

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

### Identification of Prostate Cancer Subtypes Based on Immune Signature Scores in Bulk and Single Cell Transcriptomes

Canping Chen

China Pharmaceutical University Xiaosheng Wang ( ≤ xiaosheng.wang@cpu.edu.cn) China Pharmaceutical University https://orcid.org/0000-0002-7199-7093

**Research Article** 

**Keywords:** prostate cancer, tumor immune microenvironment, subtyping, unsupervised clustering, transcriptomes

Posted Date: February 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1264750/v1

**License:** (a) This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

#### Abstract

Prostate cancer (PC) is heterogeneous in the tumor immune microenvironment (TIME). Subtyping of PC based on the TIME could provide new insights into intratumor heterogeneity and its correlates of clinical features. Based on the enrichment scores of 28 immune cell types in the TIME, we performed unsupervised clustering to identify immune-specific subtypes of PC. The clustering analysis was performed in ten different bulk tumor transcriptomic datasets and in a single cell RNA-Seq (scRNA-seq) dataset, respectively. We identified two PC subtypes: PC immunity high (PC-ImH) and PC immunity low (PC-ImL), consistently in these datasets. Compared to PC-ImL, PC-ImH displayed stronger immune signatures, worse clinical outcomes, higher epithelial-mesenchymal transition (EMT) signature, tumor stemness, intratumor heterogeneity (ITH) and genomic instability, and lower incidence of TMPRSS2-ERG fusion. Tumor mutation burden (TMB) showed no significant difference between PC-ImH and PC-ImL, while copy number alteration (CNA) was more significant in PC-ImL than in PC-ImH. PC-ImH could be further divided into two subgroups, which had significantly different immune infiltration levels and clinical features. In conclusion, "hot" PCs have stronger anti-tumor immune response while worse clinical outcomes versus "cold" PCs. CNA instead of TMB plays a crucial role in the regulation of TIME in PC. TMPRSS2-ERG fusion correlates with decreased anti-tumor immune response while better disease-free survival in PC. The identification of immune-specific subtypes has potential clinical implications for PC immunotherapy.

#### Background

Prostate cancer (PC) is the second most common cancer in men worldwide, with more than 1,275,000 new diagnoses and 350,000 deaths annually. Abundant evidence has shown that PC is highly heterogeneous in genomic and phenotypic characteristics [1]. By analyzing 333 primary PCs, The Cancer Genome Atlas (TCGA) revealed that more than 70% of PCs belonged to one of seven subtypes in terms of specific gene fusions (*ERG*, *ETV1/4*, and *FLI1*) or mutations (*SPOP*, *FOXA1*, and *IDH1*) [2]. Salami et al. demonstrated transcriptomic heterogeneity in PC by analyzing 156 multifocal, low-grade, and high-grade PC samples [3]. Wilkinson et al. showed that specific molecular features, e.g., loss of chromosome 10q and *TP53* alterations, correlated with poor response to neoadjuvant androgen deprivation therapy [4].

PC treatments include two categories: local treatments and systemic treatments. Local treatments refer to surgery and radiation therapy, while systemic treatments include hormonal therapy, targeted therapy, chemotherapy, and immunotherapy. In particular, immunotherapies, such as immune checkpoint inhibitors (ICIs), have recently achieved success in treating various malignancies, including refractory and metastatic tumors [5]. Currently, several ICIs, included sipuleucel-T [6], abiraterone acetate [7], enzalutamide [8], cabazitaxel [9], radium-223 [10], and apalutamide [11], have been approved by the Food and Drug Administration (FDA) to treat androgen depletion therapy-resistant PCs.

Nevertheless, current immunotherapeutic strategies are beneficial to only a subset of cancer patients. Thus, discovery of effective markers for immunotherapeutic responses is urgently in need. Several such markers have been identified and are being used in clinical practice, including PD-L1 expression [12], DNA mismatch repair deficiency [13], and tumor mutation burden (TMB) [14]. Besides, the tumor immune microenvironment (TIME) plays a crucial role in immunotherapeutic responses [15]. Overall, the "hot" tumors with dense T cell infiltration are more likely to respond to immunotherapy, compared to the "cold" tumors with sparse T cell infiltration [16]. Thus, identification of "hot" and "cold" tumors may aid stratification of cancer patients responsive to immunotherapy. In a previous study [17], we have developed an unsupervised machine learning method to identify "hot" and "cold" tumor subtypes based on immune signature scores.

In this study, we performed unsupervised clustering based on the enrichment scores of 28 immune cell types in the TIME to identify immune-specific subtypes of PC. To demonstrate the reproducibility of this method, we performed the clustering analysis in ten different transcriptomic datasets for PC bulk tumors. Consistently, we identified two PC subtypes which exhibited significantly different molecular and clinical features. Finally, we validated our method and results in a single cell RNA-Seq (scRNA-seq) dataset. Our identification of immune-specific subtypes may provide new insights into associations between the TIME and molecular and clinical features in PC as well as potential clinical implications for the immunotherapy of this disease.

#### Methods

# Datasets

We downloaded ten gene expression profiling datasets for PC, including TCGA-PRAD, DKFZ2018, GSE21034, GSE46602, GSE54460, GSE70770, GSE107299, GSE116918, GSE141551, and GSE157547. The TCGA-PRAD data were downloaded from the genomic data commons (GDC) data portal (https://portal.gdc.cancer.gov/). The DKFZ2018 data were obtained from cBioPortal (http://www.cbioportal.org/). The other data were downloaded from the NCBI gene expression omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). From GDC, we also downloaded profiles of somatic mutations ("maf" file), somatic copy number alterations (SCNAs) ("SNP6" files), and protein expression (normalized) in TCGA-PRAD. In addition, we downloaded a single-cell RNA sequencing (scRNA-seq) dataset (GSE141445) for PC [18] from the NCBI GEO. A summary of these datasets is presented in Supplementary Table S1.

# Calculation of enrichment scores of immune signatures, pathways, and phenotypic features

We calculated the enrichment score of an immune signature, pathway, or phenotypic feature in a tumor sample by the single-sample gene-set enrichment analysis (ssGSEA) [19] of the expression profiles of its gene set. We collected the gene sets representing immune signatures, pathways, or phenotypic features from KEGG [20] or related publications, as shown in Supplementary Table S2.

# Identification of PC subtypes

We identified PC subtypes based on the enrichment scores (ssGSEA scores) of 28 immune cell types by hierarchical clustering. The 28 immune cell types included CD56-bright natural killer (NK) cells, effector memory CD4 T cells, eosinophil, CD56-dim NK cells, type 17 T helper cells, activated B cells, monocytes, memory B cells, activated CD4 T cells, type 2 T helper cells, plasmacytoid dendritic cells, neutrophils, macrophages, effector memory CD8 T cells, myeloid-derived suppressor cell (MDSC), immature B cells, T follicular helper cells, NK cells, immature dendritic cells, mast cells, type 1 T helper cells, activated dendritic cells, activated cells, activated cells, activated cells, type 2 T cells, mast cells, type 1 T helper cells, activated cells, type 2 T follicular helper cells, activated cells, gamma delta T cells, central memory CD8 T cells, regulatory T cells, activated CD8 T cells, and natural killer T cells [21].

# Calculation of tumor immune scores

We calculated immune scores of tumors by ESTIMATE [22] with the input of the expression profiles of immune genes. The immune score reflects the tumor immune infiltration level.

# Survival analysis

We compared disease-free survival (DFS) rates between PC subtypes using the Kaplan-Meier (K-M) model [23]. K-M curves were used to show the survival time differences, and log-rank tests were utilized to evaluate the significance of survival time differences.

# Pathway analysis

We first identified differentially expressed genes (DEGs) between PC subtypes using Student's *t* test with a threshold of adjusted *P*-value < 0.05 and fold change of mean expression levels > 1.5. By input of the upregulated DEGs in a PC subtype into GSEA [24], we obtained the KEGG pathways highly enriched in the subtype with a threshold of adjusted *P*-value < 0.05.

# Calculation of TMB and SCNAs

We determined a tumor sample's TMB as the total count of its somatic mutations. We calculated G-scores in tumors by GISTIC2 [25] with the input of "SNP6" files.

# **Class prediction**

We first normalized attribute values (ssGSEA scores of immune cells) by Z score and transformed all attribute values into the range between -3 and 3 by setting an attribute value as 3 if it was greater than 3 and setting an attribute value as -3 if it was less than -3. We used the Random Forest (RF) algorithm [26] to predict PC subtypes. In the RF model, the number of trees was set to 500, and the features were the 28 immune cells. We reported prediction accuracies, sensitivities, and specificities. The class prediction was performed using the R package "randomForest".

# Statistical analysis

We used the Mann–Whitney U test to compare two classes of non-normally distributed data and Student's *t* test for normally distributed data. We used the Spearman method to correlate immune scores and other variables. The Fisher's exact test was utilized to evaluate the association between two

categorical variables. To adjust for *P*-values in multiple tests, we used the Benjamini-Hochberg method [27] to calculated the false discovery rate (FDR). We performed all statistical analyses in the R programming environment (version 4.0.3).

#### Results

# Identification of PC subtypes based on immune cell enrichment scores

Based on the enrichment scores of 28 immune cell types [28], we identified PC subtypes by hierarchical clustering. We performed the clustering analysis in ten PC datasets (TCGA-PRAD, DKFZ2018, GSE21034, GSE46602, GSE54460, GSE70770, GSE107299, GSE116918, GSE141551, and GSE157547), respectively. The clustering analysis identified two PC subtypes: PC immunity high (PC-ImH) and PC immunity low (PC-ImL), consistently in the ten datasets (Fig. 1A). To verify the significantly higher immunity in PC-ImH versus PC-ImL, we compared the enrichment scores of both immunostimulatory signatures (CD8+ T cells, immune cytolytic activity, and interferon (IFN) response) and immunosuppressive signatures (CD4+ regulatory T cells, M2 macrophages, T cell exhaustion, and myeloid-derived suppressor cells (MDSCs)) between PC-ImH and PC-ImL. Interestingly, all these immune signatures displayed significantly higher enrichment scores in PC-ImH than in PC-ImL (one-tailed Mann–Whitney *U* test, *P* < 0.01) in the ten datasets (Fig. 1B). Also, the ratios of immunostimulatory to immunosuppressive signatures (CD8+/CD4+ regulatory T cells), which were the base-2 log-transformed values of the geometric mean expression levels of all marker genes of CD8+ T cells divided by those of CD4+ regulatory T cells, were significantly higher in PC-ImL (two-tailed Student's *t* test, *P* < 0.05) in most of these datasets (Fig. 1B). These results confirmed that PC-ImH had significantly stronger immune signatures than PC-ImL.

# Clinical and phenotypic features of the PC subtypes

We compared disease-free survival (DFS) time between PC-ImH and PC-ImL in three datasets (TCGA-PRAD, GSE116918, and GSE46602) with survival data available. In these datasets, PC-ImL were likely to show better DFS than PC-ImH (log-rank test, *P* = 0.014, 0.078, and 0.117, respectively) (Fig. 2A). To support that the survival difference between both subtypes has an association with their different immune infiltration levels, we compared DFS between high-immune-score (> mean) and low-immunescore (< mean) PC patients. The immune score represents the immune infiltration level in the tumor, which was calculated by the ESTIMATE algorithm [22]. We observed that low-immune-score patients tended to have a better DFS than high-immune-score patients (Fig. 2B). It supports that the survival difference between both PC subtypes correlates with their different tumor immune microenvironment.

In TCGA-PRAD, we compared several clinical features between both subtypes, including blood level of prostate-specific antigen (PSA), Gleason score, pathological stage, lymph node staging, and metastasis. The elevated level of prostate-specific antigen (PSA) is often associated with PC. We found that PSA levels were significantly higher in PC-ImH than in PC-ImL (one-tailed Mann–Whitney *U*test, *P* = 0.009)

(Fig. 2C). Tumor grade indicates how abnormal the tumor cells and the tumor tissue look under a microscope compared to normal cells and how quickly a tumor is likely to grow and spread. The Gleason score, ranging from 6 to 10, is the classical grading system for PC. We found that PC-ImH harbored a significantly higher proportion of high-grade (Gleason score of 9 to 10) tumors than PC-ImL (Fisher's exact test, P = 0.013; odds ratio (OR) = 1.68) (Fig. 2D). The pathological stage assesses how pervasive the cancer cells are within and around the prostate. We found that PC-ImH harbored a higher proportion of late-stage (T3-4) tumors than PC-ImL (P = 0.063; OR = 1.43) (Fig. 2D). Lymph node staging indicates whether cancer cells are present in nearby lymph nodes. We observed that PC-ImH harbored a significantly higher proportion of patients whose cancer cells present in nearby lymph nodes than PC-ImL (P = 0.030; OR = 1.78) (Fig. 2D). Finally, we found that PC-ImH involved a significantly higher proportion of patients. For example, in DKFZ2018, high-grade tumors had a significantly higher proportion in PC-ImL (P = 0.009; OR = 5.65).

We further compared two tumor phenotypes between PC-ImH and PC-ImL, including epithelialmesenchymal transition (EMT) signature and intratumor heterogeneity (ITH). These phenotypes are associated with tumor progression, metastasis, and drug resistance [29–31]. We found that EMT signature scores were significantly higher in PC-ImH than in PC-ImL in the ten datasets (P < 0.01) (Fig. 2E); ITH scores, which were evaluated by the DEPTH algorithm [30], were also significantly higher in PC-ImH than in PC-ImL in two datasets (P < 0.05) (Fig. 2F).

To investigate whether the worse DFS in PC-ImH versus PC-ImL was influenced by other confounding variables, we performed multivariate survival analysis with the multivariate Cox proportional hazards model. In the model, the response variable was DFS time, and the predictor variables included age, PSA level, Gleason score, and subtype. We found that the subtype PC-ImH was still a risk factor (P = 0.049; hazard ratio (HR) = 1.603 and 95% confidence interval (CI): [1.002, 2.560]) (Fig. 2G). As expected, both Gleason score (P < 0.001; HR = 2.261 and 95% CI: [1.773, 2.884]) and PSA level (P < 0.001; HR = 1.065 and 95% CI: [1.036, 1.094]) were also significant risk factors for DFS in PC. However, age showed no significant correlation with DFS time in PC (P = 0.506).

Taken together, these results suggest a worse prognosis in PC-ImH versus PC-ImL.

# Genomic features of the PC subtypes

Tumor genomic features, such as TMB and CNA, have been associated with anti-tumor immune response in cancer [32]. However, TMB showed no significant difference between PC-ImH and PC-ImL (two-tailed Mann–Whitney *U* test, P = 0.98) (Fig. 3A). In contrast, CNA were significantly different between both subtypes. For example, homologous recombination deficiency (HRD) may contribute to aneuploidy (i.e., CNA) in cancer [33]. The HRD scores [33] were significantly higher in PC-ImH than in PC-ImL (one-tailed Mann–Whitney *U* test, P = 0.04) (Fig. 3B). The G-score calculated by GISTIC2 [34] represents the amplitude of the CNA and the frequency of its occurrence across a group of samples. We found that the G-scores of copy number amplifications and deletions were significantly higher in PC-ImL than in PC-ImH (Fig. 3C). Altogether, these results suggest: (1) PC-ImH has a higher level of genomic instability than PC-ImL; and (2) CNA rather than TMB has a significant impact on the TIME in PC.

We further compared activities (enrichment scores) of nine DNA damage repair (DDR) pathways between PC-ImH and PC-ImL. The nine pathways included mismatch repair, base excision repair, nucleotide excision repair, the Fanconi anemia (FA) pathway, homology-dependent recombination, non-homologous DNA end joining, direct damage reversal/repair, translesion DNA synthesis, and damage sensor. Notably, the enrichment scores of these pathways tended to be higher in PC-ImL and PC-ImH (Fig. 3D). It could explain why PC-ImL is more genomically stable than PC-ImH in terms of DNA repair deficiencies being a primary factor responsible for genomic instability in cancer [35].

# Pathways highly enriched in the PC subtypes

We identified KEGG pathways highly enriched in PC-ImH and PC-ImL by GSEA [24]. The pathways highly enriched in PC-ImH and PC-ImL were obtained by gene set enrichment analysis of significantly upregulated genes in PC-ImH and PC-ImL, respectively. Using a threshold of adjusted P-value < 0.05, we identified 67 pathways upregulated in PC-ImH and zero in PC-ImL. The pathways highly enriched in PC-ImH were mainly involved in to immune, stromal and oncogenic signatures (Fig. 4A). The immune-related pathways included chemokine signaling, cytokine-cytokine receptor interactions, Toll-like receptor signaling, natural killer cell-mediated cytotoxicity, leukocyte transendothelial migration, antigen processing and presentation, complement and coagulation cascades, NOD-like receptor signaling, Jak-STAT signaling, Fc epsilon RI signaling, Fc gamma R-mediated phagocytosis, cytosolic DNA-sensing, T cell receptor, B cell receptor, and TGF-B signaling. The stromal signature-related pathways included cell adhesion molecules, regulation of actin cytoskeleton, focal adhesion, ECM-receptor interaction, adherens junction, and gap junction. The oncogenic pathways MAPK signaling, apoptosis, VEGF signaling, p53, and Wnt signaling. In addition, several pathways involved in neural regulation were also highly enriched in PC-ImH, including axon guidance, neuroactive ligand receptor interaction, and neurotrophin signaling. The upregulation of numerous immune-rated pathways in PC-ImH confirmed the stronger immune signatures in this subtype versus PC-ImL. Furthermore, we observed that the enrichment scores of the oncogenic pathways enriched in PC-ImH tended to have positive correlations with immune scores in the ten datasets (Spearman correlation, P < 0.05) (Fig. 4B). It indicates that the upregulation of these pathways may activate the inflammatory TME in PC.

# Mutation profiles in the PC subtypes

We found eight genes showing significantly lower mutation frequencies in PC-ImH than in PC-ImL in TCGA-PRAD (Fisher's exact test, *P* < 0.05; OR < 0.14) (Fig. 5A). The eight genes included *KIAA0907, PIPSL, DUS3L, SH3TC2, SRGAP1, TRAV9-2, AGAP10,* and *CDH12*. Notably, the mutations of these genes correlated with lower immune scores in PC (*P* < 0.05) (Fig. 5B). In contrast, 25 genes displayed significantly higher mutation frequencies in PC-ImH than in PC-ImL in TCGA-PRAD (Fisher's exact test, *P* < 0.05; OR > 3.46) (Fig. 5A). These genes included *FOXP2, MALAT1, PCDHB7, HLA-A, LYST, MYH13,* 

# *PCDHA6, DID01, SCN10A, STAB2, HECTD4, MAP1A, MY018B, TBC1D2, MN1, MUC2, PIWIL1, ZNF423, ZNF733P, ANKRD26, BCL11A, HSPD1P6, KIAA1462, LRRC66*, and *XKR4*. As expected, the mutations of these genes were likely to correlated with higher immune scores in PC (*P* < 0.05) (Supplementary Fig. S1). **Protein expression profiles of the PC subtypes**

We found 17 proteins differentially expressed between PC-ImH and PC-ImL in TCGA-PRAD (two-tailed Student's *t* test, FDR < 0.05) (Fig. 6A). Among these proteins, six displayed significant higher expression levels in PC-ImH than in PC-ImL, including Stathmin, PREX1, Cyclin\_D1, HSP70, Lck, and Syk. In fact, many of these proteins have been associated with PC. For example, Stathmin, a member of microtubule-destabilizing protein family, plays a critical role in the regulation of mitosis and is overexpressed in various cancers [36]. This protein has been suggested as a target for the treatment of PC [36]. PREX1, a guanine nucleotide exchange factor for the RHO family of small GTP-binding proteins, is overexpressed in various cancers and its overexpression promotes PC metastasis [37]. Cyclin\_D1 is a key regulator of cell cycle and its overexpression is associated with PC metastasis [38]. Hsp70, a stress-inducible protein, has been indicated as a potential biomarker and therapeutic target in PC [39, 40]. SYK, a member of the Syk family of tyrosine kinases, has been implicated as a candidate therapeutic target for the therapy of advanced PC [41]. Overall, previous studies have indicated that overexpression of these proteins is associated with PC advancement or metastasis, consistent with their upregulation in the subtype with worse prognosis. Notably, most of these proteins displayed positive expression correlations with immune scores (Spearman correlation, *P* < 0.05) (Fig. 6B), consistently with their upregulation in PC-ImH.

The 11 proteins having higher expression levels in PC-ImL than in PC-ImH included FASN, Claudin-7, ACC1, E-Cadherin, ACC\_pS79,  $\beta$ -Catenin, VEGFR2, p70S6K, mTOR, SCD1, and Bak. In contrast with the proteins upregulated in PC-ImH, downregulation of many of these proteins is associated with PC progression, such as E-Cadherin [42], mTOR [43], and Bak [44]. As expected, most of these proteins showed negative expression correlations with immune scores (Spearman correlation, *P* < 0.05) (Fig. 6B), consistently with their upregulation in PC-ImL.

# **Prediction of the PC subtypes**

To demonstrate the predictability of the immune signature scores-based classification method, we performed the class prediction of the PC subtypes based on the enrichment scores of the 28 immune cell types by the Random Forest (RF) algorithm [26]. We used a dataset as the training set and the other nine datasets as test sets by turns. When TCGA-PRAD was the training set, its 10-fold cross-validation (CV) accuracy, sensitivity, and specificity was 91.8%, 92.6%, and 92.2%, respectively. In seven of the nine test sets, the prediction accuracies were higher than 80% (higher than 90% in two test sets); in all the nine test sets, the sensitivities were higher than 80% (higher than 90% in six test sets); in five test sets, the specificities were higher than 80% (higher than 90% in three test sets) (Fig. 7). We obtained similar results when the other datasets were the training set (Fig. 7). These results indicate the predictability of the immune signature scores-based classification method.

# Validation of the subtyping method in scRNA-seq data

Using the immune signature scores-based clustering method, we analyzed a scRNA-seq dataset (GSE141445) [18]. By the t-distributed stochastic neighbor embedding (tSNE) algorithm [45], we clustered 36424 single cells (Fig. 8A). The malignant and non-malignant cells were clearly separated. Moreover, the malignant cells were clearly separated based on different patients. For each PC patient, we hierarchically clustered its tumor cells based on their enrichment scores of four immune-rated pathways, including antigen processing and presentation, PD-L1 expression pathway in cancer, apoptosis, and JAK-STAT signaling. We used the four immune signatures instead of the 28 immune cell types for clustering bulk tumors because these immune signatures are expressed in tumor cells themselves. Consistently, we classified all tumor cells in a patient into two subgroups, also termed PC-ImH and PC-ImL (Fig. 8B). Interestingly, we found that PC-ImH tumor cells were obviously closer to T cells than PC-ImL tumor cells (Fig. 8C).

Based on the single cell clustering results, we defined a PC patient as PC-ImH if most (> 50%) of its tumor cells belonged to the immunity-high group (PC-ImH), otherwise as PC-ImL. According to this rule, seven and six PC patients were classified into PC-ImH and PC-ImL, respectively (Fig. 8D). Likewise, we compared clinical, phenotypic, and molecular features between both subtypes in this dataset. Consistent with previous results, PSA levels were significantly higher in PC-ImH than in PC-ImL (P = 0.009) (Fig. 8E). The stemness scores, and cell cycle activity (ssGSEA scores of the cell cycle pathway) were also significantly higher in PC-ImH than in PC-ImH that PC-ImH that PC-ImH that a worse prognosis than PC-ImL.

# Identification of two subgroups in PC-ImH

In TCGA-PRAD and GSE116918, there were significantly different survival prognosis between PC-ImH and PC-ImL (Fig. 2A). We further analyzed both datasets by the graph learning-based dimensionality reduction analysis [46]. We found that PC-ImH could be further divided into two subgroups, termed PC-ImH-A and PC-ImH-B (Fig. 9A). Compared to PC-ImH-B, PC-ImH-A showed stronger immune signatures (Fig. 9B) but worse DFS (Fig. 9C); PC-ImH-A harbored a higher proportion of late-stage (T3-4) tumors than PC-ImH-B (*P* < 0.001; OR = 2.48) (Fig. 9D). Moreover, PC-ImH-A had higher stemness scores, ITH levels and *MKI67* expression levels than PC-ImH-B (Fig. 9E). Overall, these results are consistent with the previous results of the comparisons between PC-ImH and PC-ImL, supporting the negative association between immune activities and clinical outcomes in PC.

# Association between TMPRSS2-ERG fusion and immune signatures in PC

TMPRSS2-ERG fusion has been indicated as a biomarker for PC [47]. Interestingly, in two datasets (TCGA-PRAD and DKFZ2018) with TMPRSS2-ERG fusion data available, TMPRSS2-ERG fusion was consistently more frequent in PC-ImL than in PC-ImH (Fisher's exact test, P < 0.05; OR > 1.5) (Fig. 10A). It indicates a negative association between TMPRSS2-ERG fusion and anti-tumor immune response. Indeed, in both datasets, the tumors with TMPRSS2-ERG fusion displayed significantly lower immune signature scores than those without TMPRSS2-ERG fusion (Fig. 10B). Moreover, the tumors with TMPRSS2-ERG fusion showed better DFS than those without TMPRSS2-ERG fusion (Fig. 10C). It suggests that TMPRSS2-ERG fusion could be a positive prognostic factor for PC patients, while it is a biomarker for PC tumorigenesis.

#### Discussion

We identified two immune-specific subtypes of PC based on the enrichment scores of 28 immune cell types. Compared to PC-ImL, PC-ImH displayed higher levels of immune infiltration and stromal contents, worse prognosis and tumor progression, higher scores of EMT signature, tumor stemness, and ITH, and higher level of genomic instability. Our data suggest that "hot" PCs have stronger anti-tumor immune response while worse clinical outcomes compared to "cold" PCs. This finding is consistent with that from a previous study [48]. The main reason could be that the strong anti-tumor immune response is in fact the tumor progression-promoting inflammation in PC [49]. The negative association between anti-tumor immune response and clinical outcomes have also been indicated in some cancer types, such as gliomas [50]. However, in many other cancer types, such as head and neck squamous cell cancer [51], gastric cancer [52], and triple-negative breast cancer [53], elevated anti-tumor immune response is associated with better clinical outcomes. Thus, the association between anti-tumor immune response and clinical outcomes.

Interestingly, although high TMB is a predictive biomarker for the active response to immunotherapy for its increased production of neoantigens to incite anti-tumor immune response, TMB showed no significant difference between PC-ImH and PC-ImL. However, CNA, an immunosuppressive marker [32], is more significant in PC-ImL than in PC-ImH. It suggests that CNA instead of TMB plays a crucial role in the regulation of TIME in PC.

Interestingly, PD-L1, an immunosuppressive signature, also showed higher mRNA expression levels in PC-ImH than in PC-ImL in eight of the ten datasets (two-tailed Student's *t* test, P < 0.05). The possible reason could be that PD-L1 is also expressed on immune cells, which are more abundant in PC-ImH. Indeed, by analyzing the scRNA-seq dataset (GSE141445) [18], we found that *PD-L1* was more highly expressed in immune cells than in tumor cells (P < 0.001). Because both PD-L1 expression [54] and high level of immune infiltration [16] are indicators of an active response to ICIs, PC-ImH would respond better to ICIs than PC-ImL. This hypothesis needs to be approved by clinical data, whereas they are not currently publicly available.

In conclusion, the identification of immune-specific subtypes has potential clinical implications for the prognosis and immunotherapy of PC.

#### Abbreviations

PC prostate cancer TIME tumor immune microenvironment scRNA-seq single cell RNA-Seq PC-ImH PC immunity high PC-ImL PC immunity low EMT epithelial-mesenchymal transition ITH intratumor heterogeneity TMB Tumor mutation burden CNA copy number alteration **TCGA** The Cancer Genome Atlas **ICIs** immune checkpoint inhibitors FDA Food and Drug Administration GDC genomic data commons GEO gene expression omnibus **SCNAs** somatic copy number alterations ssGSEA single-sample gene-set enrichment analysis NK natural killer **MDSC** myeloid-derived suppressor cell DFS disease-free survival K-M Kaplan-Meier

#### DEGs

differentially expressed genes RF Random Forest **FDR** false discovery rate IFN interferon **PSA** prostate-specific antigen OR odds ratio HR hazard ratio CI confidence interval HRD homologous recombination deficiency CV cross-validation CAFs cancer-associated fibroblasts **tSNE** t-distributed stochastic neighbor embedding.

#### Declarations

#### Conflict of Interest Statement(s)

All authors have read the journal's authorship agreement and reviewed and approved the final manuscript. All authors have read the journal's policy on conflicts of interest and have none to declare.

#### **Funding Statement**

This work was supported by the China Pharmaceutical University (grant number 3150120001 to XW). The sponsors had no role in study design, the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

#### CRediT authorship contribution statement

**Canping Chen:** Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing - review & editing. **Xiaosheng Wang:** Conceptualization, Methodology, Resources, Investigation, Writing -

original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

#### References

- 1. Haffner MC, et al. Genomic and phenotypic heterogeneity in prostate cancer. Nat Rev Urol. 2021;18(2):79–92.
- 2. Cancer Genome Atlas Research. N., The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163(4):1011–25.
- 3. Salami SS, et al., *Transcriptomic heterogeneity in multifocal prostate cancer*. JCI Insight, 2018. 3(21).
- 4. Wilkinson S, et al. Nascent Prostate Cancer Heterogeneity Drives Evolution and Resistance to Intense Hormonal Therapy. Eur Urol; 2021.
- 5. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science. 2018;359(6382):1350–5.
- 6. Caram ME, et al., *Factors associated with use of sipuleucel-t to treat patients with advanced prostate cancer.* 2019. 2(4): p. e192589-e192589.
- 7. Fizazi K, et al., *Abiraterone acetate for treatment of metastatic castration-resistant prostate cancer: final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study.* 2012. 13(10): p. 983–992.
- Beer TM, et al. Enzalutamide in metastatic prostate cancer before chemotherapy. 2014;371(5):424– 33.
- 9. Paller CJ, E.S.J.D.d. Antonarakis, development, and therapy, *Cabazitaxel: a novel second-line treatment for metastatic castration-resistant prostate cancer.* 2011. **5**: p. 117.
- 10. Parker C, et al., *Alpha emitter radium-223 and survival in metastatic prostate cancer.* 2013. 369(3): p. 213–223.
- 11. Smith MR, et al., *Apalutamide treatment and metastasis-free survival in prostate cancer*. 2018. 378(15): p. 1408–1418.
- 12. Aguiar PN Jr, et al. PD-L1 expression as a predictive biomarker in advanced non-small-cell lung cancer: updated survival data. Immunotherapy. 2017;9(6):499–506.
- 13. Xiao Y, Freeman GJ. The microsatellite instable subset of colorectal cancer is a particularly good candidate for checkpoint blockade immunotherapy. Cancer Discov. 2015;5(1):16–8.
- 14. Goodman AM, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. Mol Cancer Ther. 2017;16(11):2598–608.
- 15. Murciano-Goroff YR, Warner AB, Wolchok JD. The future of cancer immunotherapy: microenvironment-targeting combinations. Cell Res. 2020;30(6):507–19.
- 16. Haanen J. Converting Cold into Hot Tumors by Combining Immunotherapies. Cell. 2017;170(6):1055–6.

- 17. He Y, et al., *Classification of triple-negative breast cancers based on Immunogenomic profiling*. 2018. 37(1): p. 1–13.
- 18. Chen S, et al. Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. Nat Cell Biol. 2021;23(1):87–98.
- 19. Hänzelmann S, Castelo R. and J.J.B.b. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data*. 2013. 14(1): p. 1–15.
- 20. Kanehisa M, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353–61.
- 21. Charoentong P, et al., *Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade*. 2017. 18(1): p. 248–262.
- 22. Yoshihara K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
- 23. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ. 1998;317(7172):1572.
- 24. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–50.
- 25. Mermel CH, et al., *GISTIC2. 0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.* 2011. **12**(4): p. 1-14.
- 26. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32.
- 27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B. 1995;57:289–300.
- 28. Charoentong P, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017;18(1):248–62.
- 29. Miranda A, et al. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. Proc Natl Acad Sci U S A. 2019;116(18):9020–9.
- 30. Li M, et al. An algorithm to quantify intratumor heterogeneity based on alterations of gene expression profiles. Commun Biol. 2020;3(1):505.
- 31. Iwatsuki M, et al. Epithelial-mesenchymal transition in cancer development and its clinical significance. Cancer Sci. 2010;101(2):293–9.
- 32. Davoli T, et al., *Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy*. Science, 2017. 355(6322).
- 33. Knijnenburg TA, et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep. 2018;23(1):239–54 e6.
- 34. Mermel CH, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011;12(4):R41.
- 35. Tubbs A, Nussenzweig A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. Cell. 2017;168(4):644–56.

- 36. Mistry SJ, Bank A, Atweh GF. Targeting stathmin in prostate cancer. Mol Cancer Ther. 2005;4(12):1821–9.
- 37. Qin J, et al. Upregulation of PIP3-dependent Rac exchanger 1 (P-Rex1) promotes prostate cancer metastasis. Oncogene. 2009;28(16):1853–63.
- 38. Drobnjak M, et al. Overexpression of cyclin D1 is associated with metastatic prostate cancer to bone. Clin Cancer Res. 2000;6(5):1891–5.
- 39. Abe M, et al. Plasma levels of heat shock protein 70 in patients with prostate cancer: a potential biomarker for prostate cancer. Clin Prostate Cancer. 2004;3(1):49–53.
- 40. Dong J, et al. Hsp70 Binds to the Androgen Receptor N-terminal Domain and Modulates the Receptor Function in Prostate Cancer Cells. Mol Cancer Ther. 2019;18(1):39–50.
- 41. Ghotra VP, et al. SYK is a candidate kinase target for the treatment of advanced prostate cancer. Cancer Res. 2015;75(1):230–40.
- 42. Umbas R, et al. Expression of the cellular adhesion molecule E-cadherin is reduced or absent in highgrade prostate cancer. Cancer Res. 1992;52(18):5104–9.
- 43. Stelloo S, et al. mTOR pathway activation is a favorable prognostic factor in human prostate adenocarcinoma. Oncotarget. 2016;7(22):32916–24.
- 44. Fox JL, Storey A. BMX Negatively Regulates BAK Function, Thereby Increasing Apoptotic Resistance to Chemotherapeutic Drugs. Cancer Res. 2015;75(7):1345–55.
- 45. Van der Maaten L. and G.J.J.o.m.l.r. Hinton, Visualizing data using t-SNE. 2008. 9(11).
- 46. Trapnell C, et al. The dynamics regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. 2014;32(4):381–6.
- 47. Sanguedolce F, et al. Urine TMPRSS2: ERG Fusion Transcript as a Biomarker for Prostate Cancer: Literature Review. Clin Genitourin Cancer. 2016;14(2):117–21.
- 48. Thorsson V, et al. The Immune Landscape of Cancer. Immunity. 2018;48(4):812–30 e14.
- 49. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell. 2010;140(6):883-99.
- 50. Antunes ARP, et al., *Understanding the glioblastoma immune microenvironment as basis for the development of new immunotherapeutic strategies*. Elife, 2020. 9.
- 51. Lyu H, et al. Correlate the TP53 Mutation and the HRAS Mutation with Immune Signatures in Head and Neck Squamous Cell Cancer. Comput Struct Biotechnol J. 2019;17:1020–30.
- 52. Jiang Z, et al. Immunogenomics Analysis Reveals that TP53 Mutations Inhibit Tumor Immunity in Gastric Cancer. Transl Oncol. 2018;11(5):1171–87.
- 53. Liu Z, et al. A Comprehensive Immunologic Portrait of Triple-Negative Breast Cancer. Transl Oncol. 2018;11(2):311–29.
- 54. Patel SP, Kurzrock R. PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. Mol Cancer Ther. 2015;14(4):847–56.

#### Figures

#### Figure 1

Hierarchical clustering of prostate cancer (PC) in ten transcriptomic datasets based on the enrichment scores of 28 immune cell types. (A) There are two PC subtypes identified: PC immunity high (PC-ImH) and PC immunity low (PC-ImL). The enrichment score of an immune cell type in a tumor sample was calculated by the single-sample gene-set enrichment analysis (ssGSEA) [19] of the expression profiles of its marker gene set. (B) Comparisons of the enrichment scores of immunostimulatory signatures (CD8+ T cells, immune cytolytic activity, and interferon (IFN) response), immunosuppressive signatures (CD4+ regulatory T cells, M2 macrophages, T cell exhaustion, and *myeloid-derived suppressor cells* (*MDSCs*)), and the ratios of immunostimulatory to immunosuppressive signatures (CD8+/CD4+ regulatory T cells), between PC-ImH and PC-ImL in TCGA-PRAD.

#### Figure 2

**Comparisons of clinical and phenotypic features between the PC subtypes.** Kaplan–Meier curves to compare DFS time between the PC subtypes (A) and between high-immune-score (> mean) and low-immune-score (< mean) PC patients (B). The log-rank test P-values are shown in (A, B). Comparisons of PSA levels (C), proportion of high-grade (Gleason score of 9 to 10) tumors, proportion of late-stage (T3-4) tumors proportion of patients with cancer cells present in nearby lymph nodes, and proportion of patients with metastasis (D), scores of EMT signature (E), and ITH (F) between the PC subtypes. The one-tailed Mann–Whitney U test P values are shown in (C, E, F), and the Fisher's exact test P values and odds ratios are shown in (D). (G) Cox proportional hazards regression analysis with the response variable "DFS time" and the predictor variables "age", "PSA level", "Gleason score" and "subtype PC-ImH". The scores of EMT signature and stemness were the ssGSEA scores of their marker gene sets, and ITH scores were calculated by the DEPTH algorithm {Li, 2020 #770}. The immune scores calculated by ESTIMATE [19] represent immune infiltration levels in tumors. DFS: disease-free survival. PSA: prostate-specific antigen. EMT: epithelial-mesenchymal transition. ITH: intratumor heterogeneity. HR: hazard ratio. CI: confidence interval. \* P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001, ns P  $\ge$  0.05. It also applies to the following figures.

#### Figure 3

**Comparisons of genomic features between the PC subtypes in TCGA-PRAD.** Comparisons of TMB (**A**), HRD scores (**B**), G-scores of copy number amplifications and deletions (**C**), and enrichment scores of nine DNA damage repair pathways ((**D**) between the PC subtypes. The one-tailed Mann–Whitney *U* test *P* 

values are shown in (**A**, **B**, **D**). TMB is the total number of somatic mutations. The G-scores were calculated by GISTIC2 [34]. TMB: tumor mutation burden. HRD: homologous recombination deficiency. CNA: copy number alteration.

#### Figure 4

**Pathways enriched in the PC subtypes. (A)** The KEGG pathways involved in immune, stromal, and oncogenic signatures highly enriched in PC-ImH relative to PC-ImL in TCGA-PRAD. These pathways were identified by GSEA [24]. **(B)** Spearman correlations between the oncogenic pathways' scores and immune scores. FDR: false discovery rate. The Spearman correlation coefficients ( $\rho$ ) and *P*-values are shown.

#### Figure 5

*Genes differentially mutated between* the PC subtypes in TCGA-PRAD. (A) *8 and 25 genes with significantly lower and higher mutation frequencies in* PC-ImH than in PC-ImL, respectively (Fisher's exact test, *P* < 0.05). (B) The mutations of the 8 genes with *lower mutation frequencies in* PC-ImH correlate with lower immune scores in PC. The one-tailed Mann–Whitney *U* test *P* values are shown.

#### Figure 6

**Proteins** *differentially expressed between* the PC subtypes in TCGA-PRAD. (A) Heatmap showing 6 and 11 proteins with significant higher and lower expression levels in PC-ImH than in PC-ImL, respectively (two-tailed Student's *t* test, FDR < 0.05). (B) Spearman correlations between the expression levels of these proteins and immune scores in PC. The Spearman correlation coefficients ( $\rho$ ) and *P*-values are shown.

#### Figure 7

**Prediction performance for the PC subtypes based on the enrichment scores of 28 immune cells by the Random Forest algorithm.** TCGA-PRAD was the training set and the other nine datasets were the test sets. For TCGA-PRAD, the 10-fold cross-validation results are shown.

#### Figure 8

**Validation of the subtyping method in a** *scRNA*-seq dataset. (A) Clustering of 18792 malignant (tumor) cells and 17632 non-malignant cells by the tSNE algorithm [45]. (B) Hierarchical clustering of tumor cells based on the enrichment scores of 4 immune-related signatures identifies two subgroups in each of 13 PC patients. (C) Mapping 18792 tumor cells into immunity-high (PC-ImH) and immunity-low (PC-ImL) subgroups based on the hierarchical clustering of tumor cells in each patient. (D) Proportions of tumor cells classified into PC-ImH and PC-ImL in each of 13 PC patients. (E) PSA levels, stemness scores, ITH levels, and cell cycle activity (ssGSEA scores of the cell cycle pathway) were significantly higher in PC-ImH than in PC-ImL in the *scRNA*-seq dataset. The one-tailed Mann–Whitney *U* test or two-tailed Student's *t* test *P*-values are shown.

#### Figure 9

**Identification of two subgroups in PC-ImH.** (**A**) PC-ImH can be further divided into two subgroups: PC-ImH-A and PC-ImH-B, by the graph learning-based dimensionality reduction analysis (Trapnell et al., 2014) of PC-ImH in TCGA-PRAD and GSE116918. Comparisons of immune signatures (**B**), DFS (**C**), proportion of late-stage (T3-4) tumors (**D**), and stemness scores, ITH levels and *MKI67* expression levels (**E**) between PC-ImH-A and PC-ImH-B.

#### Figure 10

#### Association between TMPRSS2-ERG fusion and immune signatures in PC.

(A) More frequent incidence of TMPRSS2-ERG fusion in PC-ImL than in PC-ImH in TCGA-PRAD and DKFZ2018, which have TMPRSS2-ERG fusion data available. The Fisher's exact test *P*-values and odds ratios are shown. (B) The tumors with TMPRSS2-ERG fusion showing significantly lower immune signature scores than those without TMPRSS2-ERG fusion. (C) Kaplan–Meier curves showing that the tumors with TMPRSS2-ERG fusion have better DFS than those without TMPRSS2-ERG fusion.

#### **Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFig.S1.pdf
- SupplementaryTableS1.pdf
- SupplementaryTableS2.pdf