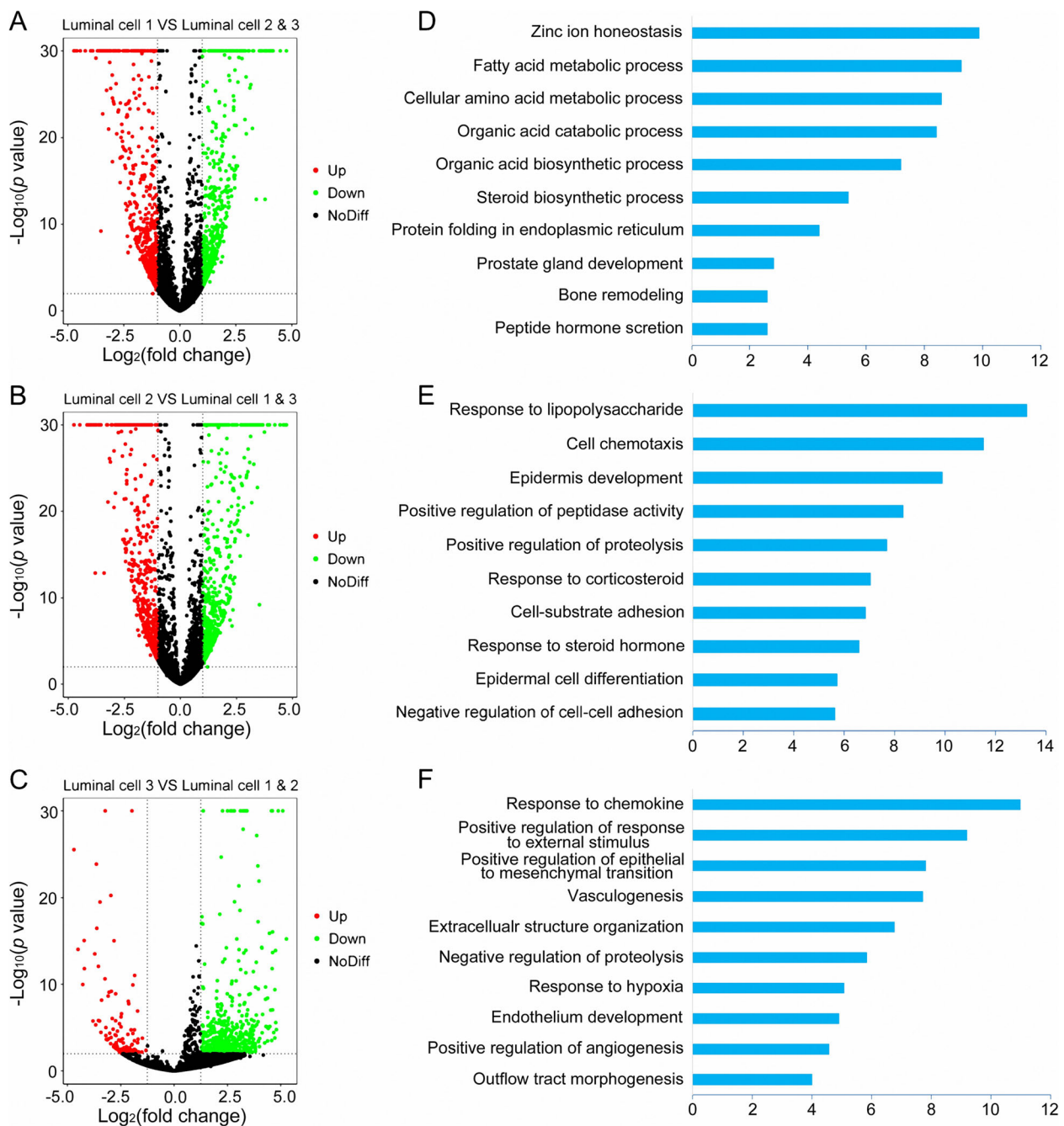


Fig. 3 DEGs of each luminal cluster and the enriched biological processes. **a, b, c** The DEGs in each luminal cluster were identified using edgeR package with the comparison to the other two luminal clusters. Scatter plots showing DEGs profiles of type 1, 2, 3 luminal clusters in PCa, respectively. Red spots indicate upregulated genes; green spots indicate downregulated genes; black spots indicate no significant change in genes. **d, e, f** The enriched biological processes for DEGs in type 1, 2, 3 luminal clusters, respectively

was further performed using basal cells and 3 types of luminal cells. Basal cells and type 2 luminal cells were shown in the beginning of the trajectory (Fig. 4a, b). Type 3 luminal cells appeared at the end of the trajectory branch 1 and type 1 luminal cells were at both ends of the trajectory branch 2 (Fig. 4a, b). These findings indicated that both basal cells and type 2 luminal cells could be the tumor-initiating cells and may transition to 2 different types of malignant luminal cells during the progression of PCa.

We further analyzed the dynamic expression changes of genes along the trajectory to determine the genes critical for PCa progression. Genes with the most significant expression changes were identified: Acid Phosphatase, Prostate (*ACPP*), Kallikrein Related Peptidase 2 (*KLK2*), *KLK3*, Keratin 15 (*KRT15*), Matrix Metalloproteinase 7 (*MMP7*) and *NPY* (Fig. 4c). In addition, we performed clustering of the top 100 genes with pseudotemporal expression pattern and analyzed the functional enrichments of each cluster. Genes in cluster 1 were highly

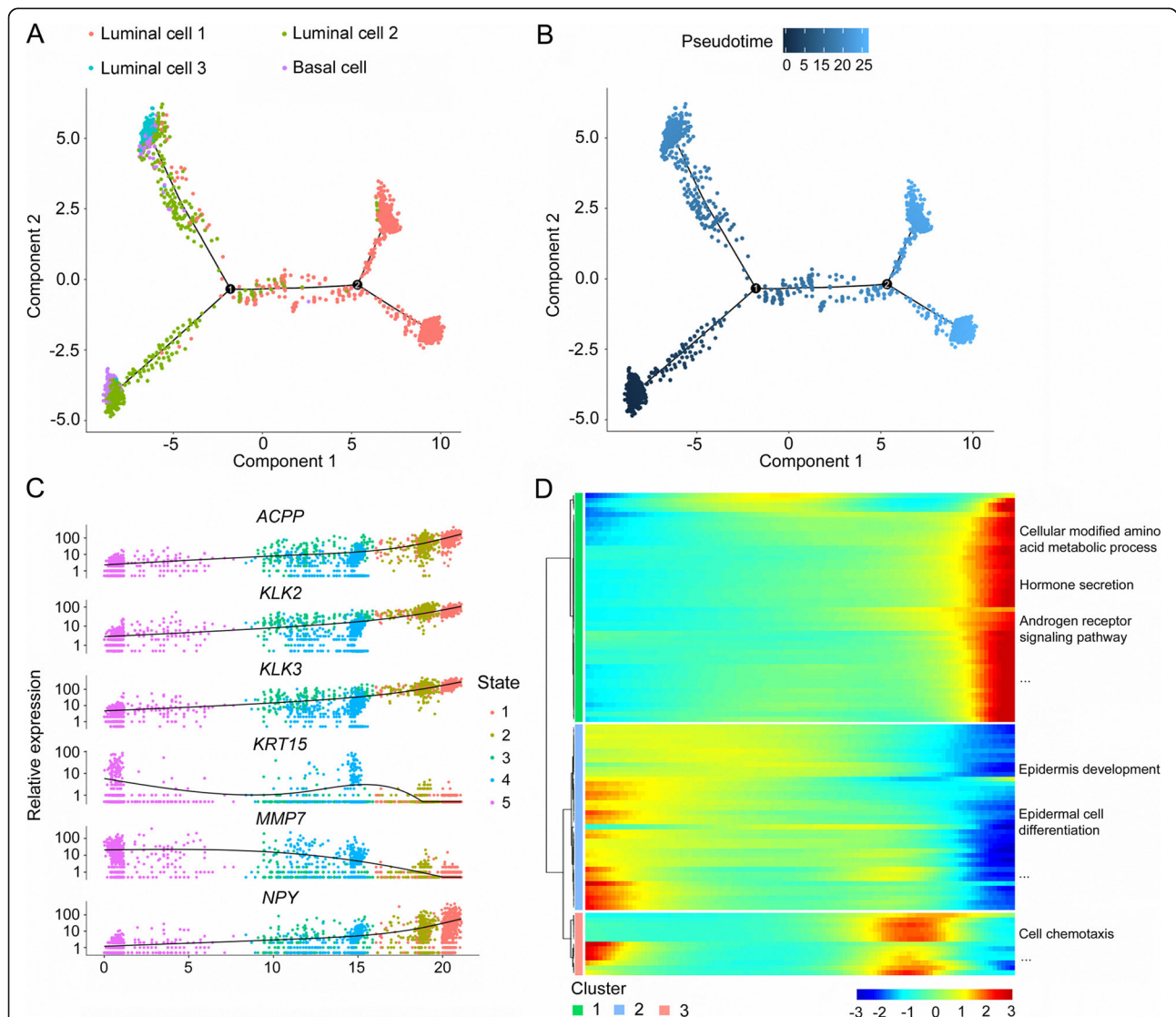


Fig. 4 Reconstructing the pseudotime trajectory of cancer cells using basal and luminal cells, and identifying genes varied during the trajectory. **a** Pseudotime trajectory of basal cells and 3 types of luminal cells was generated by Monocle2. Red spots represent type 1 luminal cells; green spots represent type 2 luminal cells; blue spots represent type 3 luminal cells; purple spots represent basal cells. **b** Pseudotime was colored in a gradient from dark to light blue. The start of pseudotime is indicated by dark blue, the end of pseudotime by light blue. **c** Basal and luminal cells were divided into 5 states by featured gene expression profiles. Top six DEGs with expression levels that changed the most over pseudotime trajectory were identified and shown as dot plots representing as expression level. **d** Top 100 DEGs with expression levels that changed the most over the pseudotime trajectory were divided into 3 clusters based on their expression trend, and the representative processes of each cluster are shown. Color key from blue to red indicates relative expression levels of top 100 DEGs from low to high

expressed at the end stage and they were mainly enriched in PCa initiation- and progression-related processes including cellular modified amino acid metabolic process, hormone secretion, and AR signaling pathways (Fig. 4d). Genes in cluster 2 showed high expression level at the beginning stage and were mainly involved in epidermis development and epidermal cell differentiation, suggesting a potential role for normal prostate development (Fig. 4d). Most genes in cluster 3 had increased expression at the intermediate stage of the pseudotime trajectory and enriched in cell chemotaxis (Fig. 4d).

Sub-clustering of malignant luminal cells in PCa

We first analyzed the expression levels of each luminal marker genes in the public PCa patients from TCGA to examine their clinical relevance to PCa initiation and progression. In comparison, most type 1 luminal marker genes had increased expression in PCa with low malignancy (Gleason score = 6 or 7) than high malignancy (Gleason score ≥ 8) or normal prostate, indicating a potential role in PCa initiation and early development (Supplementary Fig. 5). We therefore performed sub-clustering of type 1 luminal cells to identify the subpopulation responsible for PCa initiation. The type 1 luminal cells were divided into 5 subgroups by PCA (Fig. 5a, b). Functional enrichment for marker genes of each subgroup was analyzed and subgroup 1 was associated with biomolecule metabolic process essential for the preparation of tumor cells mitosis, such as lipid metabolic process and organic acid metabolic process (Fig. 5c). Compared with other subgroups, subgroup 2 was related to cell motility, tube morphogenesis and negative regulation of cell death, the processes critical for cancer migration and growth (Fig. 5c). Subgroup 3 was mainly involved in the G protein coupled receptor signaling pathway and muscle tissue development (Fig. 5c). Subgroup 4 correlated with PCa-related processes, including regulation of hormone level, hormone transport and peptide hormone transcription (Fig. 5c). In comparison, the unique function of subgroup 5 was mainly enriched in cell growth (Fig. 5c). VlnPlot and FeaturePlot showed that *AMACR*, Glycine-N-Acyltransferase Like 1 (*GLYATL1*) and *PCA3* were specifically expressed in subgroup 5 (Fig. 5d, e). Interestingly, *AMACR* is a peroxisomal enzyme highly expressed in cancerous prostate and involved in β -oxidation of branched-chain fatty acids and bile acid intermediates, and is critical for PCa initiation and progression [36, 37]. *PCA3* is a non-coding RNA with high expression levels in PCa cell lines, primary tissues and even urine from PCa patients [38]. These two markers of subgroup 5 have been used for PCa diagnosis and stratification for decades, implying that

subgroup 5 might be a distinct bunch of cells essential for PCa diagnosis and stratification in clinic.

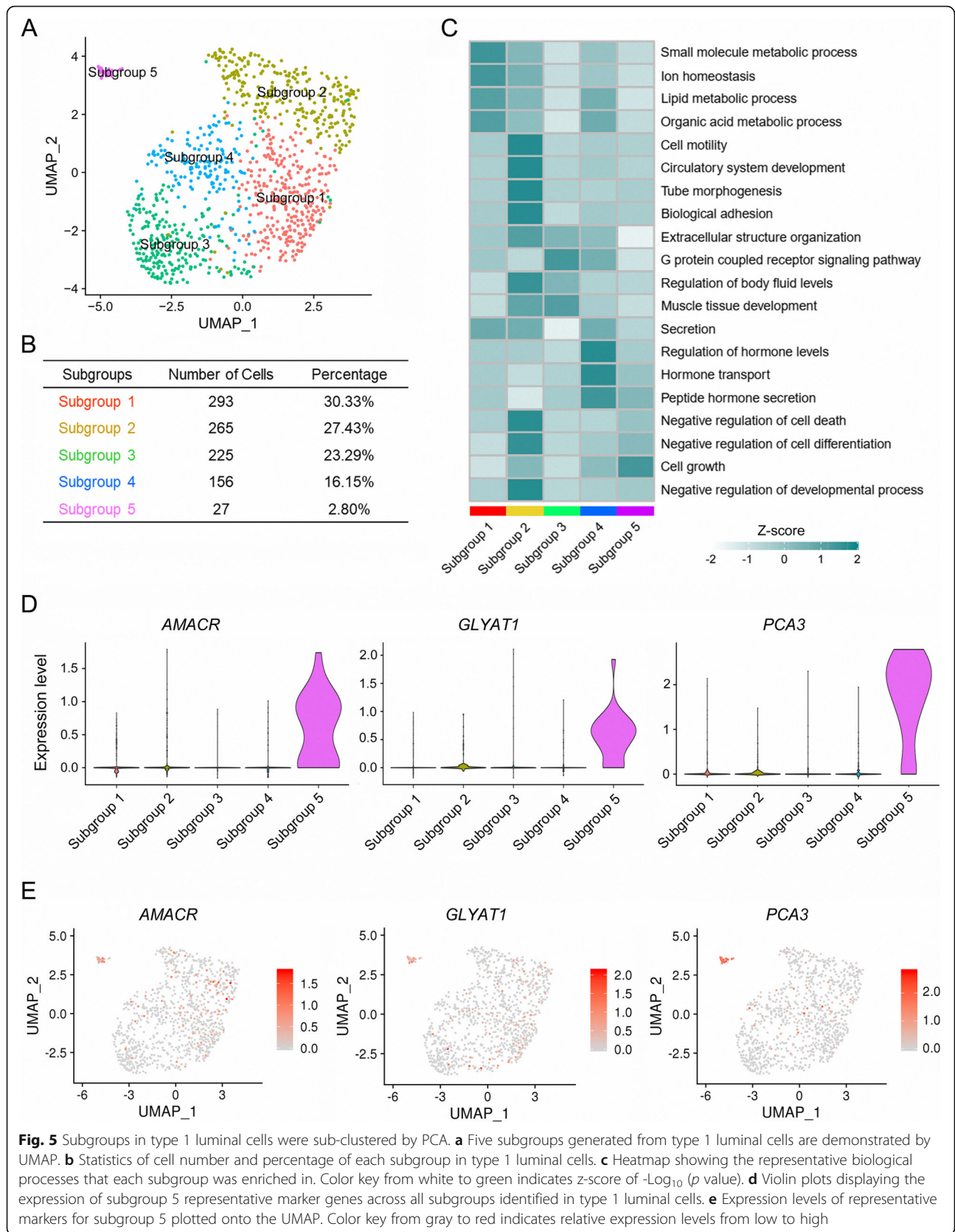
Subgroup 5 in type 1 luminal cell is critical for PCa diagnosis and stratification

Clinical correlations of type 1 luminal subgroups were analyzed with the expression patterns of their marker genes in PCa patients from TCGA. The expression levels of most subgroup 1–4 marker genes in normal prostate were similar to or even higher than those in PCa tissues (Supplementary Fig. 6). In contrast, marker genes of subgroup 5 were highly expressed in PCa tissues, especially in patients with a Gleason score at 6 or 7, suggesting a potential role in PCa initiation and early development (Fig. 6a). We further identified 6 marker genes of subgroup 5 with the highest specificity and sensitivity in distinguishing normal prostate from cancerous prostate by ROC analysis, including *AMACR*, *GLYATL1*, *HPN*, *PCA3*, Prostate Cancer Associated Transcript 18 (*PCAT18*) and Phospholipase A2 Group VII (*PLA2G7*) (Fig. 6b). ROC analysis based on the 6-gene set showed a strong PCa predictive ability with an AUC score of 0.937 (Supplementary Fig. 7A). In addition, the expression patterns of these genes not only distinguished PCa from normal prostate, but also distinguished early-stage PCa (Gleason score = 6) from highly malignant PCa (Gleason score ≥ 8) (Supplementary Fig. 8).

Numerous PCa patients exhibit therapy resistance within 1–2 years and relapse into advanced PCa, such as CRPC. Subgroup 5 may contribute to PCa recurrence. However, Kaplan-Meier analysis revealed that these genes showed no significant predictive abilities in biochemical recurrence, either individually or as a gene set (Fig. 6c, Supplementary Fig. 7B). In summary, malignant cells in subgroup 5 were critical for PCa diagnosis and stratification, but not relevant to biochemical recurrence.

HPN is a qualified biomarker for PCa diagnosis and stratification

Specifically, *HPN* as one of the subgroup 5 marker genes presented enormous potential to distinguish PCa tissue from normal prostate with an AUC score as high as 0.930, which is only down by 0.007 compared with the 6-gene set (Fig. 6b, Supplementary Fig. 7A). In addition, *HPN* showed different distinguishing ability in patients with various pathology grading as shown in Supplementary Fig. 8. To further validate the bioinformatic prediction of *HPN* as a qualified biomarker for PCa diagnosis and stratification, we performed immunostaining on PCa tissue array with 55 normal prostates and 95 cancer tissues with different pathology grading demonstrated by Gleason scores from 6 to 9. We found that *HPN* was highly expressed in cancer tissues compared with normal prostate, and the staining intensities seemed to vary



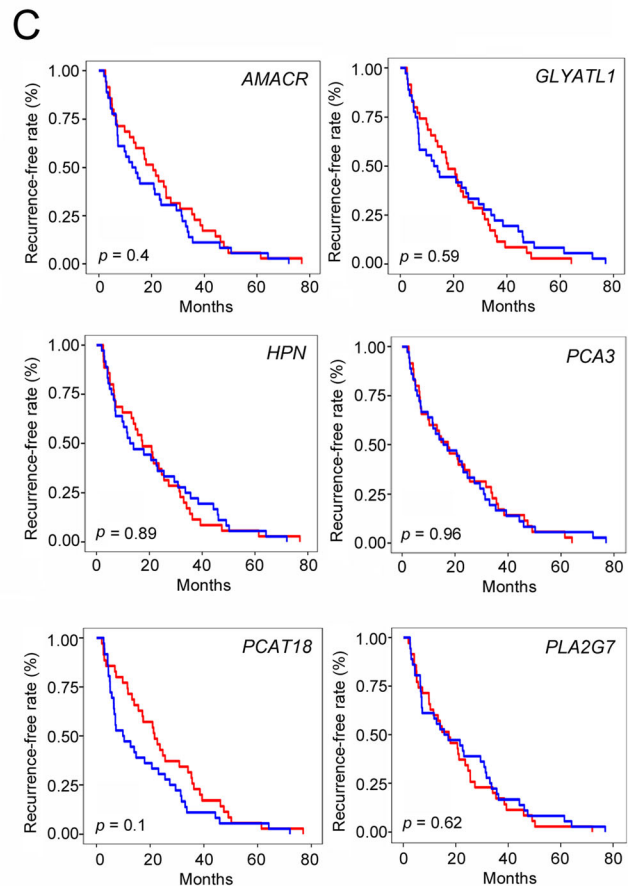
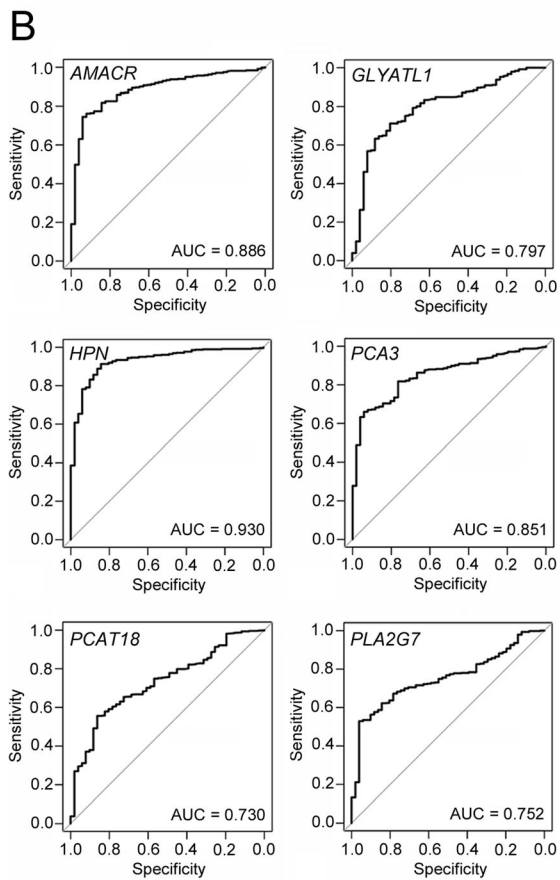
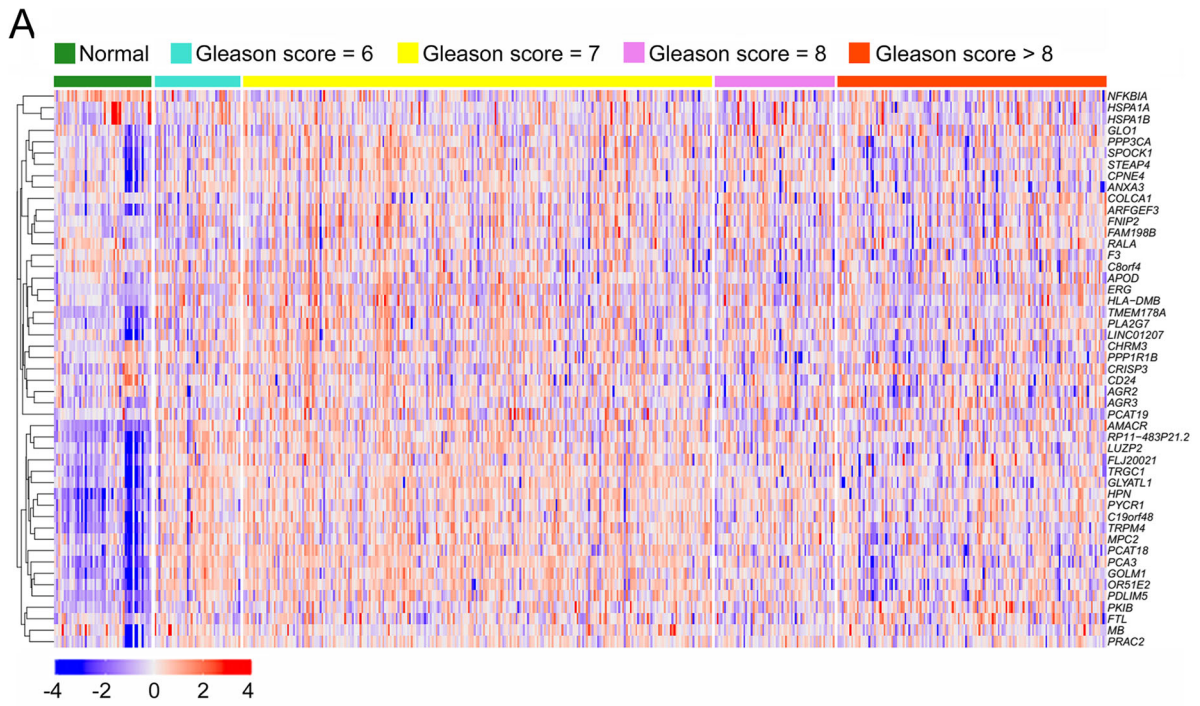


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Clinical relevance of subgroup 5 to PCa diagnosis and stratification. **a** Clustering heatmap demonstrating the correlation of PCa status to subgroup 5 marker gene expression using TCGA data. **b** ROC curves for top 6 marker genes of subgroup 5 in distinguishing normal prostate and cancerous prostate using TCGA data. **c** Kaplan-Meier plot predicting recurrence-free rate of PCa patients based on the expression changes of top 6 subgroup 5 marker genes

between the cancer tissues with different pathology grading (Fig. 7a). To examine the intensity differences, we further analyzed the immunostaining scoring with H-score, and found that the expression level of HPN was significantly enhanced in cancer tissues (Fig. 7b). In addition, the expression level of HPN in cancer tissues with high pathology grading (Gleason score = 7, 8, 9) was significantly higher than that with low pathology grading (Gleason score = 6) (Fig. 7c). Therefore, we believe that HPN might be a more specific and sensitive biomarker for clinical PCa diagnosis and stratification.

To explore the role of HPN in predicting PCa recurrence, we divided the PCa patients from TCGA into 2 groups as LOW and HIGH based on HPN expression level, thereafter analyzed the expression of therapy-resistant markers including ATP Binding Cassette Subfamily G Member 2 (*ABCG2*), Aldehyde Dehydrogenase 1 (*ALDH1*), Receptor Tyrosine Kinase (*KIT*), Activated Leukocyte Cell Adhesion Molecule (*ALCAM*), Cadherin 1 (*CDH1*), C-X-C Motif Chemokine Receptor 4 (*CXCR4*), Epithelial Cell Adhesion Molecule (*EPCAM*), Enhancer Of Zeste 2 (*EZH2*), Integrin Subunit Beta 1 (*ITGB1*) and Transglutaminase 2 (*TGM2*) in these two groups. We found no significant differences between PCa patients with low and high expression of HPN, suggesting that HPN was not suitable as a predictive marker for PCa recurrence (Fig. 7d).

Discussion

Numerous genomic and transcriptomic studies have been performed to obtain specific and sensitive markers for precise stratification and early detection of PCa. However, PCa is a highly heterogeneous tumor consisting of multiple types of cells. Except for PSA, few clinically relevant biomarkers have been identified and established using novel techniques such as next generation sequencing. In the context of these sequencing strategies, significant gene expression differences in specific cell population might be normalized or even hidden by “less important” genes expressed in abundance. Compared to conventional RNA-seq, scRNA-seq examines cancerous gene expression at the individual cell level and has been recently used to investigate heterogeneity in many types of tumors [14–17]. However, single-cell analysis of PCa has only been reported in liquid biopsies and cultured prostate epithelial cells [39, 40]. To our knowledge, studies using scRNA-seq to examine PCa heterogeneity with primary PCa tissue has yet been

reported. In this study, we performed scRNA-seq using PCa tissues obtained from two patients and identified 15 clusters, including 3 different types of luminal clusters. Type 1 luminal cells, identified as highly malignant cells according to CNV level and pseudotime trajectory analysis, might be closely related to PCa initiation and early development.

Multiple PCa initiation- and progression-related genes that have been used as potential markers for PCa diagnosis can also be detected in different subgroup of type 1 luminal cells. Subgroup 1 expressed several of these genes including *NPY* and *KLK3*. *NPY* is a secreted protein highly expressed in early PCa and usually associated with worse clinical outcome [41]. *KLK3* is the coding gene of PSA, which has been used in prostate health index (PHI) and 4 K score to evaluate PCa early development [42]. Subgroup 2 highly expressed E74-Like Factor 3 (*ELF3*), the expression of which is usually elevated in PCa for tumor growth [43]. In contrast, no PCa candidate marker genes have been identified in subgroup 3 based on our findings. Subgroup 4 marker genes contained Golgi Membrane Protein 1 (*GOLM1*) and *FOLH1*. Golgi protein *GOLM1*, consistently upregulated in PCa, has been used as a urine marker of PCa [44]. *FOLH1* encodes prostate specific membrane antigen (PSMA), a membrane bound glycoprotein overexpressed in PCa and serves as an important marker and target for drug delivery [45]. Subgroup 5 expressed marker genes *AMACR* and *PCA3*, which have been widely used for PCa diagnosis and stratification in decades due to their more specific expression in PCa cells compared to previously used biomarkers like *KLK3* and *FOLH1* [46, 47]. Therefore, we speculated that subgroup 5 was a distinct cell cluster critical for PCa diagnosis and stratification.

We further generated an ROC curve for more subgroup 5 marker genes and found that *HPN* exhibited the potential ability in distinguishing normal prostate from cancerous prostate in PCa patients with various pathology grading. Particularly, *HPN* showed an AUC score of 0.930, which was higher than that of most reported candidate marker genes for PCa diagnosis (Supplementary Fig. 9). To validate our findings, we performed immunostaining with anti-*HPN* antibody on PCa tissue array and verified its expression in normal prostate and cancerous prostate with various pathology grading. *HPN* is a membrane serine protease identified as one of the most overexpressed molecules in PCa [48]. Goel et al. stated that *HPN* should be regarded as a novel immunohistochemical marker for the histopathological diagnosis

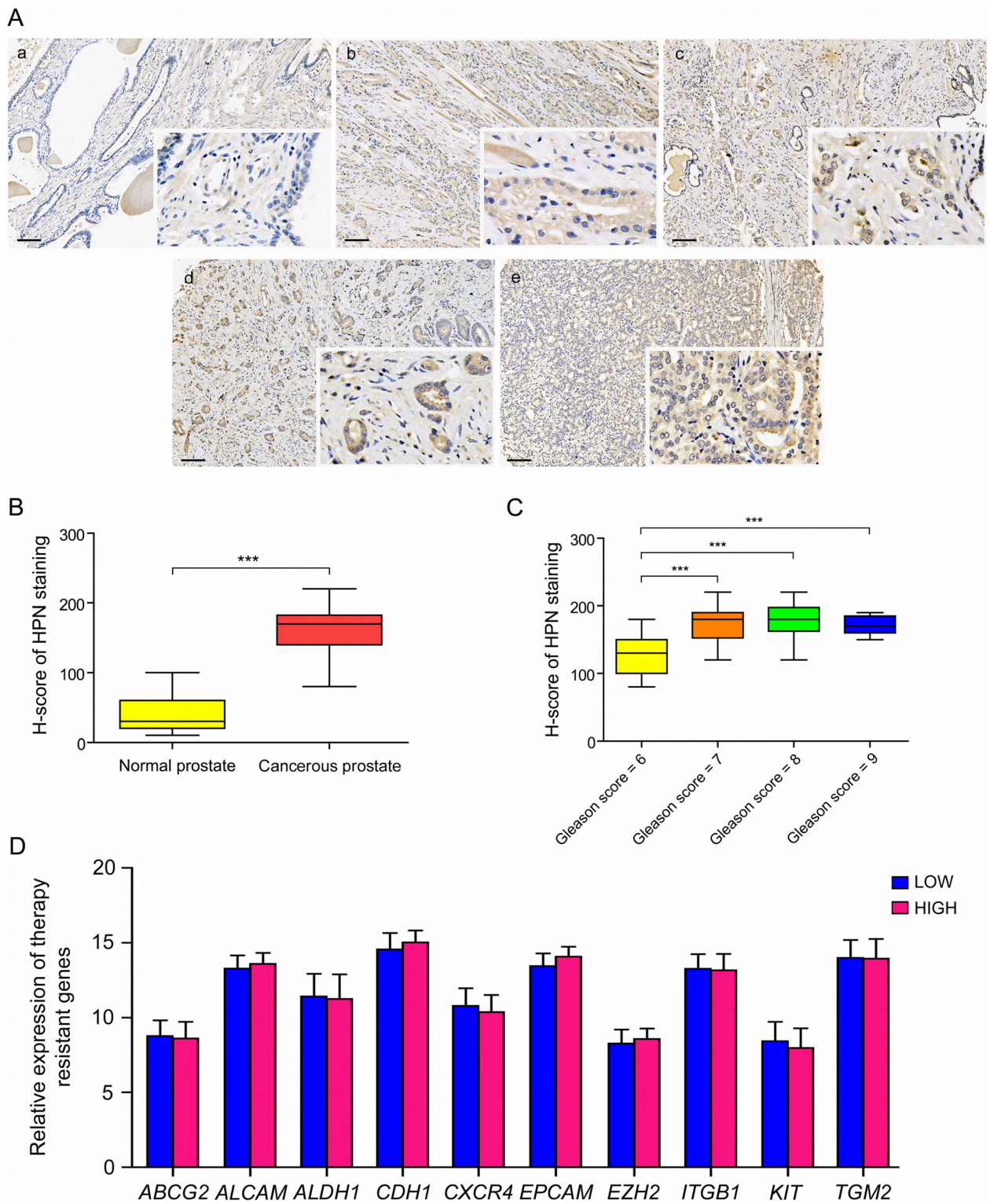


Fig. 7 Validation of HPN expression in PCa tissue array. **a** Immunostaining of HPN in normal prostate and cancerous prostate with different pathology grading. Positive signals with anti-HPN were stained in brown. Cell nucleus were stained with hematoxylin and presented blue in PCa tissue sections. **a** normal prostate, **b** cancerous prostate with a Gleason score of 6, **c** cancerous prostate with a Gleason score of 7, **d** cancerous prostate with a Gleason score of 8, **e** cancerous prostate with a Gleason score of 9. **b** H-score of HPN staining in normal prostate and cancerous prostate. **c** H-score of HPN staining in PCa tissues with different pathology grading. **d** Relative expression of therapy-resistant markers in PCa patients with low and high expression levels of HPN, **LOW** patients with low expression of HPN; **HIGH** patients with high expression of HPN

of PCa based on the immunostaining intensity among normal prostate, low-scoring and high-scoring PCa [49]. Furthermore, HPN as a cell surface marker could be a drug target in PCa treatment. Tang et al. have reported that targeted inhibition of HPN by small-molecule inhibitor HepIn-13 could attenuate PCa progression and metastasis [50]. Taken together, HPN may be a potential biomarker for clinical PCa diagnosis and stratification, even a potential target for PCa treatment.

Conclusions

In conclusion, this is the first report of PCa heterogeneity examined by scRNA-seq of primary PCa tissue to our knowledge. We found that PCa tissue consisted of 3 different types of luminal cells with distinct roles in PCa initiation and progression. A distinct subgroup of luminal cells critical for PCa diagnosis and stratification was identified along with its marker gene HPN. Our findings are potentially valuable in not only advancing the current understanding of PCa initiation and progression, but also the translational use of markers for PCa diagnosis and stratification.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12943-020-01264-9>.

Additional file 1 Supplementary Figure 1. PCa pathology grading diagnosis by histology observation and the expression of AMACR and TP63. **A, D** HE staining of PCa tissues. **B, E** Immunostaining of AMACR on PCa tissues. **C, F** Immunostaining of TP63 on PCa tissues. Bar = 50 μ m. **Supplementary Figure 2** CNV analysis of different type of cells in PCa tissues. A All clusters in PCa tissues; B Three types of luminal clusters in PCa tissues. **Supplementary Figure 3 Highly expressed genes distributed in PCa.** **A-C** Violin plots displaying the expression of highly expressed genes in each luminal cluster across the cell types identified in PCa (from luminal 1 to luminal 3, respectively); **D-F** Expression levels of highly expressed genes in each luminal cluster plotted onto the UMAP (from luminal 1 to luminal 3, respectively), color key from gray to red indicates relative expression levels from low to high. **Supplementary Figure 4** GO enrichment of differentially expressed genes (DEGs) of each luminal cluster analyzed by pairwise comparison. **A, B** The enriched GO terms for DEGs between type 1 and type 2 luminal cells; **C, D** The enriched GO terms for DEGs between type 1 and type 3 luminal cells; **E, F** The enriched GO terms for DEGs between type 2 and type 3 luminal cells. **Supplementary Figure 5** Clustering heatmap demonstrating the correlation between PCa status and the marker gene expression of each luminal cluster using TCGA data. **Supplementary Figure 6** Clustering heatmap demonstrating the correlation between PCa status and the marker gene expression of subgroup 1–4 using TCGA data. **Supplementary Figure 7** Clinical correlations of 6-gene set from subgroup 5 marker genes were analyzed with their expression patterns in PCa patients from TCGA. **A** ROC analysis for 6-gene set from subgroup 5 marker genes in distinguishing normal prostate from cancerous prostate; **B** Kaplan–Meier analysis predicting recurrence-free rate of PCa patients based on the expression changes of 6-gene set from subgroup 5 marker genes. **Supplementary Figure 8** Heatmap showing different distinguishing abilities of subgroup 5 marker genes in patients with various pathology gradings. **Supplementary Figure 9** ROC analysis of reported candidate marker genes for PCa diagnosis.

Abbreviations

ABCG2: ATP Binding Cassette Subfamily G Member 2; ACPP: Acid Phosphatase, Prostate; AGR2: Anterior Gradient 2; ALCAM: Activated

Leukocyte Cell Adhesion Molecule; ALDH1: Aldehyde Dehydrogenase 1; AMACR: Alpha-Methylacyl-CoA Racemase; AR: Androgen Receptor; ATM: Ataxia Telangiectasia Mutated; AUC: Area Under The Curve; B4GALT1: Beta-1,4-Galactosyltransferase 1; BPH: Benign Prostatic Hyperplasia; BRCA1/2: Breast Cancer 1/2; CCL3: C-C Motif Chemokine Ligand 3; CCND1: Cyclin D1; CDH1: Cadherin 1; CDK12: Cyclin Dependent Kinase 12; CDKN1B: Cyclin Dependent Kinase Inhibitor 1B; CNV: Copy Number Variation; CP: Ceruloplasmin; CRABP2: Cellular Retinoic Acid Binding Protein 2; CRPC: Castration Resistant Prostate Cancer; CT: Computed Tomography; CXCL1: C-X-C Motif Chemokine Ligand 1; CXCR4: C-X-C Motif Chemokine Receptor 4; DKK1: Dickkopf WNT Signaling Pathway Inhibitor 1; DAB: 3,3' Diaminobenzidine; DBS: DNA Double Strand Breaks; DEGs: Differentially Expressed Genes; DRE: Digital Rectal Examination; ELF3: E74-Like Factor 3; EPCAM: Epithelial Cell Adhesion Molecule; EZH2: Enhancer Of Zeste 2; FOLH1: Folate Hydrolase 1; FGFR1: Fibroblast Growth Factor Receptor 1; GLYA TL1: Glycine-N-Acyltransferase Like 1; GO: Gene Ontology; GOLM1: Golgi Membrane Protein 1; HPN: Hepsin; IFI6: Interferon Alpha Inducible Protein 6; ITGB1: Integrin Subunit Beta 1; KIT: Receptor Tyrosine Kinase; KLK2: Kallikrein Related Peptidase 2; KLK3: Kallikrein Related Peptidase 3; KLK4: Kallikrein Related Peptidase 4; KRT8: Keratin 8; KRT18: Keratin 18; LYN: Latexin; MMP7: Matrix Metalloproteinase 7; MRI: Magnetic Resonance Imaging; MYC: Myelocytomatosis; NPY: Neuropeptide Y; PBS: Phosphate Buffer Saline; PCa: Prostate Cancer; PCA: Principal Component Analysis; PCA3: Prostate Cancer Associated 3; PCAT18: Prostate Cancer Associated Transcript 18; PET: Positron Emission Tomography; PLA2G7: Phospholipase A2 Group VII; PSA: Prostate Specific Antigen; PSMA: Prostate Specific Membrane Antigen; PTEN: Phosphatase And Tensin Homolog; RB1: Retinoblastoma Transcriptional Corepressor 1; RNA-seq: RNA sequencing; ROC: Receiver Operating Characteristic; scRNA-seq: Single-cell mRNA Sequencing; SLC45A3: Solute Carrier Family 45 Member 3; SLP1: Secretory Leukocyte Peptidase Inhibitor; TCGA: The Cancer Genome Atlas; TGM2: Transglutaminase 2; TMPRSS2-ERG: Transmembrane Serine Protease 2-Erythroblast Transformation Specific Related Gene; TP53: Tumor Protein P53; TP63: Tumor Protein P63; TRUS: Transrectal Ultrasound; UMAP: Uniform Manifold Approximation and Projection

Acknowledgments

We would like to thank Dr. Yangmin Chen, PharmD, for editorial assistance.

Authors' contributions

Chang Zou, Xiaoshi Ma, Jinan Guo and Kaisheng Liu conceived and designed the experiments; Jinan Guo and Kefeng Xiao collected the patients' samples; Xiaoshi Ma, Pan Zhao, and Fengyan Hu performed the experiments; Xiaoshi Ma, Lipeng Chen, Dale Liu, Shaowei Dong, Jinqian Xia, Qiaoyun Long, and Yongjian Yue analyzed the data; Xiaoshi Ma wrote the paper; Kaisheng Liu, Zhongqiang Xiao, Xinghua Pan, Zunfu Ke, Zhe-Sheng Chen and Chang Zou revised the paper; Zhiqiang Cheng, Zunfu Ke, Zhe-Sheng Chen and Chang Zou provide funding and supervision. All authors read and approved the final manuscript.

Funding

This study was funded by the Science and Technology Foundation of Shenzhen (JCYJ20180305164128430, GJHZ20180928171602104), the Shenzhen Economic and Information Committee "Innovation Chain and Industry Chain" integration special support plan project (20180225112449943), the Basic and Applied Basic Research of Guangdong Province (2019B1515120033), the Guangdong Provincial Natural Science Foundation (2018A030313743), and the Shenzhen Public Service Platform on Tumor Precision Medicine and Molecular Diagnosis.

Availability of data and materials

All data generated during this study are included in this published article and its supplementary files. Raw sequencing data and processed gene expression data were deposited at the Gene Expression Omnibus (GEO) under accession number GSE157703.

Ethics approval and consent to participate

This study was approved by the Shenzhen People's Hospital Research Ethics Committee (NO. LL-KY-2020128).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, Guangdong, China. ²Shenzhen Public Service Platform on Tumor Precision Medicine and Molecular Diagnosis, Shenzhen, Guangdong, China. ³Key Laboratory of Medical Electrophysiology of Education Ministry, School of Pharmacy, Southwest Medical University, Luzhou, Sichuan, China. ⁴Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, China. ⁵Guangdong Provincial Key Laboratory of Single Cell Technology and Application, Guangzhou, Guangdong, China. ⁶Guangdong-Hongkong-Macao Great Bay Area Center for Brain Science and Brain-Inspired Intelligence, Guangzhou, Guangdong, China. ⁷Department of Pathology, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, Guangdong, China. ⁸Department of Pharmaceutical Sciences, College of Pharmacy and Health Sciences, St. John's University, Queens, New York, USA.

Received: 10 July 2020 Accepted: 25 September 2020

Published online: 08 October 2020

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, A. J. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394–424.
- Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P, Davis M, Peters TJ, Turner EL, Martin RM, et al. 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N Engl J Med.* 2016; 375:1415–24.
- Komura K, Sweeney CJ, Inamoto T, Ibuki N, Azuma H, Kantoff PW. Current treatment strategies for advanced prostate cancer. *Int J Urol.* 2018;25:220–31.
- Candefjord S, Ramsar K, Lindahl OA. Technologies for localization and diagnosis of prostate cancer. *J Med Eng Technol.* 2009;33:585–603.
- Pathak H, Gupta M, Kumar D, Lakhe V, Rahangdale P, Fulari N. Prostate Cancer: PSA and other diagnostic modalities for the early diagnosis. *Indian Pract.* 2014;67:436–40.
- Kijviki K. Digital rectal examination, serum prostatic specific antigen or transrectal ultrasonography: the best tool to guide the treatment of men with benign prostatic hyperplasia. *Curr Opin Urol.* 2009;19:44–8.
- De Visser P, Oosterlinck W, De Meerleer G, Villeirs G. Clinical and imaging tools in the early diagnosis of prostate cancer, a review. *Jbr-btr.* 2010;93:62–70.
- Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. Integrative clinical genomics of advanced prostate cancer. *Cell.* 2015;161:1215–28.
- Rubin MA, Girelli G, Demichelis F. Genomic correlates to the newly proposed grading prognostic groups for prostate Cancer. *Eur Urol.* 2016;69: 557–60.
- Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell.* 2010;18:11–22.
- Abeshouse A, Ahn J, Akbani R, Ally A, Zmuda E. The molecular taxonomy of primary prostate cancer. *Cell.* 2015;163:1011–25.
- Mateo J, Carreira S, Sandhu S, Miranda S, Mossop H, Perez-Lopez R, Nava Rodrigues D, Robinson D, Omlin A, Tunariu N, et al. DNA-repair defects and Olaparib in metastatic prostate Cancer. *N Engl J Med.* 2015;373:1697–708.
- Tolkach Y, Kristiansen G. The heterogeneity of prostate Cancer: A practical approach. *Pathobiology.* 2018;85:108–16.
- Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, Zhou A, Eyob H, Balakrishnan S, Wang CY, et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature.* 2015;526:131–5.
- Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, Fisher JM, Rodman C, Mount C, Filbin MG, et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendrogloma. *Nature.* 2016; 539:309–13.
- Peng J, Sun BF, Chen CY, Zhou JY, Chen YS. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* 2019;29:725–38.
- Young MD, Mitchell TJ. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science.* 2018;361:594–9.
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoekius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. *Cell.* 2019;177:1888–902 e1821.
- Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* 2019;47:D721–d728.
- Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science.* 2014;344:1396–401.
- Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32:2847–9.
- Pirker R, Pereira JR, von Pawel J, Krzakowski M, Ramlau R, Park K, de Marinis F, Eberhardt WE, Paz-Ares L, Störkel S, et al. EGFR expression as a predictor of survival for first-line chemotherapy plus cetuximab in patients with advanced non-small-cell lung cancer: analysis of data from the phase 3 FLEX study. *Lancet Oncol.* 2012;13:33–42.
- Barclay WW, Axanova LS, Chen W, Romero L, Maund SL, Soker S, Lees CJ, Cramer SD. Characterization of adult prostatic progenitor/stem cells exhibiting self-renewal and multilineage differentiation. *Stem Cells.* 2008;26:600–10.
- Puram SV, Tirosh I, Parkh AS, Patel AP, Yizhak K, Gillespie S, Rodman C, Luo CL, Mroz EA, Emerick KS, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell.* 2017;171: 1611–24 e1624.
- Haberhorn U, Eder M, Kopka K, Babich JW, Eisenhut M. New strategies in prostate Cancer: prostate-specific membrane antigen (PSMA) ligands for diagnosis and therapy. *Clin Cancer Res.* 2016;22:9–15.
- Ueda K, Tatsuguchi A, Saichi N, Toyama A, Tamura K, Furihata M, Takata R, Akamatsu S, Igarashi M, Nakayama M, et al. Plasma low-molecular-weight proteome profiling identified neuropeptide-Y as a prostate cancer biomarker polypeptide. *J Proteome Res.* 2013;12:4497–506.
- Thompson M, Lapointe J, Choi YL, Ong DE, Higgins JP, Brooks JD, Pollack JR. Identification of candidate prostate cancer genes through comparative expression-profiling of seminal vesicle. *Prostate.* 2008;68:1248–56.
- Seed RI, Taurozzi AJ, Wilcock DJ, Nappo G, HHH E. The putative tumour suppressor protein Latexin is secreted by prostate luminal cells and is downregulated in malignancy. *Sci Rep.* 2019;9:5120.
- Bu H, Bormann S, Schäfer G, Horninger W, Massoner P, Neeb A, Lakshmanan VK, Maddalo D, Nestl A, Sultmann H, et al. The anterior gradient 2 (AGR2) gene is overexpressed in prostate cancer and may be useful as a urine sediment marker for prostate cancer detection. *Prostate.* 2011;71:575–87.
- Browne AJ, Gobel A, Thiele S, Hofbauer LC, Rauner M, Rachner TD. p38 MAPK regulates the Wnt inhibitor Dickkopf-1 in osteotropic prostate cancer cells. *Cell Death Dis.* 2016;7:e2119.
- Benelli R, Stigliani S, Minghelli S, Carlone S, Ferrari N. Impact of CXCL1 overexpression on growth and invasion of prostate cancer cell. *Prostate.* 2013;73:941–51.
- Qian W-j, Shi T, Liu AY. Prostate cancer-associated secreted proteins. *United States; 2020.*
- Okada H, Tsubura A, Okamura A, Senzaki H, Naka Y, Komatz Y, Morii S. Keratin profiles in normal/hyperplastic prostates and prostate carcinoma. *Virchows Arch A Pathol Anat Histopathol.* 1992;421:157–61.
- Parsons JK, Gage WR, Nelson WG, De Marzo AM. p63 protein expression is rare in prostate adenocarcinoma: implications for cancer diagnosis and carcinogenesis. *Urology.* 2001;58:619–24.
- Goldstein AS, Huang J, Guo C, Garraway IP, Witte ON. Identification of a cell of origin for human prostate cancer. *Science.* 2010;329:568–71.
- Wu X, Daniels G, Lee P, Monaco ME. Lipid metabolism in prostate cancer. *Am J Clin Exp Urol.* 2014;2:111–20.
- Zheng SL, Chang BL, Faith DA, Johnson JR, Isaacs SD, Hawkins GA, Turner A, Wiley KE, Bleecker ER, Walsh PC, et al. Sequence variants of alpha-methylacyl-CoA racemase are associated with prostate cancer risk. *Cancer Res.* 2002;62:6485–8.
- Ploussard G, de la Taille A. The role of prostate cancer antigen 3 (PCA3) in prostate cancer detection. *Expert Rev Anticancer Ther.* 2018; 18:1013–20.

39. Lambros MB, Seed G, Sumanasuriya S, Gil V, Crespo M, Fontes M, Chandler R, Mehra N. Single-Cell Analyses of Prostate Cancer Liquid Biopsies Acquired by Apheresis. *Clin Cancer Res.* 2018;24:5635–44.
40. McCray T, Moline D, Baumann B, Vander Griend DJ, Nonn L. Single-cell RNA-Seq analysis identifies a putative epithelial stem cell population in human primary prostate cells in monolayer and organoid culture conditions. *Am J Clin Exp Urol.* 2019;7:123–38.
41. Rasiah K. The identification of novel biomarkers in the development and progression of early prostate Cancer. University of New South Wales; 2006.
42. Kretschmer A, Tilki D. Biomarkers in prostate cancer – current clinical utility and future perspectives. *Crit Rev Oncol Hematol.* 2017;120:180–93.
43. Wang H, Ru Y, Sanchez-Carbayo M, Wang X, Kieft JS, Theodorescu D. Translation initiation factor eIF3b expression in human cancer and its role in tumor growth and lung colonization. *Clin Cancer Res.* 2013;19:2850–60.
44. Varambally S, Laxman B, Mehra R, Cao Q, Dhanasekaran S, Tomlins S, Granger J, Vellaichamy A, Sreekumar A, Yu J, et al. Golgi Protein GOLM1 Is a Tissue and Urine Biomarker of Prostate Cancer 1 2. *Neoplasia* (New York, NY). 2008;10:1285–94.
45. Ghosh A, Heston WDW. Tumor target prostate specific membrane antigen (PSMA) and its regulation in prostate cancer. *J Cell Biochem.* 2004;91:528–39.
46. Ferro M, Bruzzese D, Perdonà S, Marino A, Mazzarella C, Perruolo G, d'esposito V, Cosimato V, Buonerba C, Lorenzo G, et al. Prostate health index (phi) and prostate Cancer antigen 3 (PCA3) significantly improve prostate Cancer detection at initial biopsy in a Total PSA range of 2-10 ng/ml. *PLoS One.* 2013;8:e67687.
47. Alinezhad S, Väänänen RM, Ochoa NT, Vertosick EA, Bjartell A, Boström PJ, Taimen P, Pettersson K. Global expression of AMACR transcripts predicts risk for prostate cancer - a systematic comparison of AMACR protein and mRNA expression in cancerous and noncancerous prostate. *BMC Urol.* 2016;16:10.
48. Wu Q, Parry G. Hepsin and prostate cancer. *Front Biosci.* 2007;12:5052–9.
49. Goel MM, Agrawal D, Natu SM, Goel A. Hepsin immunohistochemical expression in prostate cancer in relation to Gleason's grade and serum prostate specific antigen. *Indian J Pathol Microbiol.* 2011;54:476–81.
50. Tang X, Mahajan SS, Nguyen LT, Béliveau F, Leduc R, Simon JA, Vasioukhin V. Targeted inhibition of cell-surface serine protease Hepsin blocks prostate cancer bone metastasis. *Oncotarget.* 2014;5:1352–62.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

