Systematical assessment of cancer stemness traits via one-class logistic regression (OCLR) algorithm and associations with tumor immune infiltrates in prostate cancer based on seven cohorts

Hongxi Chen

Hunan Normal University

Jinliang Xie

Xiangya hospital

Peng Jin (✉ jinpeng2077@sina.com)

Central South University Xiangya School of Medicine <https://orcid.org/0000-0003-4266-0034>

Primary research

Keywords: stemness indices, mRNAsi, prognosis, immunosuppression, Connectivity Map, prostate cancer

Posted Date: June 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-35734/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Prostate cancer stemness (PCS) cells have been reported to drive tumor progression, recurrence and drug resistance. However, there is lacking systematical assessment of stemness traits and associations with immunological properties in prostate adenocarcinoma (PRAD).

Methods We collected 7 PRAD cohorts with 1465 men and calculated the stemness indices for each sample using the innovative one-class logistic regression (OCLR) machine learning algorithm. We selected the mRNAsi to quantify the stemness traits that correlated significantly with prognosis and accordingly identified 21 PCS-related CpG loci and 13 pivotal signature. Meanwhile, we conducted consensus clustering and classified the total cohorts into 5 PCS clusters with distinct outcomes based on the 13-gene panel. Additionally, we implemented the CIBERSORT algorithm to infer the differential abundance across 5 PCS clusters. Lastly, we used the Connectivity Map (CMap) resource to screen potential compounds for targeting PRAD stemness.

Results: The 13-gene based PCS model possessed high predictive significance for progression-free survival (PFS) that was trained and validated in 7 independent cohorts. We found that PCScluster5 possessed the highest stemness fractions and suffered from the worst prognosis. Immune infiltration analysis shows that the activated immune cells (CD8⁺ T cell and dendritic cells) infiltrated significantly less in PCScluster5 than other clusters, especially PCScluster1, supporting the negative regulations between stemness and anticancer immunity. High mRNAsi was also found to be associated with up-regulation of immunosuppressive checkpoints, like PDL1. Finally, several potential compounds, including the top hits of cell cycle inhibitor and FOXM1 inhibitor, were identified for targeting PRAD stemness.

Conclusion: Our study comprehensively evaluated the PRAD stemness traits based on large cohorts and established a 13-gene based classifier for predicting prognosis or potential strategies for stemness treatment.

1. Background

Prostate adenocarcinoma (PRAD) was the most common urological malignancy and a leading cause of tumor-related death in men worldwide. The latest cancer statistics reported that the estimated new cases in 2020 may reach up to 191,930 and account for 21%, ranking the highest among the all body sites, in the United States(1). Though radical prostatectomy has become the main strategy for resecting the localized primary prostate tumors, nearly 20–35% of PRAD patients inevitably progressed into advanced stages with refractory recurrence and poor prognosis within 10 years (2, 3). Besides, PRAD has a disposition to metastasize to bone tissues and most of patients have to suffer from distal bone metastases (4). To date, Androgen deprivation therapy (ADT), such as enzalutamide and abiraterone, was an effective treatment for patients with early stages, but patients could not avoid to develop into androgen-resistant tumors within several years (5, 6). Radiotherapy or chemotherapy, like cyclophosphamide and methotrexate, were commonly used for metastatic and hormone-refractory PRAD

with limited efficacy on prolonging progression-free survival and various side effects (7, 8). As a result, it was essential to deeply understand the underlying mechanisms that contribute to disease progression and metastasis in PRAD, as well as novel targets for prognostic prediction and therapeutic improvements.

Stemness was defined as the potentiality to differentiate from the cell of origin and make normal cells to grow into all cell types to constitute human organism (9, 10). As reported, gradual loss of a differentiated ability and gain of stem-cell-like characteristics were the main reasons for driving tumor progression(10). As to undifferentiated PRAD, it was more likely for cancer cells to spread to distal metastases, disease recurrence, as well as refractory drug resistance. Neuroendocrine prostate cancer (NEPC) is an aggressive subpopulation of PRAD with no effective targeted drugs and worse prognosis. There were obvious features and hallmarks of NEPC incorporating (i) loss of AR signaling, p53 crosstalk, (ii) oncogenic reprogramming driven by master neural BRN2 transcription factor and (iii) elevated stemness traits associated with over-expressed SOX2 and EMT processes(11–13). Inhibition of PRAD self-renewal capacity and tumorigenicity significantly suppressed the tumor growth and NEPC differentiation process (14). Given that high hazards of relapse in PRAD has been attributed to the maintenance of PRAD stemness cells (PSC), possessing various stem cell properties that led to therapy resistance, there was an urgent to develop prognostic and/or predictive biomarkers associated with stemness. Recently, cloud computing methods and valuable bioinformatics algorithms were developed quickly in dealing with emerging big data, especially in cancer discovery. Biological top journals is reporting the pan-cancer analysis of whole genomes (PCAWG) plans and proposed several novel insights on mutational drivers, structural variations or tumor evolution, indicating the importance of high-throughput sequencing in cancer investigation (15–17). Considering the above facts, there was less reported researches to systematically depict the stemness traits in PRAD based on genome-sequencing data and large cohorts. The one-class logistic regression (OCLR) machine-learning method was a useful algorithm to quantify cancer stemness by two independent stemness indices (18, 19). One is gene expression-based stemness index (mRNAsi) that reflects the gene expression, and the another is mDNAsi which is reflective of epigenetic features (20). Previous studies have demonstrated the effective application of OCLR algorithm in multiple malignancies, and it was worthy to conduct it in assessment of PRAD stemness. Of note, PRAD include diverse contexts to together form the complexity of TME, like normal tumor cells, PCS, infiltrating immune cells, tumor-associated fibroblasts and endothelial cells (21, 22). However, there was lacking of comprehensive reports in investigating the underlying interactions or associations between PRAD stemness with tumor infiltrating immune environment.

In the current study, we totally collected 7 independent PRAD cohorts to systematically assess the PRAD stemness using OCLR algorithm (N = 1474 men). We evaluated the prognostic significance of the two indices (mRNAsi, mDNAsi), and the associations with molecular features. We identified the most hazard PCS-related CpG loci and pivotal signature, which was trained and validated in several datasets. Meanwhile, we also discussed the associations of stemness with TMB and classified distinct subclusters based on stemness traits. We further observed that high mRNAsi correlated with reduced CD8⁺ T cell

infiltrations and PDL1 expression levels. Lastly, Connectivity Map (CMap) was utilized to discover useful compounds for PRAD stemness treatment (23, 24).

2. Methods And Materials

2.1 Acquisition of PRAD cohorts and data processing

The PRAD cohorts were systematically screened and checked, which were publicly available and matched with full clinical annotations. We excluded the patients with incomplete survival information. Totally, we obtained seven treatment-naïve PRAD cohorts in our study: DKFZ-PRAD cohort, SU2C/PCF-PRAD cohort, CPC/GENE-PRAD cohort, MSKCC-PRAD cohort, GSE70769 cohort, GSE116918 cohort and TCGA-PRAD cohort. The expression profiles of TCGA cohort were obtained through data portal (<https://portal.gdc.cancer.gov/>), along with the 450K methylation data, somatic mutation data and copy number variation (CNV). Besides, the raw data of microarray datasets, based on Affymetrix and Illumina platforms, were obtained from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>). The sequencing data with clinical information of other cohorts were all downloaded from the cBioportal dataset (<https://www.cbioportal.org/>). The RMA algorithm in the Affy software package was utilized for background adjustment and quantile normalization (25). We used the *lumi* software package to process the raw data from the Illumina platform (26). Moreover, we also transformed the RNA-sequencing data (FPKM values) into transcripts per kilobase million (TPM) type, making it more comparable with those results generated from microarrays (27). The summarization of all cohorts' information was collected in Supplementary Table S1. For the clinical data that was not apparently available, three methods were adopted as following: 1. using the *GEOquery* package in R; 2. from the supplementary materials attached to the relevant literature; 3. directly downloading from the GEO website; 4. contacting with the corresponding authors for further information when necessary. The biological recurrence in PRAD patients was defined as tumor progression, and overall survival (OS) or progression-free survival (PFS) was all recognized as the end-point event.

2.2 Calculations of stemness indices for PRAD and screening of prognostic CpG islands, PCS-related signature

We utilized the one-class logistic regression (OCLR), a machine learning algorithm as previously reported, to calculating the PCS data for each TCGA-PRAD sample (18). The mRNAsi, mDNAsi and combined EREG-mRNAsi were all inferred and matched with PFS follow-up information in Table S2, respectively. The prognostic significance was assessed based on Kaplan-Meier analysis, and the correlations between mRNAsi and clinical variables were calculated based on the Pearson's correlation coefficients. The methylation status of each CpG from TCGA-PRAD cohort was defined as a beta-value ($\beta = M/(M + U)$), ranging from 0 to 1, and we utilized the minifi package to remove low-quality probes. Then, differential analysis was conducted based on limma package and the Cox regression models were used to screen the prognostic CpG sites or signature based on survival package. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed via org.Hs.eg.db, clusterprofiler

and GOplot packages. In order to evaluate the functional mechanisms that the PCS-related differentially expressed genes might participate in, GSEA analysis (Version:3.0; <http://software.broadinstitute.org/gsea/index.jsp>) was performed to identify the difference of the pathways. The annotated gene set file (msigdb.v6.2.symbols.gmt) was adopted for our analysis as the reference, and the significance was also based on the threshold of $P < 0.05$.

2.3 Evaluation of mRNAsi with TMB in TCGA-PRAD

We downloaded and prepared somatic variants with the Mutation Annotation Format (MAF) and used the “maftools” package to perform the visualization. Tumor mutation burden (TMB) was referred to the somatic genes, base substitutions, insertions or deletions, and we calculated it as: (mutation frequency with number of variants)/the length of exons (38 million) for each TCGA-PRAD sample. The associations between mRNAsi and TMB data were calculated based on Pearson’s correlation coefficients. We further classified the PRAD patients into high-TMB and low-TMB groups according to the median data, and performed Kaplan-Meier analysis based on survival package.

2.4 Consensus clustering and prognostic analysis based on PCS-related signature

Firstly, we utilized two methods of elbow method and gap statics to identify the optimized K categories, meanwhile considering the associations between PCSclusters and survival outcomes. The ConsensusClusterPlus was used and we totally repeated 1000 times to ensure the stability of classification. Then, we assessed the prognosis in each PCScluster via Kaplan-Meier analysis, respectively. Meanwhile, we also calculated the PCS indices using the OCLR algorithm across the 1474 PRAD men and conducted the survival analysis. The proportions of PCS indices (High vs. Low) were also calculated and compared in the PCSclusters, respectively.

2.5 Assessing the associations between immune infiltrates and PCS in PRAD-TME

CIBERSORT (<https://cibersort.stanford.edu/>) is a developed algorithm utilized for deconvolving the expression matrix of immune cell subtypes based on the principle of linear support vector regression (28–30). Therefore, we processed the RNA-seq (TPM normalized) data to infer the proportion of immune cells in tumor samples using the CIBERSORT algorithm and the LM22 gene signature. To guarantee the accuracy of results, we only selected the samples with $P < 0.05$ (N = 1100 from total 1424 samples). We conducted the cluster analysis across the 5 PCSclusters and analyzed the distributions of immune cells. Additionally, we utilized the ABSOLUTE package (https://software.broadinstitute.org/cancer/cga/absolute_download) to conduct the tumor purity and ploidy analysis based on the CNV results in the TCGA-PRAD cohort.

2.6 Screening of potential compounds targeting the PRAD stemness signature

The connectivity Map (CMap) dataset was a comprehensive resources to explore the underlying associations among genes, chemicals, and biological conditions. We wanted to utilize the CMap to screen potentially candidate compounds that targeting crosstalk associated with PRAD stemness. We calculated a list of differentially expressed genes between high and low mRNAsi groups and selected the top 300 to query the dataset. The query results were calculated and ranked with scores from -100 to 100. Overall, we selected the relative compounds that have an enrichment score ≤ -95 , which might be capable of targeting the PRAD stemness signature.

2.7 Statistical analysis

Student *t* test was used to compare two groups with normally distributed variables, while Wilcoxon rank-sum test was conducted to compare groups with non-normally distributed variables. As to comparisons of more than two groups, Kruskal-Wallis tests and one-way analysis of variance were selected as non-parametric and parametric methods, respectively. Correlation coefficients were assessed by Spearman analysis. Differential analysis was used based on limma package, and Kaplan-Meier analysis was conducted via survival package. Univariate- or multivariate Cox regression models were established based on survival package. The area under the curve (AUC) was utilized for assessing the predictive accuracy based on pROC curve. All statistical analyses were performed using R (Version 3.6.1) or SPSS software (version 25.0), and the statistical significance was defined with $P < 0.05$.

3. Results

3.1 Calculation of PRAD stemness indices based on mRNA expression

We collected a total of 7 PRAD cohorts with expression data and corresponding survival information to systematically characterize the stemness features of PRAD as the flowchart (Fig. 1A). The overall characteristics of PRAD cohorts were summarized in Table 1. We obtained the stemness indices from three aspects using the one-class logistic regression (OCLR) algorithm in the TCGA-PRAD cohort, including mRNAsi, mDNAsi and EREG-mRNAsi (Table S1). The stemness indices ranged from 0 to 1, and tumor samples were divided into high and low groups according to the median data of three features, respectively. Firstly, we assessed the prognostic values of three features using the Kaplan-Meier analysis, and found that patients with higher mRNAsi, mDNAsi suffered from worse progression-free survival (PFS) outcomes (Fig. 1B and 1C). However, no significance was observed in EREG-mRNAsi (Fig. 1D). We then selected the mRNAsi and mDNAsi to explore further and ranked the respective scores with well distributions (Fig. 1E and 1F). Combined with somatic mutation data, we observed that higher mRNAsi scores correlated with more frequent hallmark genes alterations, including SPOP and PTEN mutations, TP53 variations, as well as MYC amplifications (Fig. 1E). Besides, similar results were observed in Fig. 1F. In addition, we found that mRNAsi was also associated with higher TN stages, and Gleason scores, in line with the prognostic role of mRNAsi in PRAD. To assess the accuracy of inferred stemness indices, we

selected several representative stemness signature and observed the well fitness among mRNAsi and SOX2, CDC20, FOXM1, NANOG expression levels (Fig. 1H).

Table 1
Basic information of PRAD cohorts used in this study for estimating stemness traits.

Cohort Names	Platform used	Number of input	Age (Mean ± SD)	Source	Survival Outcome
DKFZ-PRAD	Illumina HumanHT-12 V3.0 expression beadchip	105	52.78 ± 9.56	OICR and ICGC	PFS
SU2C/PCF-PRAD	Illumina HumanHT-12 V3.0 expression beadchip	270	61.22 ± 8.09	Multiple countries	OS
CPC/GENE-PRAD	Illumina HumanHT-12 V3.0 expression beadchip	130	61.24 ± 7.52	Barbieri and Baca	PFS
MSKCC-PRAD	Affymetrix Human Exon 1.0 ST Array	140	NA	MSKCC	PFS
GSE70769	Illumina HumanHT-12 V4.0 expression beadchip	92	NA	Cancer Research UK Cambridge Institute	PFS
GSE116918	Almac Diagnostics Prostate Disease Specific Array (DSA) (GPL25318)	248	59.43 ± 7.17	the Northern Ireland Biobank	PFS
TCGA-PRAD	Illumina HumanHT-12 V4.0 expression beadchip	489	61.09 ± 6.82	National cancer institute	PFS

3.2 Identification of PCS-related hazard CpG loci and signature in PRAD

Based on the OCLR algorithm, mDNAsi was calculated to reflect the epigenetic features, and the mDNAsi scores correlated with prognostic outcomes. Thus, we intended to screen the PCS-related risk CpG islands. We collected the 450K methylated data and the methylation level of each methylated site was quantified by the β value, ranging from 0 (unmethylated) to 1 (fully methylated). We excluded the CpG sites for which data was missing more than 75% of the samples. We imputed the remaining missing values using the k-nearest neighbors (KNN) imputation procedure. Overall, we derived a total of 21121 CpG sites in 553 samples in our following analysis. We illustrated the top100 differential methylated levels in normal samples, low mRNAsi Pca samples and high mRNAsi Pca samples (Fig. 2A). Then, we merged the 536 PCS-associated differential CpG sites with PFS survival data in TCGA-PRAD cohort. Univariate Cox regression analysis revealed a total of 21 prognostic PCS-associated CpG islands with $P < 0.001$ (Fig. 2A, Table S2). Besides, we further compared the differential genes between low mRNAsi patients and high mRNAsi patients in heatmap (Fig. 2B). Accordingly, we screened a total of 100

differentially prognostic genes across the two groups with $P < 0.01$ and identified 13 hub PCS-related signature using the LASSO regression models, including FOXM1, CCNA2, ASF1B, AURKB, CHTF18, CDC20, CDKN3, EZH2, TRIP13, PLK1, PTTG1, MYBL2 and KIF4A (Fig. 2C, Table S3). Among them, FOXM1, EZH2 and TRIP13 were well-known master regulators mediating the stemness processes across multiple cancers. Meanwhile, we selected the top300 differentially expressed genes in two mRNAsi groups to conduct the Gene Ontology (GO) analysis and these genes were mainly enriched in stemness-associated biological items, such as notch developmental processes, cell cycle or self-renew processes (Fig. 2D). We further conducted the GSEA analysis between two groups using the expression data and found oncogenic stemness pathways were significantly enriched in high mRNAsi group, incorporating notch signaling pathway, WNT signaling pathway, oxidation phosphorylation crosstalk and EMT associated axis (Fig. 2E), further demonstrating that our signature was identified as roust PCS-associated genes.

3.3 Training and validation of 13-gene PCS-related signature in 7 PRAD cohorts

To assess the clinically predictive significance of the pivotal PCS-related signature, we collected the seven PRAD cohorts incorporating totally 1474 men to demonstrate this point. Firstly, we conducted the multivariate Cox regression analysis to calculate the PCS-related scores (PCSS) as following: $PCSS = \sum (EXP_i * \beta_i)$ ($i = 1-13$), where EXP represented the respective expression levels of genes, β represented the calculated coefficients. We classified the entire cohorts into two groups, incorporating the TCGA-PRAD cohort as the training cohort and the other cohorts as independent testing datasets. In the training cohort, we found that PCSS possessed well clinical significance for predicting progression based on the ROC curves in Fig. 3A (1 – year:0.85, 3 – year:0.78, 5 – year:0.7). Besides, patients with higher PCSS levels suffered from worse prognostic outcomes with faster disease progression than those with lower PCSS levels (log-rank test $P < 0.0001$), indicating that PCSS could be a potent indicator for hazard classification of PRAD patients (Fig. 3H). In addition, we evaluated the respective efficiency of PCSS in each PRAD testing cohort, respectively (Fig. 3B-G). In line with the training cohort, the PCSS still exhibited high predictive significance for 1-year-, 3-year-, 5-year- and 7-year-pogression with average AUC > 0.7 . Of note, we did not predict the 7-year progression in DKFZ-PRAD cohort, for the follow-up time of this cohort did not reach over 7 years. Furthermore, Kaplan-Meier analysis demonstrated the significant differential survival outcomes of PRAD cohorts between high PCSS group and low PCSS group in 6 testing datasets with $P < 0.0001$ (Fig. 3I-N). Collectively, these results suggested that the 13 hub PCS-related signature significantly correlated with PRAD progression, and the PCSS calculated based on the genes possessed robust predictive efficiency for patients.

3.3 High mRNAsi correlates with genome instability, tumor mutation burden (TMB)

We downloaded the somatic mutation data of 504 PRAD patients and calculated the TMB for each case using the formula. The TMB was merged with follow-up time and vital status in Table S4. Then, we used

the maftools package to visualize the mutational profiles and summarized them in Fig. 4A. Totally, we classified these mutation data into different categories, where missense mutation occupied the most part, single nucleotide polymorphism (SNP) mutates the most frequently and C > T was the top type of single nucleotide variants (SNV) in PRAD. Meanwhile, we exhibited the top 10 mutated genes, including well-known TP53, SPOP, MUC16 and FOXA1. In addition, we further compared the mutational difference between two groups and found that the top PRAD associated signature mutated more frequently in high mRNAsi samples than low mRNAsi samples (Fig. 4B). Moreover, we conducted correlation analysis and observed a significantly positive association between TMB and mRNAsi ($R^2 = 0.278$, $P < 0.001$). Kaplan-Meier analysis further indicated that patients with high TMB suffered from high progression risk. Collectively, we considered that mRNAsi was associated with genome TMB, which is a hazard factor in PRAD.

3.4 Consensus clustering to identify distinct stemness prognostic subgroups in total cohorts

Firstly, we collected and merged the expression data from the 7 cohorts including 1465 men, normalized to TPM type. We utilized the 13 PCS-related signature to conduct consensus clustering and identified stemness molecular subgroups of PRAD for prognostic analysis. We used the ConsensusClusterPlus package to iterate 1000 times for the stabilization of classification categories. We obtained the most optimized classification of the samples when $K = 5$ (Fig. 5A and 5B), and we illustrated the 5 clusters in Fig. 5C. We utilized the OCLR algorithm to calculate the mRNAsi for each sample and conducted the Kaplan-Meier analysis across the 5 clusters, where PCScluster5 had the worst PFS prognosis and PCScluster1 had the most favorable prognosis, in agreement with previous results (Fig. 5D). Besides, we combined the PCSclusters, mRNAsi levels (high vs. low) and vital status to draw a comprehensive sankey diagram, from which PCScluster5 exhibited more frequent relationships with high mRNAsi and progression status (Fig. 5E). Kaplan-Meier analysis also indicated that patients with high mRNAsi suffered from worse prognosis in relative to them with low mRNAsi (log-rank test $P < 0.0001$, Fig. 5F). Lastly, we found that patients with high mRNAsi occupied the most proportions in PCScluster5, and exhibited the least proportions in PCScluster1. Conversely, patients with low mRNAsi exhibited the opposite results (Fig. 5G). In total, these analysis suggested that the 13 PCS-related signature could guide molecular classifications, by which the PCSclusters had distinct prognostic outcomes and possessed different stemness proportions.

3.5 High mRNAsi correlated with lower CD8⁺ T cell infiltrates and immune suppression in PRAD-TME

Previous studies have indicated that stemness indices were associated with immune infiltrations in tumor micro-environment, altering the responses to immune checkpoint therapy for patients. Given the underlying relationships between stemness traits and immune cells, we collected and normalized the expression data of all PRAD cohorts to estimate the immune cell fractions using the CIBERSORT algorithm. After filtering out the samples with $P > 0.05$, we finally included 1000 samples to infer the

proportions of 22 immune cells in PRAD-TME. We illustrated the differentially infiltrating patterns across the 5 PCScusters and we could observe the distinct immune fractions especially between PCScuster 5 and PCScuster1 (Fig. 6A). Besides, we also described the underlying relationships across these immune cells and correlations with prognosis, among which the node sizes represented the associations with survival and line colors represented the favorable or poor outcomes (Fig. 6B). Meanwhile, the heatmap also confirmed that there was a significant difference of TIL fractions between PCScuster5 and PCScuster1, incorporating the cytotoxic cells, T cells, dendritic cells and CD8⁺ T cells. Previous results have indicated the distinct mRNAsi levels in PCScuster5 and PCScuster1, and we thus speculated that high mRNAsi correlated with suppressive immune activity in PRAD-TME. As a result, we particularly extracted the fractions of CD8⁺ T cells inferred from CIBERSORT and corresponding mRNAsi value from the eligible 750 men. Correlation analysis verified the expectation that CD8⁺ T cells infiltration was positively associated with mRNAsi levels in Fig. 6D ($R^2 = 0.81$, $P < 0.001$). In addition, we further detected that PDL1 expression levels were significantly decreased in high mRNAsi group than that in low mRNAsi group with $P < 0.0001$ (Fig. 6E). Of note, we also estimated the tumor ploidy and found that there was no significant difference between two mRNAsi groups (Fig. 6F).

3.6 Screening of potential compounds or inhibitors targeting the stemness signature based on the Connectivity Map (CMap) analysis

Connectivity Map (CMap) dataset was a data-driven, comprehensive resource for investigating relationships across genes, chemicals, and biological conditions. We thus intended to screen for candidate compounds, especially targeting the stemness-associated signature in PRAD. The top 300 differentially expressed genes were obtained from the high mRNAsi and low mRNAsi groups. The top 78 compounds that potentially target the DEGs were exhibited in Fig. 7, along with 54 mechanisms of action from the mode-of-action (MoA) analysis. The top hits revealed that 14 compounds (alvocidib, antimycin-a, apicidin, AZD-8055, BMS-345541, cercosporin, CGP-60474, chromomycin-a3, droxinostat, HG-5-113-01, LDN-193189, methylene-blue, palbociclib, PI-103) shared the MoA of cell cycle inhibitor. Meanwhile, 11 compounds shared the MoA of CDK inhibitor and 9 compounds shared the MoA of FOXM1 inhibitor, followed by mTOR inhibitor with 7 compounds and PI3K inhibitor with 6 compounds (Fig. 7).

4. Discussion

Cancer stemness cells were commonly regarded to be responsible for tumor progression and persistence, tumor recurrence and resistance to conventional therapies (9, 10, 31). We utilized the large cohorts with expression data to calculate the stemness indices and found the mRNAsi was tightly associated with progression-free survival (PFS), possessing the well fitness and quantification of stemness in PRAD. High mRNAsi were found to correlate with well-known PRAD features, including somatic SPOP mutations, PTEN loss, MYC amplifications and TP53 variations. We identified the most hazard 12 PCS-related CpG

islands and 13 vital signature. We focused on the 13-gene based model that correlates with notch signaling pathway, cell cycle and self-renew process, which was trained and validated in 1465 men to possess superior predictive value in 1-year, 3-year and 5-year PFS. Patients with high mRNAsi suffered from high TMB levels, which is a risk indicator in PRAD. We considered the tumor heterogeneity and classified the total cohorts into 5 PCSclusters based on the 13-gene signature. We observed the distinct survival outcomes across 5 PCSclusters and detected the differential mRNAsi proportions in each PCScluster. Particularly, PCScluster5 samples possessed the highest mRNAsi fractions and suffered from the worst PFS outcomes compared with other clusters. We also utilized the CIBERSORT algorithm to infer the abundance of immune infiltrating cells in PRAD-TME and found the differential infiltration patterns across 5 clusters. Importantly, we speculated that high mRNAsi correlated significantly with low infiltrations of immune activated cells, especially including CD8⁺ T cells, dendritic cells, as well as low PDL1 expression levels. Lastly, we screened the potential compounds for the PCS-related signature and found the top 3 potential drugs for stemness treatment, incorporating cell cycle inhibitor, CDK inhibitor and FOXM1 inhibitor.

Stemness signatures were already identified in several malignancies and had different prognostic value in AML, breast cancer, lung cancer or colon cancers (32-34). In our screening in multiple cohorts, the 13-gene signature was never been reported in other previous studies that could predict the PFS outcomes, which was convenient to implement in clinical practice. Cross-species regulatory network analysis confirmed that the synergistic interaction between FOXM1 and CENPF in driving drives prostate cancer malignancy (35). Interestingly, FOXM1 was the top hit in the LASSO regression results, and the CMap analysis also indicated the FOXM1 inhibitor in targeting the tumor progression. The latest research verified that FOXM1 could activate the autophagy pathway via AMPK/mTOR axis and affect the docetaxel response in CRPC, indicating the master regulatory role in altering the PRAD stemness traits (36). Given that CCNA2, CDKN3, ASF1B and KIF4A were all hub mediators in cell cycle, we further considered the combination of FOXM1 inhibitor and cell cycle inhibitor in treating PRAD stemness. Meanwhile, other pivotal signature were also tightly associated with stemness and tumorigenesis, like CCDC20, EZH2 and TRIP13. Functional analysis further demonstrated that these PCS-related DEGs were mainly enriched in notch signaling pathway, EMT and cell cycle crosstalk. Moreover, the 13-gene signature could also guide the molecular classifications and distinguish differential subpopulations with distinct prognosis. The most advantage of predictive genes is no need for further high-through sequencing and somatic mutation data, which could be optimized to a practical gene panel especially for indicating stemness and individualized classification for treatment.

Recent studies have just reported that immune cell exclusion is widely associated with the persistence of a stem cell-like phenotype across 21 solid cancer types. Pan-cancer analysis revealed that stemness indices correlated positively with higher intra-tumor heterogeneity, and tended to limit anti-tumor immune responses (37). Besides, the Polycomb Repressor Complex 1 (PRC1) was recently identified to coordinate stemness with immune suppression in double-negative prostate cancer, indicating the negative regulations between stemness and immune activation (38). In line with the current pan-cancer findings,

we also detected the differential immune contexts in distinct PCSclusters and found the PCScluster5, possessing the highest mRNAsi fraction, exhibited the significantly immune suppression. Bases on the expression data of 7 cohorts, we observed that mRNAsi correlated negatively with infiltrating levels of cytotoxic T cells. Meanwhile, high mRNAsi was associated with up-regulation of PDL1, well-known immunosuppressive checkpoints. Additionally, we also found generally positive relationships between stemness and tumor mutational load, in consistence with previous findings that mutations accumulate in normal adult stem cells. However, the efficacy of immunotherapy was relatively limited in PRAD treatment and the underlying mechanisms was not completely definite so far. Whether stemness mediated the poor immunotherapy response and could be a superior indicator for immune treatment than TMB was warranted to investigate more.

However, there existed several limitations in our work that need to be optimized in the future. To assess the well fitness between calculated mRNAsi with actual stemness traits in PRAD-TME, we conducted the correlation analysis and found the positive associations between mRNAsi with representative stem-cell identity signature, like SOX2, FOXM1 and CDC20. It was still remains unclear whether stemness indices from the bulk tumor sequencing could be utilized to represent all the PCS in samples from diverse genetic background. Besides, the 13-gene based PCS-related signature should be further validated in large samples from multi-centre PRAD patients to identify the associations with not only survival outcomes, but conventional drug responses, like ADT or chemotherapy. Considering that we found the underlying associations between mRNAsi and Gleason scores, whether stemness traits mediated the AR status or inversely altered by AR axis was intriguing to figure out. Lastly, we analyzed the CMap dataset and proposed the combination of cell cycle inhibitor with FOXM1 inhibition for targeting the PRAD stemness. However, the specific experimental validations need to be designed to assess the real inhibitory effect. Lastly, the robust associations between stemness and immune suppression was needed to demonstrate in large samples, which was an essential and worthy issue for guiding immunotherapy strategies in individualized PRAD management.

5. Conclusions

Collectively, our study systematically assessed the PRAD stemness indices based on multiple independent cohorts, providing robust quantified mRNAsi reflective of stemness, and associations with mutational burden, immune infiltrations. The 13-gene based PCS-related signature proved to be superior in predicting PFS prognosis and could be a well molecular classifier for uncovering distinct stemness clusters. Compounds screening indicated the potential directions for treating PRAD stemness, including cell cycle and FOXM1 inhibitors.

Abbreviations

PRAD, prostate adenocarcinoma; PCS, prostate cancer stemness; OCLR, one-class logistic regression; PFS, progression-free survival; OS, overall survival; CMap, Connectivity Map; ADT; androgen deprivation therapy; NEPC; neuroendocrine prostate cancer; TME, tumor microenvironment; SNP, single nucleotide

polymorphism; CNV, copy number variation; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes

Declarations

Acknowledgements

Not applicable.

Authors' contributions

Hongxi Chen, Jinliang Xie, Peng Jin conceived the concept of the paper. Hongxi Chen and Jinliang Xie conducted the analysis and draft writing.

Funding

This research was funded by the Natural Science Foundation of Hunan Province [No. 2018JJ3813]

Availability of data and materials

Additional data not presented in the manuscript can be obtained by contacting the authors.

Ethics approval and consent to participant

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020.
2. Litwin MS, Tan H-J. The Diagnosis and Treatment of Prostate Cancer: A Review. JAMA. 2017;317(24):2532-42.
3. Bill-Axelson A, Holmberg L, Garmo H, Taari K, Busch C, Nordling S, et al. Radical Prostatectomy or Watchful Waiting in Prostate Cancer - 29-Year Follow-up. The New England journal of medicine. 2018;379(24):2319-29.

4. Gartrell BA, Coleman R, Efstathiou E, Fizazi K, Logothetis CJ, Smith MR, et al. Metastatic Prostate Cancer and the Bone: Significance and Therapeutic Options. *European urology*. 2015;68(5):850-8.
5. Spratt DE, Zumsteg ZS, Feng FY, Tomlins SA. Translational and clinical implications of the genetic landscape of prostate cancer. *Nat Rev Clin Oncol*. 2016;13(10):597-610.
6. Yamamoto Y, Lin PJC, Beraldi E, Zhang F, Kawai Y, Leong J, et al. siRNA Lipid Nanoparticle Potently Silences Clusterin and Delays Progression When Combined with Androgen Receptor Cotargeting in Enzalutamide-Resistant Prostate Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2015;21(21):4845-55.
7. Rodríguez-Ruiz ME, Perez-Gracia JL, Rodríguez I, Alfaro C, Oñate C, Pérez G, et al. Combined immunotherapy encompassing intratumoral poly-ICLC, dendritic-cell vaccination and radiotherapy in advanced cancer patients. *Ann Oncol*. 2018;29(5):1312-9.
8. Kholia S, Jorfi S, Thompson PR, Causey CP, Nicholas AP, Inal JM, et al. A novel role for peptidylarginine deiminases in microvesicle release reveals therapeutic potential of PAD inhibition in sensitizing prostate cancer cells to chemotherapy. *J Extracell Vesicles*. 2015;4:26192.
9. Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. *Molecular cancer*. 2017;16(1):41.
10. Seguin L, Desgrosellier JS, Weis SM, Cheresch DA. Integrins and cancer: regulators of cancer stemness, metastasis, and drug resistance. *Trends in cell biology*. 2015;25(4):234-40.
11. Mu P, Zhang Z, Benelli M, Karthaus WR, Hoover E, Chen C-C, et al. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science (New York, NY)*. 2017;355(6320):84-8.
12. Bishop JL, Thaper D, Vahid S, Davies A, Ketola K, Kuruma H, et al. The Master Neural Transcription Factor BRN2 Is an Androgen Receptor-Suppressed Driver of Neuroendocrine Differentiation in Prostate Cancer. *Cancer discovery*. 2017;7(1):54-71.
13. Rickman DS, Beltran H, Demichelis F, Rubin MA. Biology and evolution of poorly differentiated neuroendocrine tumors. *Nature medicine*. 2017;23(6).
14. Yasumizu Y, Rajabi H, Jin C, Hata T, Pitroda S, Long MD, et al. MUC1-C regulates lineage plasticity driving progression to neuroendocrine prostate cancer. *Nature communications*. 2020;11(1):338.
15. Pan-cancer analysis of whole genomes. *Nature*. 2020;578(7793):82-93.
16. Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature*. 2019;570(7759):71-6.
17. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793).
18. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell*. 2018;173(2).
19. Lian H, Han Y-P, Zhang Y-C, Zhao Y, Yan S, Li Q-F, et al. Integrative analysis of gene expression and DNA methylation through one-class logistic regression machine learning identifies stemness features

- in medulloblastoma. *Mol Oncol*. 2019;13(10):2227-45.
20. Sokolov A, Paull EO, Stuart JM. ONE-CLASS DETECTION OF CELL STATES IN TUMOR SUBTYPES. *Pac Symp Biocomput*. 2016;21:405-16.
21. Cavarretta I, Ferrarese R, Cazzaniga W, Saita D, Lucianò R, Ceresola ER, et al. The Microbiome of the Prostate Tumor Microenvironment. *European urology*. 2017;72(4):625-31.
22. Sfanos KS, Yegnasubramanian S, Nelson WG, De Marzo AM. The inflammatory microenvironment and microbiome in prostate cancer development. *Nature reviews Urology*. 2018;15(1):11-24.
23. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*. 2017;171(6).
24. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science (New York, NY)*. 2006;313(5795):1929-35.
25. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004;20(3):307-15.
26. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008;24(13):1547-8.
27. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci*. 2012;131(4):281-5.
28. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature methods*. 2015;12(5):453-7.
29. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nature medicine*. 2015;21(8):938-45.
30. Zhao SG, Lehrer J, Chang SL, Das R, Erho N, Liu Y, et al. The Immune Landscape of Prostate Cancer and Nomination of PD-L2 as a Potential Therapeutic Target. *J Natl Cancer Inst*. 2019;111(3):301-10.
31. Saygin C, Matei D, Majeti R, Reizes O, Lathia JD. Targeting Cancer Stemness in the Clinic: From Hype to Hope. *Cell Stem Cell*. 2019;24(1):25-40.
32. Wang T, Fahrman JF, Lee H, Li Y-J, Tripathi SC, Yue C, et al. JAK/STAT3-Regulated Fatty Acid β -Oxidation Is Critical for Breast Cancer Stem Cell Self-Renewal and Chemoresistance. *Cell metabolism*. 2018;27(1).
33. Tao Y, Kang B, Petkovich DA, Bhandari YR, In J, Stein-O'Brien G, et al. Aging-like Spontaneous Epigenetic Silencing Facilitates Wnt Activation, Stemness, and Braf-Induced Tumorigenesis. *Cancer cell*. 2019;35(2).
34. Seneviratne AK, Xu M, Henao JJA, Fajardo VA, Hao Z, Voisin V, et al. The Mitochondrial Transacylase, Tafazzin, Regulates for AML Stemness by Modulating Intracellular Levels of Phospholipids. *Cell Stem Cell*. 2019;24(4).
35. Aytes A, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, et al. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that

- drives prostate cancer malignancy. *Cancer cell*. 2014;25(5):638-51.
36. Lin J-Z, Wang W-W, Hu T-T, Zhu G-Y, Li L-N, Zhang C-Y, et al. FOXM1 contributes to docetaxel resistance in castration-resistant prostate cancer by inducing AMPK/mTOR-mediated autophagy. *Cancer Lett*. 2020;469:481-9.
37. Miranda A, Hamilton PT, Zhang AW, Pattnaik S, Becht E, Mezheyski A, et al. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. *Proceedings of the National Academy of Sciences of the United States of America*. 2019;116(18):9020-9.
38. Su W, Han HH, Wang Y, Zhang B, Zhou B, Cheng Y, et al. The Polycomb Repressor Complex 1 Drives Double-Negative Prostate Cancer Metastasis by Coordinating Stemness and Immune Suppression. *Cancer cell*. 2019;36(2).

Figures

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6