Heliyon 9 (2023) e22648

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Development and validation of a circulating tumor cells-related signature focusing on biochemical recurrence and immunotherapy response in prostate cancer

Xuezhou Zhang [1], Baoan Hong [1], Zhipeng Sun, Jiahui Zhao, Mingchuan Li, Dechao Wei, Yongxing Wang, Ning Zhang [*]

Department of Urology, Beijing Anzhen Hospital, Capital Medical University, Beijing, China

ARTICLE INFO

Keywords:
prostate cancer
circulating tumor cells
biochemical recurrence
Immunotherapy
single cell data

ABSTRACT

Background: Studies have shown that the circulating tumor cells (CTCs) play a key role for invasion and formation of distant metastases in prostate cancer (PCa). However, few CTCs-related genes (CRGs) have been developed for biochemical recurrence (BCR) prediction and clinical applications of PCa patients.

Materials and methods: Bioinformatics analysis with public PCa datasets were used to investigate the relationship between the differentially expressed CRGs and BCR. Lasso-COX regression analysis was used to constructed and validated a CRGs-based BCR prediction signature for PCa. Single-cell data were used to validate the expression levels of signature genes in different cell types and then explored the cell-cell communication relationships. Finally, the expression levels of signature genes were verified and the CRGs involved in immunotherapy response were further identified.

Results: Thirteen CRGs were differentially expressed and closely associated with BCR in PCa. Then we constructed and validated a BCR prediction signature for PCa patients based on 3 differentially expressed CRGs (EMID1, SPP1 and UBE2C), and the signature was an independent factor to predict BCR for PCa. Single-cell data showed the specific expression patterns of the signature genes, while the SPP1 pathway plays an important role in cell-cell communication. Further analyses suggested UBE2C was highly expressed in BCR group and high expression of UBE2C had a better response for patients who received immunotherapy. Moreover, the expression levels of UBE2C in CTCs were higher than other cells and tissues, indicated that UBE2C may affect the BCR event of PCa patients through CTCs.

Conclusion: Our findings demonstrated that CRGs were significantly associated with BCR and immunotherapy efficacy in PCa and CRGs may influence the BCR event through CTCs.

1. Introduction

Worldwide, prostate cancer (PCa) is one of the most common malignancies in men. According to the American Cancer Society

* Corresponding author. Department of Urology, Beijing Anzhen Hospital, Capital Medical University, 2 Anzhen Road, Chaoyang District, Beijing, 100020, China.
E-mail address: Azurology_zn@126.com (N. Zhang).
[1] The authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2023.e22648
Received 31 July 2023; Received in revised form 11 November 2023; Accepted 15 November 2023
Available online 21 November 2023

statistics in 2023, PCa accounted for 288,300 new cases and 34,700 deaths with ranking first in incidence and second in mortality among male tumors [1]. Worryingly, the incidence and mortality rate of PCa is still increasing year by year, and it has been a serious danger in the world. According to the International Agency for Research on Cancer (IARC), PCa accounted for 7.1 % of new tumor patients worldwide in 2018 [2]. And in 2020, the percentage of PCa patients has risen to 7.3 % [3]. Radical prostatectomy (RP) is currently the main clinical treatment option for early stage PCa [4]. However, according to relevant statistics, biochemical recurrence (BCR) occurs in 30%–50 % of patients after RP treatment [5,6]. BCR is defined as two consecutive follow-up with both PSA>0.2 ng/ml after RP, which is a sign of PCa progression with clinical local recurrence or distant metastasis. Therefore, it is urgent and challenging to explore the factors associated with BCR in PCa.

The typical feature of advanced PCa is the development of bone metastases, which makes it difficult for patients to undergo tumor biopsy [7]. Therefore, making the search for appropriate blood tumor markers is of great significance. Circulating tumor cells (CTCs) play a key role in tumor progression and formation of distant metastases [8]. CTCs are tumor cells that were shed from the primary tumor and then invaded into the circulatory system, which can metastasize through the blood and form distant metastases at other sites [9,10]. Conventional wisdom holds that CTCs are more easily detected in patients with metastatic PCa. However, recent study showed that CTCs were detected in 25 of 27 patients with early screening PCa, suggesting that CTCs perform a relevant biological function early in the disease progression [11]. Lowes et al. included 55 PCa patients treated with RP and enumerate the patients' CTCs at 0, 6, 12, and 24 months after the treatment. The results showed that the presence of CTCs positively correlated with positive surgical margins; patients with CTCs present at any time had a shorter BCR time compared to patients with negative CTCs at any time [12]. These studies demonstrated that CTCs may be present and play a biological role in both early and late stages of PCa, even after RP. With the continuous improvement of CTCs capture technology, CTCs-based liquid biopsy technology is now widely used in clinical and basic research, but the CTCs-related genes (CRGs) associated with BCR of PCa have been rarely explored. The search for CRGs that could predict BCR will largely improve the clinical diagnosis and treatment for PCa.

Therefore, in this research we focused on exploring the value of CRGs in PCa and developed a signature to predict BCR based on CRGs (EMID1, SPP1, and UBE2C). EMID1 is an extracellular matrix (ECM) protein that enhances intracellular signaling for cell growth by inhibiting cell motility and promoting cell cycle. Extracellularly secreted EMID1 inhibits cell adhesion to ECM, which may be a novel target for tumor therapy [13]. SPP1 is a multifunctional adhesion protein that is expressed by a variety of tissue cells and plays an important role in cell fusion, motility and invasion. Studies have shown that down-regulation of SPP1 significantly inhibits the proliferation and migration ability of stomach adenocarcinoma cells [14]. SPP1 can also promote Enzalutamide resistance and epithelial-mesenchymal transition (EMT) activation in castration-resistant prostate cancer (CRPC) through PI3K/AKT and ERK1/2 pathways [15]. UBE2C is a key factor in protein ubiquitination modification, which plays an important role in tumor progression [16]. We found for the first time that high expression of UBE2C was closely associated with BCR of PCa. In addition, we collected clinical data from IMvigor210 cohort who had received immunotherapy in recent years, demonstrated that high expression of UBE2C may closely related to better response of immunotherapy in PCa.

2. Materials and methods

2.1. Data collection and acquisition

RNA-seq expression information of primary tumor samples (n = 3) and CTC samples (n = 3) of PCa were downloaded from Gene Expression Omnibus (GEO) cohort (GSE106363). The GSE147493 cohort with 99 PCa samples (no metastasis-M0: 37; metastasis-M1: 62) was included in the research. Meanwhile, the GSE116918 cohort with 248 PCa patients (non-BCR: 192; BCR: 56) also includes detailed clinical characteristics such as age, T stage, gleason grade, BCR event and follow-up time to BCR. To further differentiate cell subtypes and explore the distribution of genes in different single-cell subtypes, we used a PCa single-cell RNA-seq cohort (GSE143791) for analysis. The novel GSE70770 cohort with 205 PCa samples (non-BCR: 141; BCR: 64) was included in the research to validate the relationship between the gene level of the constructed signature and BCR event. The IMvigor210 study is a single-arm phase II clinical study of the programmed cell death-Ligand 1 (PD-L1) immune checkpoint inhibitor (ICI) Atezolizumab treatment in locally progressive or metastatic urothelial cancer after failure of platinum-based chemotherapy [17]. We present the data set used in this study in detail in Table 1.

Table 1
Characteristics of the databases used in this study.

Dataset ID	Description	Samples
GSE106363	Primary tumor vs CTCs	Primary tumor samples (n = 3) and CTC samples (n = 3)
GSE147493	No metastasis vs metastasis	M0 samples (n = 37) and M1 samples (n = 62)
GSE116918	Identification of BCR event	Non-BCR samples (n = 192) and BCR samples (n = 56)
GSE143791	Single-cell data	ScRNA-seq samples (n = 5)
GSE70770	Validation dataset	Non-BCR samples (n = 141) and BCR samples (n = 64)
IMvigor210	PD-L1 ICI treatment cohort	Immunotherapy samples (n = 348)

2.2. Differentially expressed CRGs identification

The differentially expressed genes (DEGs) between primary PCa samples and CTC samples in the GSE106363 cohort were screened by using R software package "limma" with false discovery rate (FDR) < 0.05 and $|\log2FC| > 1$ as the selection criteria. Similarly, DEGs between M0 and M1 were obtained in the same way in the GSE147493 cohort. GO enrichment analysis was performed separately for the above two gene sets. Then the intersection genes-CRGs between the two sets were obtained using a Venn diagram.

2.3. Identification of BCR-related CRGs

The GSE116918 cohort was further used to validated the correlation between the CRGs expressions and BCR event in PCa patients. Univariate Cox regression analysis was performed to determine the BCR-related CRGs by using R software. The copy number (gene-level) dataset of GDC (Genomic Data Commons) TCGA (The Cancer Genome Atlas) Prostate Cancer were downloaded from UCSC XENA database (https://xena.ucsc.edu/), the copy number variation frequency (CNV.frequency%) and the chromosome region and alteration of the BCR-related CRGs were screened using R package "RCircos".

2.4. Validation and consensus clustering analysis of CRGs

To determine the relationship between the expression levels of CRGs and the BCR event of PCa patients, we classified PCa patients into different groups by consensus clustering analysis by using the R package “ConsensusClusterPlus”. Kaplan-Meier analysis was used to compare the BCR outcomes difference between different clusters. PCA, tSNE and UMAP analyses were used to test the accuracy of the above clustering. Heatmap of BCR-related CRGs expression and corresponding clinicopathological features was analyzed and boxplot was used to show the expression patterns of those CRGs in different clusters. Immune infiltration patterns in different subtype clusters were also showed. Gene Set Variation Analysis (GSVA) and Gene Set Enrichment Analysis (GSEA) focused on the differential enrichment of KEGG pathways between the above clusters by using the R package “GSVA” and “GSEABase” [18].

2.5. Construction of BCR-related CRGs-based prediction signature

R software package "survival" was used to perform univariate Cox regression analysis to determine the BCR-related CRGs. These genes were included in a Lasso-Cox model for cross-validation and establishing a BCR-related CRGs-based signature for predicting the BCR event of PCa patients by using the R package “glmnet” and the CRGs-based signature risk score $= \Sigma(\beta i \times Expi)$ ($i =$ the number of CRGs). LASSO regression is a common construction method for clinical prediction models, which is applicable to genomics screening. Genomics usually has tens to hundreds of group features, and Lasso regression can make the coefficients of some features smaller, thus enhancing the generalization ability and adaptability of the model [19].

Patients in the cohort were also divided into high-risk group and low-risk group based on the median risk score. Kaplan-Meier analysis was used to compare the biochemical relapse-free survival (bRFS) of high- and low-risk PCa patients in the two groups. The accuracy of the signature was determined by the receiver operating characteristic (ROC) area under curve (AUC) values. Furthermore, patients with complete information on clinical characteristics were selected to assess the independence of the signature in BCR prediction. Multivariate Cox regression analysis was used to evaluate BCR prediction significance of the CRGs-based signature. A heat map which indicated the relationship between risk scores and the genes of the signature was also showed.

2.6. Identification and evaluation of a BCR predictive nomogram

Clinicopathological characteristics and risk scores were used to construct the nomogram and Time-C index was used to validate the predictive performance of the nomogram. In addition, calibration curves were plotted to assess the concordance between actual and predicted BCR and decision curve analysis (DCA) was performed to assess the clinical net benefit for PCa [20].

2.7. Single cell data and drug sensitivity analysis

Single-cell sequencing allows the study of rare cells at the cellular level or the exploration of phenotypic changes in homocytic populations and is currently widely used in tumor research [21]. To further differentiate cell subtypes and explore the distribution of CRGs-based signature genes in different single-cell subtypes, we used a PCa scRNA-seq cohort (GSE143791) for analysis. Five of the samples were randomly selected and the data were subjected to rigorous quality control using the R package “Seurat” [22]. The data were then normalized and genes with large coefficients of variation were linearly downscaled using principal component analysis (PCA). Using the R package “SingleR” [23], tSNE clustering analysis and subtype annotation were performed to determine the expression levels of CRGs-based signature genes.

Interactions between cells in different cell types and tissues provide new directions for tumor research. We used the R package “CellChat” to infer the probability of biologically meaningful cell-cell communication by integrating gene expression and interactions between signaling ligands, receptors and their cofactors [24]. The cell-cell communication network was visualized using chord diagrams. Then we used bubble diagrams to observe cell-cell communication mediated by multiple ligand receptors or signaling pathways. Similarly, we used the R package "CellChat" to identify the main senders, receivers, mediators and influencers in the intercellular communication network.

Drug sensitivity was an important indicator for predicting drug response to treatment. Based on the Genomics of Drug Sensitivity in Cancer (GDSC) database [25], the antineoplastic drugs response of each PCa sample in the high- and low-risk group was predicted by using R package "oncoPredict" and"ggplot2".

2.8. Validation of the signature gene and the relationship with immunotherapy response

The novel GSE70770 cohort (GPL-10558) with 205 PCa samples (non-BCR: 141; BCR: 64) was included in the research to validate the relationship between the gene level of the constructed signature and BCR event. Subsequently, we used the HPA (Human Protein Atlas) database to verify the expression of the above signature genes at the protein level. Then, we analyzed the signature genes and their relationship with the response of immunotherapy in IMvigor210 cohort. Finally, we analyzed the differential expression of the target gene in primary tumors, CTCs and leukocytes (white blood cells, WBC), and the differential expression of the target gene in primary tumors, CTCs and metastatic sites.

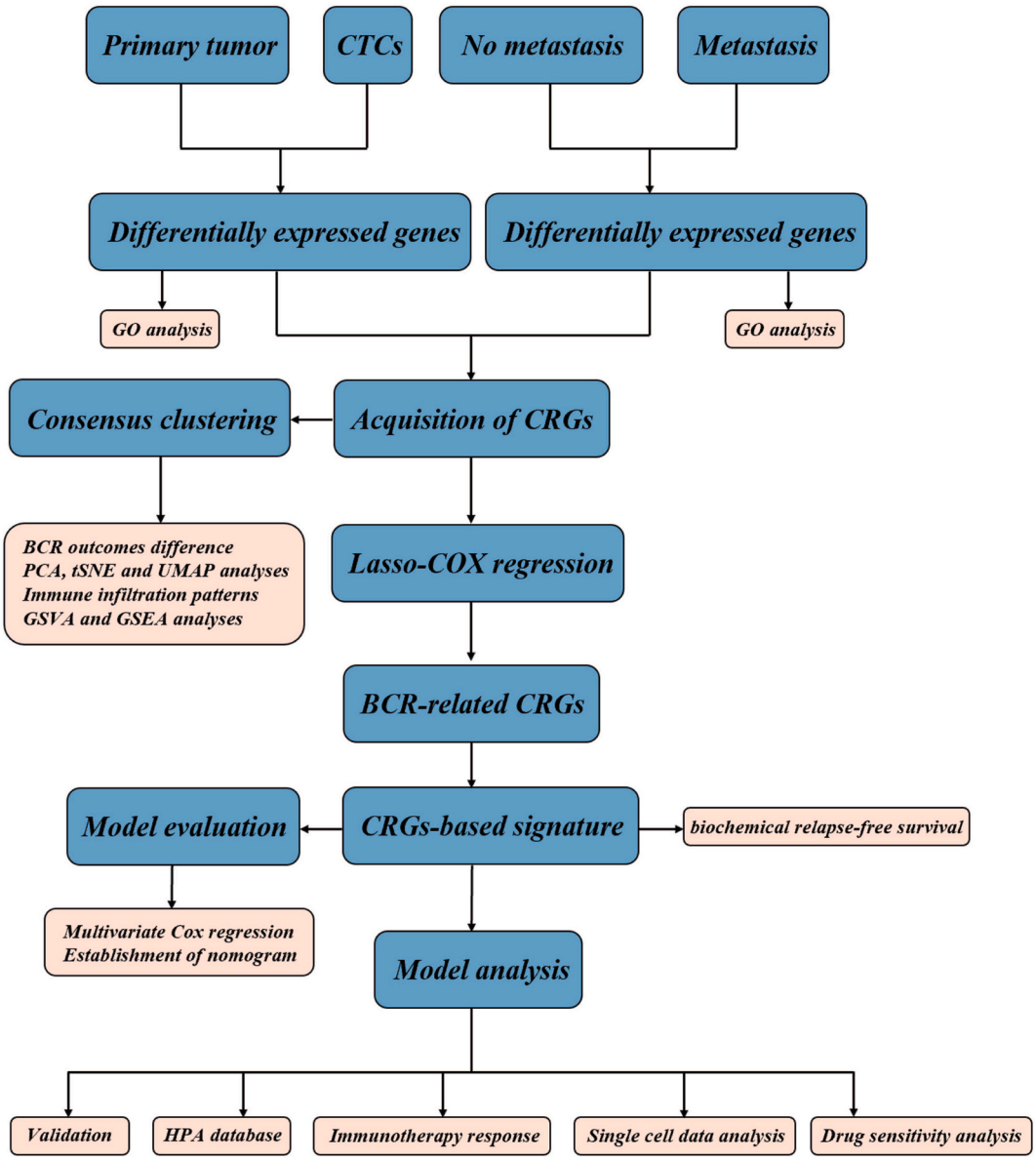
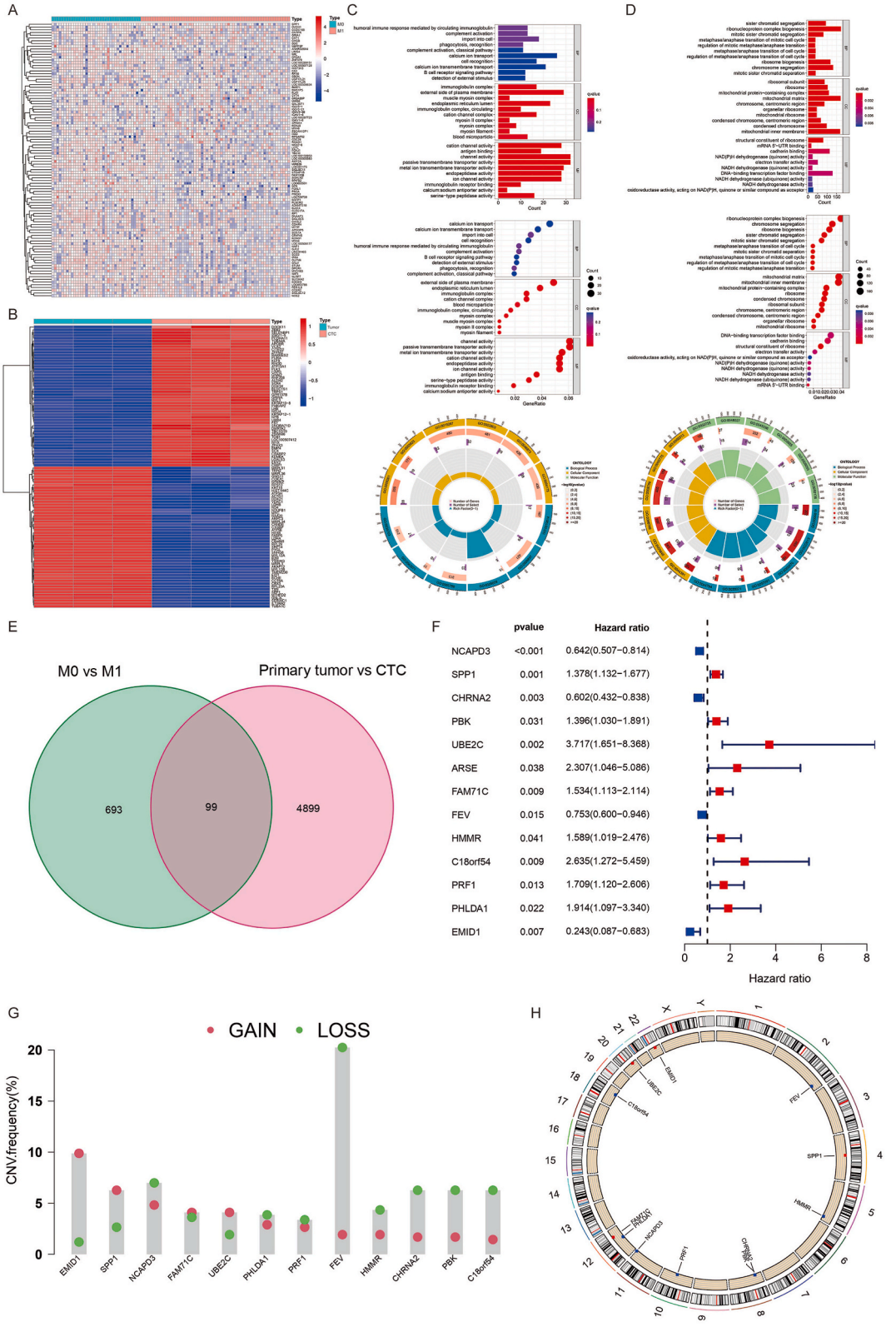


Fig. 1. The workflow provides a comprehensive description of our research.



(caption on next page)

Fig. 2. (A) The top 100 genes with the most significant variability between no metastasis PCa patients and metastasis PCa patients; **(B)** The top 100 genes with the most significant variability between primary tumors and CTCs; **(C)** The results of GO enrichment analysis in the M0 vs M1 set; **(D)** The results of GO enrichment analysis in the primary tumors vs CTCs set; **(E)** The intersection 99 genes-CRGs between the two sets with using a Venn diagram; **(F)** Univariate Cox regression analysis showed that 13 CRGs had significant BCR event value in PCa ($p < 0.05$); **(G)** The copy number variation frequency (CNV.frequency%) of the above BCR-related CRGs; **(H)** The chromosome region and alteration of the above BCR-related CRGs.

2.9. Statistical analysis

Statistical analysis was carried out using R software (version 4.2.2), and $p < 0.05$ was considered statistically significant. In this study, in order to ensure the transparency and reliability of the statistical results, we applied the multiple test correction in the differential expression analysis and the batch correction in the analysis of different datasets.

3. Results

3.1. Identification of differentially expressed genes in PCa

The workflow Fig. 1 provides a comprehensive description of our research. We got 792 DEGs between no metastasis PCa patients and metastasis PCa patients by using R package "limma" with false discovery rate (FDR) < 0.05 and $|\log2FC| > 1$ as the selection criteria. The top 100 genes with the most significant variability were shown in Fig. 2A. Similarly, we got 4998 DEGs between primary tumors and CTCs. The top 100 genes with the most significant variability were also shown in Fig. 2B.

GO enrichment analysis was performed separately for the above two gene sets. In the M0 vs M1 set, DEGs are mainly enriched in the "calcium ion transport" in the BP (biological process) part. The results of CC (cellular component) showed that DEGs were mainly concentrated in the "external side of plasma membrane". The main MF (molecular function) performed by DEGs was "channel activity" and "passive transmembrane transporter activity" (Fig. 2C). Additionally, in the primary tumors vs CTCs set, DEGs are mainly enriched in the "ribonucleoprotein complex biogenesis" in the BP part. The results of CC showed that DEGs were mainly concentrated in the "mitochondrial matrix" and "mitochondrial inner membrane". The main MF results performed by DEGs was "DNA-binding transcription factor binding" and "cadherin binding" (Fig. 2D). GO enrichment analysis circle maps were also shown. According to the above results of GO enrichment analyses, we found that the biological functions of these DEGs involved in transport and channel activity which were closely related to the recurrence and metastasis of tumors. Then we got the intersection 99 genes-CRGs between the two sets with using a Venn diagram (Fig. 2E).

3.2. Identification of CRGs related to BCR in PCa

To investigate the correlation between CRGs and BCR of PCa patients, we performed analysis of the above 99 genes and Univariate Cox regression analysis showed that 13 CRGs had significant BCR event value in PCa ($p < 0.05$) (Fig. 2F). The details were shown in Supplementary Table 1. The copy number variation frequency (CNV.frequency%) and the chromosome region and alteration of the above BCR-related CRGs were screened (Fig. 2G and H). The results showed that the number of gain in the four genes EMID1, SPP1, FAM71C and UBE2C was greater than the number of loss, while other genes were the opposite.

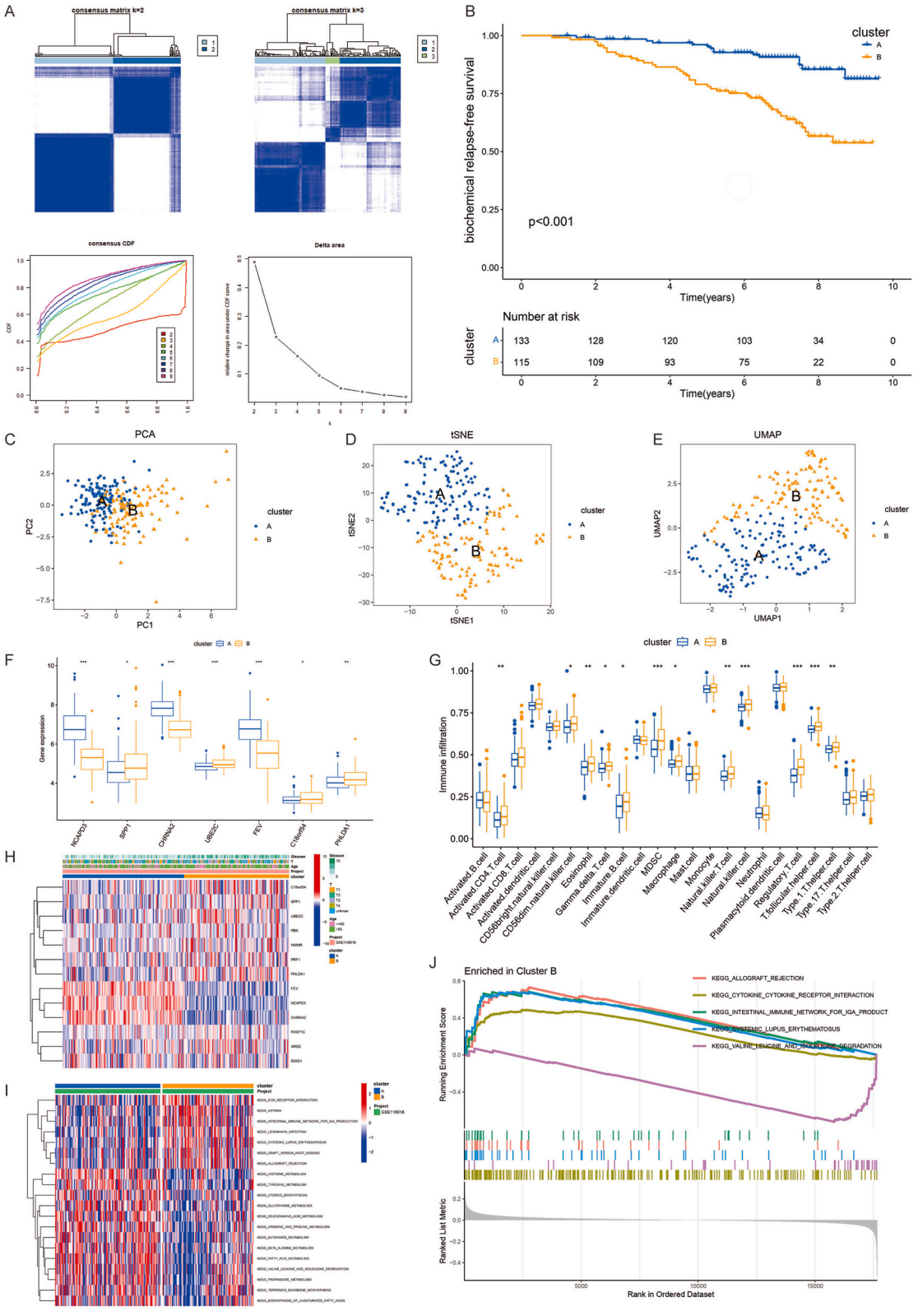
3.3. Consistent clustering of PCa molecular subgroups

To better understand the role of CRGs in PCa, we used the above 13 BCR-related CRGs for Consensus Clustering by using the R package "Consensus Cluster Plus". As shown in Fig. 3A, when $k = 2$, the cohort could be well classified into two subtypes. Biochemical relapse-free survival (bRFS) analysis showed a significant difference between the two subtypes ($p < 0.001$) and CRG-cluster B has a worse bRFS (Fig. 3B). PCA, tSNE and UMAP analyses were used to test the accuracy of this clustering and the results showed that the two clustering subtypes could be well identified (Fig. 3C–E).

Boxplot was also used to show the expression patterns of CRGs between the two subtypes (Fig. 3F). Immune infiltration patterns in the two subtype clusters were identified in Fig. 3G. In addition, the heat map was showed of the above CRGs expression and corresponding clinicopathological features of two subtypes (Fig. 3H). To explore the overall distribution of the above CRGs in the two clusters, we applied the "GSVA" R package to focus on the differential enrichment of KEGG pathways between clusters A and B (Fig. 3I). Cluster B, the poor bRFS group, was mainly involved in "ECM receptor interaction" pathway, which was crucial pathway for tumor cell invasion and metastasis. We then performed GSEA enrichment analysis and the results showed "cytokine-cytokine receptor interaction" was a significant pathway with Cluster B (Fig. 3J).

3.4. Construction and validation of BCR-related CRGs-based prediction signature

Lasso-COX regression analysis were used to construct a BCR-related CRGs-based prediction signature (Fig. 4A and B), which contained 3 CRGs (EMID1, SPP1 and UBE2C) and the coefficient of the prognostic risk score as follows: ($-1.571 \times$ EMID1's expression level) + ($0.258 \times$ SPP1's expression level) + ($1.450 \times$ UBE2C's expression level). Patients in the cohort were also divided into high-risk group and low-risk group based on the median risk score. Kaplan-Meier analysis was used to compare the bRFS of high- and low-risk



(caption on next page)

Fig. 3. (A) We used the above 13 BCR-related CRGs for Consensus Clustering by using the R package "Consensus Cluster Plus", when k = 2, the cohort could be well classified into two subtypes; **(B)** Biochemical relapse-free survival (bRFS) analysis showed a significant difference between the two subtypes ($p < 0.001$) and CRG-cluster B has a worse bRFS; **(*C*–E)** PCA, tSNE and UMAP analyses were used to test the accuracy of this clustering and the results showed that the two clustering subtypes could be well identified; **(F)** Boxplot was also used to show the expression patterns of CRGs between the two subtypes, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; **(G)** Immune infiltration patterns in the two subtype clusters were identified, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; **(H)** The heat map was showed of the above CRGs expression and corresponding clinicopathological features of two subtypes; **(I)** GSVA analysis with differential enrichment of KEGG pathways between clusters A and B; **(J)** GSEA enrichment analysis showed "cytokine-cytokine receptor interaction" was a significant pathway with Cluster B.

PCa patients in the two groups and the results showed that the CRGs-based signature was closely related to the bRFS of PCa patients ($p < 0.001$) (Fig. 4C). Patients in CRGs-based signature with high-risk score had a significantly shorter 5-year bRFS rate than patients with low-risk score. ROC curve analysis for bRFS at 1-, 3-, and 5- years showed the AUC values which indicated a good predictive performance of the CRGs-based signature (Fig. 4D).

Multivariate Cox regression analysis confirmed that the CRGs-based signature was an independent BCR event factor for PCa ($p = 0.0035$) (Fig. 4E). The heat map indicated the expression patterns of the 3 signature CRGs (Fig. 4F). The risk score was significantly different among the two previous subtypes (Fig. 4G), clusterB had a higher risk score ($p < 0.001$).

3.5. Establishment of a BCR predictive nomogram for PCa patients

We combined the CRGs-based signature with clinicopathological information to construct the nomogram (Fig. 4H). The calibration plot showed the nomogram had a good predictive performance (Fig. 4I). Decision curve analysis (DCA) is a common method for evaluating clinical predictive signatures and the results showed the above nomogram exhibited as a good method for predicting BCR in PCa patients (Fig. 4J).

3.6. Single cell data and drug sensitivity analysis of DFRGs

We used a PCa scRNA-seq cohort (GSE143791) to further differentiate cell subtypes and explore the distribution of CRGs-based signature genes in different single-cell subtypes. In the PCa scRNA-seq dataset, there were 17 cell clusters and 10 medium cell types, and the distribution and number of various cell types were shown in Fig. 5A and B. The bubble plots demonstrated the expression levels of signature genes in different cell types (Fig. 5C). The expression levels of the gene signatures of each cluster were shown in Fig. 5D and the cell subtypes can be well distinguished. Expressions and percentages showed that SPP1 is mainly expressed in macrophage while UBE2C is mainly expressed in pre-B cell $CD34^{-}$ (Fig. 5E).

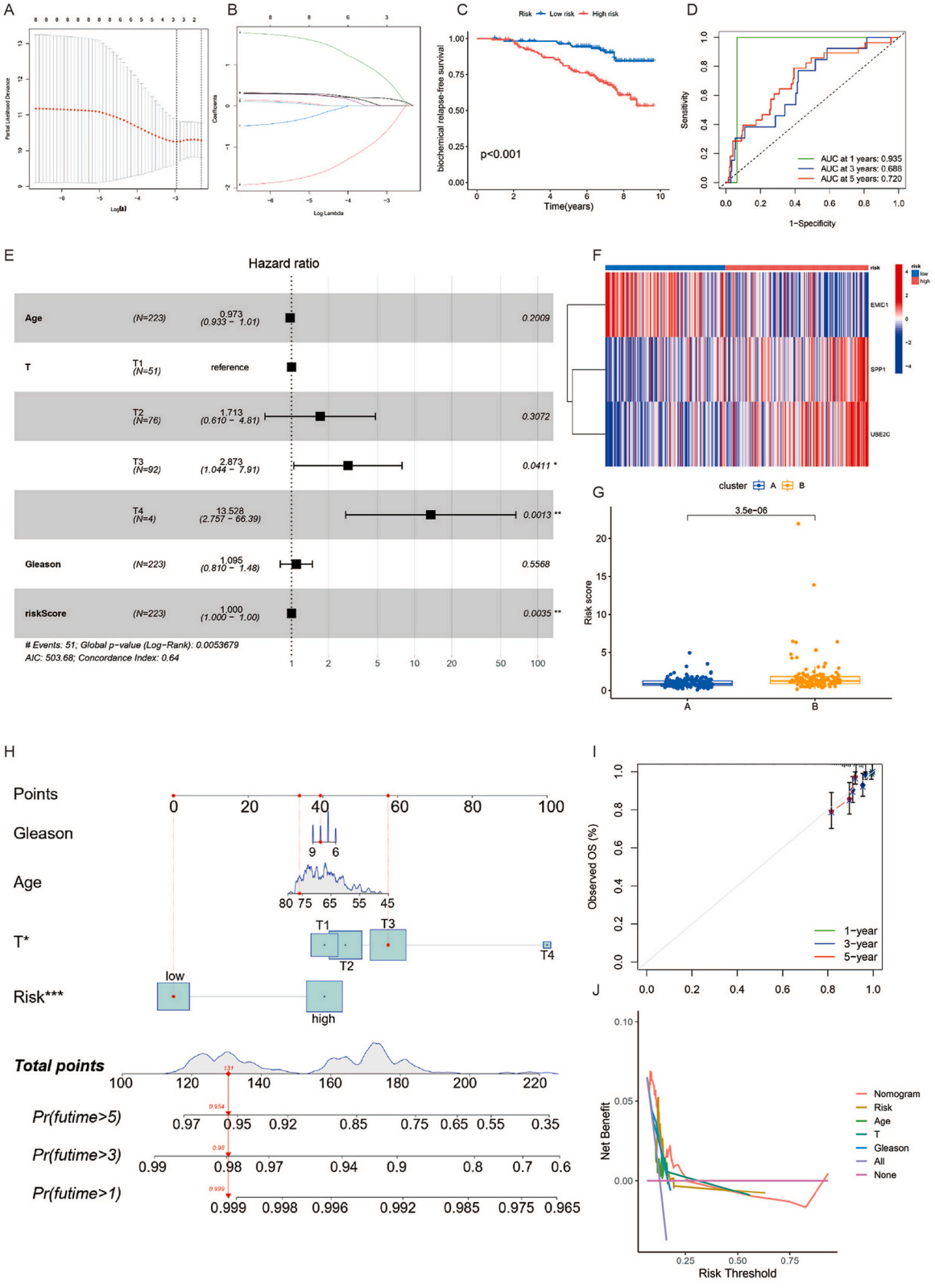
Single-cell analysis identified ten cell types. To further investigate the potential interactions between different cell types in PCa, analysis was performed with using the R package "CellChat". The results revealed that many interactions existed between different cell subpopulations, while macrophage and tissue stem cells showed the highest number of interactions with other cells (Fig. 6A). Fig. 6B demonstrated the strength of the interactions between different cell types, with thicker lines and larger nodes representing stronger interaction. Different Ligand-Receptor (LR) pairs play important roles in different cell-cell communications, and Fig. 6C showed the magnitude of the roles with different LR pairs in detail. Fig. 6D showed the interactions between each cell and other different cell types. The details were shown in Supplementary Table 2. Fig. 6E and F indicated that only macrophage can act as sender, while all cells except B cells can act as receivers to receive signals. Fig. 6G demonstrated the relationship between the action of the SPP1 pathway and LP pairs, and the results showed that macrophage can transmit signals to other cells through different LP pairs. The details were shown in Supplementary Table 3.

Next, we explored the potential sensitivity of high- and low-risk group to clinical drugs (Supplementary Fig. 1A-M). The results showed that the high-risk group was more sensitive to Paclitaxel ($p < 0.001$) and Tozasertib ($p < 0.001$).

3.7. Validation of UBE2C and the relationship with immunotherapy response

The novel GSE70770 cohort was included in the research to validate the relationship between the gene level of the constructed signature and BCR event. The results showed that EMID1 was highly expressed in non-BCR group ($p < 0.001$) (Fig. 7A), SPP1 expression levels were not significantly different between the two groups (Fig. 7B), while UBE2C was highly expressed in BCR group ($p < 0.001$) (Fig. 7C). The above results again demonstrated that high expression of UBE2C may be associated with BCR in PCa. Subsequently, we used the HPA database to verify the expression of the above signature genes at the protein level. The results revealed that UBE2C was highly expressed in PCa tissues (Fig. 7D–F).

Meanwhile, we found that high expression of UBE2C ($p < 0.001$) had a significant better response for patients who received immunotherapy, while EMID1 and SPP1 expression levels were not significantly different between the two groups (Fig. 7G–I), which maybe providing potential predictor for clinical response to ICI immunotherapy in PCa. Finally, we analyzed the differential expression of UBE2C in primary tumors, CTCs and WBC (Fig. 7J), and the differential expression of UBE2C in primary tumors, CTCs and metastatic sites (Fig. 7K). The results suggested that the expression levels of UBE2C in CTCs were higher than other cells and tissues, indicated that UBE2C may affect the BCR event of PCa patients through CTCs. Finally, the correlation between UBE2C and different biological functions was analyzed. The results showed that the expression level of UBE2C was closely related to "cell cycle" ($p < 0.01$) and "proliferation" ($p < 0.05$) (supplementary Fig.1N), suggested that UBE2C may be closely related to tumor cell proliferation



(caption on next page)

Fig. 4. (A,B) Lasso-COX regression analysis were used to construct a BCR-related CRGs-based prediction signature; **(C)** The CRGs-based signature was closely related to the bRFS of PCa patients ($p < 0.001$), patients in CRGs-based signature with high-risk score had a significantly shorter 5-year bRFS rate than patients with low-risk score; **(D)** ROC curve analysis for bRFS at 1-, 3-, and 5- years showed the AUC values which indicated a good predictive performance of the CRGs-based signature; **(E)** Multivariate Cox regression analysis confirmed that the CRGs-based signature was an independent BCR event factor for PCa ($p = 0.0035$); **(F)** The heat map indicated the expression patterns of the 3 signature CRGs; **(G)** The risk score was significantly different among the two previous subtypes, clusterB had a higher risk score ($p = 3.5e\text{-}06$); **(H)** We combined the CRGs-based signature with clinicopathological information to construct the nomogram; **(I)** The calibration plot showed the nomogram had a good predictive performance; **(J)** Decision curve analysis (DCA) showed the above nomogram exhibited as a good method for predicting BCR in PCa patients.

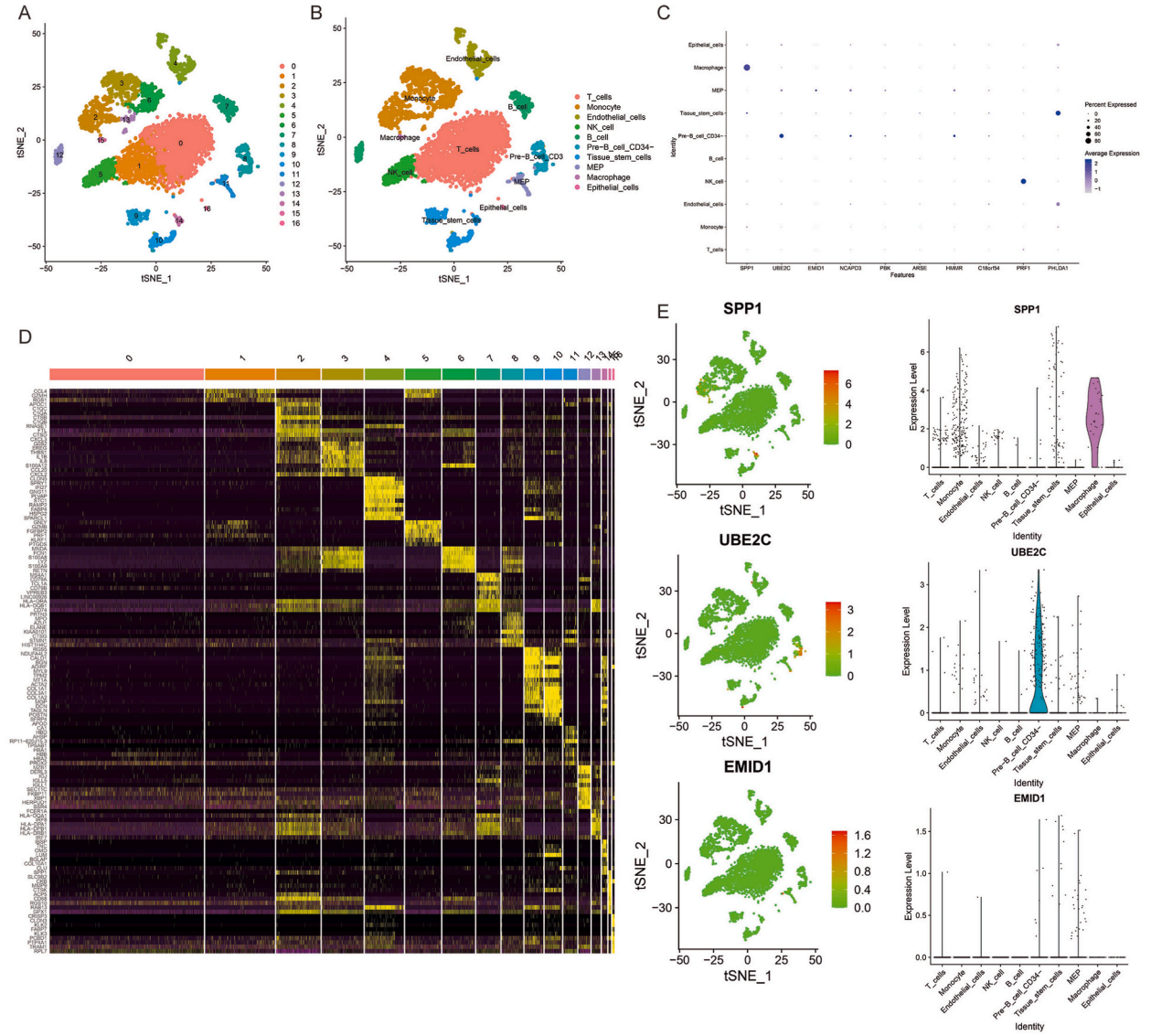
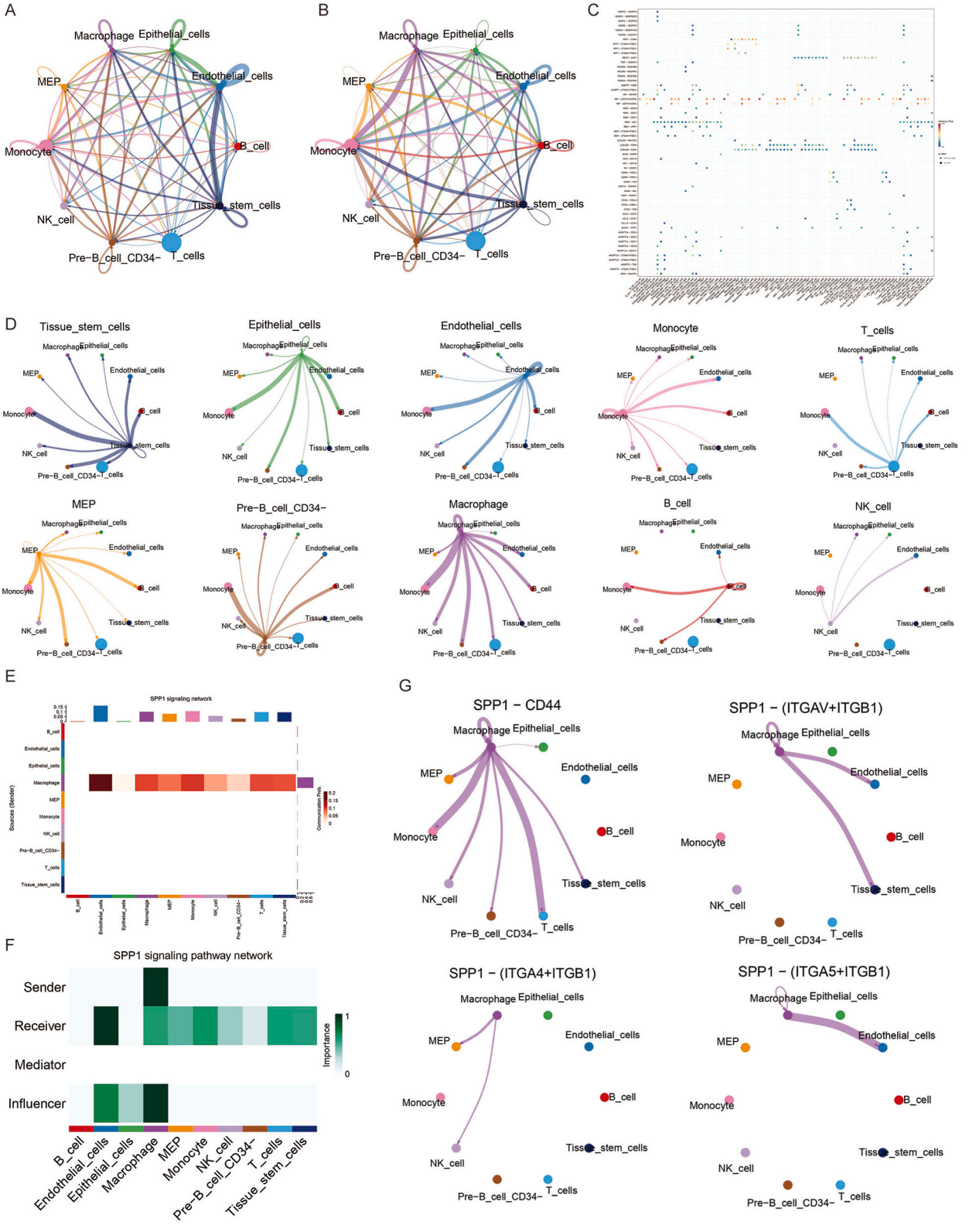


Fig. 5. (A,B) In the PCa scRNA-seq dataset, there were 17 cell clusters and 10 medium cell types, and the distribution and number of various cell types were shown; **(C)** The bubble plots demonstrated the expression levels of signature genes in different cell types; **(D)** The expression levels of the gene signatures of each cluster and the cell subtypes can be well distinguished; **(E)** Expressions and percentages showed that SPP1 is mainly expressed in macrophage while UBE2C is mainly expressed in pre-B cell CD34^{-}.

and metastasis.

4. Discussion

Recurrence and metastasis are the direct causes of death in malignant tumor patients, and hematogenous transmission is the main



(caption on next page)

Fig. 6. (A) Cell-cell communications results revealed that many interactions existed between different cell subpopulations, while macrophage and tissue stem cells showed the highest number of interactions with other cells; **(B)** The strength of the interactions between different cell types, with thicker lines and larger nodes representing stronger interaction; **(C)** The magnitude of the roles with different Ligand-Receptor (LR) pairs in different cell-cell communications; **(D)** The interactions between each cell and other different cell types in detail; **(E,F)** Only macrophage can act as sender, while all cells except B cells can act as receivers to receive signals; **(G)** The relationship between the action of the SPP1 pathway and LP pairs, while macrophage can transmit signals to other cells through different LP pairs.

route, which is a prerequisite for tumor formation of metastases [26,27]. In 1867, Australian medical scientist Ashworth first discovered blood cells similar to tumor cells in the peripheral blood of tumor patients, thus introducing the concept of CTCs [28]. In recent years, research on CTCs has gradually intensified, and several studies have shown that CTCs are necessary for the formation of invasive and distant metastases in PCa [29–31]. It is well known that the typical feature of advanced PCa is the formation of bone metastases, in which CTCs play a very important role. PCa exhibits temporal and spatial heterogeneity at different time points and in different tumor regions during disease progression [32]. However, such studies on the differences in "spatial and temporal heterogeneity" of PCa are often conducted between primary tumor foci and metastatic foci, and there is a lack of in-depth studies on CTCs, the intermediate bridge between primary foci and metastatic foci. CTCs, which are active and intact tumor cells that carry multiple histological information (genome, transcriptome, proteome, metabolome, etc.) [33]. By comparing the multi-omics features of primary foci, CTCs and metastatic foci, the key molecules in the evolution of PCa could be revealed. In this study, we innovatively performed DEGs screening based on PCa CTCs, which provides new insights into the applications of CTCs in PCa. Compared with invasive tissue biopsy, CTCs have the advantages of being noninvasive, having easy access to samples and being repeatable at different time points to dynamically assess changes in the molecular characteristics of PCa and predict efficacy and BCR.

Early stage PCa can be treated by RP. However, according to relevant statistics, a significant number of patients still develop BCR after surgery [5,6]. BCR is a sign of recurrence and metastasis of PCa. There were two types of BCR in clinical practice, one could be detected by imaging or other examinations of recurrent foci or metastases, and the other was clinically insidious and more difficult to observe. Therefore, there is an urgent need for novel and reliable biomarkers to predict BCR and metastasis in PCa. Some studies have shown that a significant proportion of patients developed metastatic recurrence after radical tumor resection, but metastases were not detected during follow-up, indicating the presence of minimal residual disease (MRD) in patients [34], which led to the widespread use of “liquid biopsy” techniques represented by CTCs [35]. CTCs had shown immeasurable clinical application in a variety of solid tumors such as colorectal cancer, breast cancer, and gastric cancer [36–38]. However, there were few reports about CTCs associated with BCR of PCa patients. In this research, we screened for CRGs associated with BCR based on DEGs of CTCs and constructed a signature, which could better predict BCR in PCa. Our findings may highlight the satisfactory clinical value of CTCs in PCa.

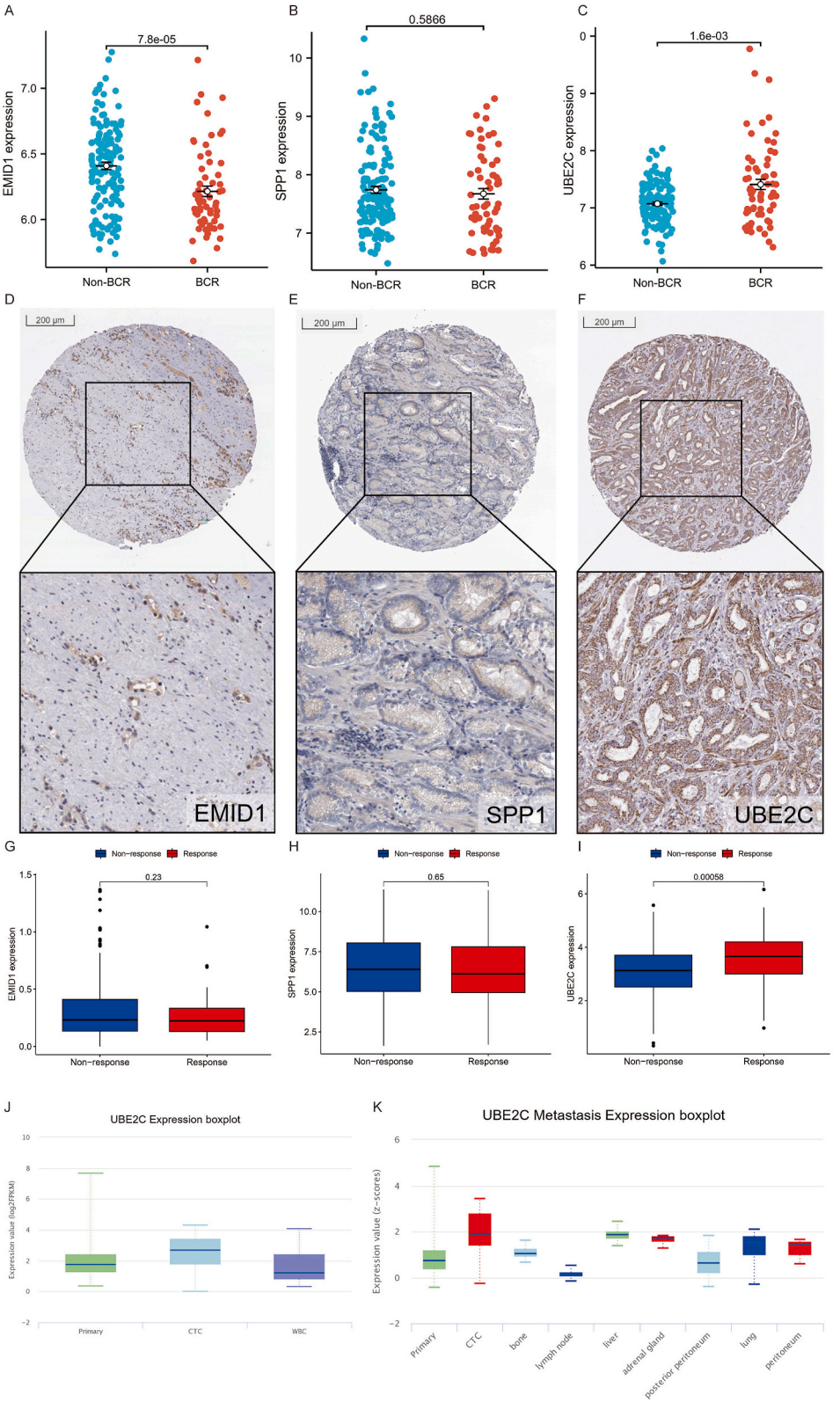
With the development of sequencing and cell separation technologies, single-cell sequencing, which explores the genome at the individual cell level, has gradually emerged, and it could better resolve the heterogeneity of genetic variation among different cells [39]. In this study, we applied the public PCa single-cell sequencing dataset and found that SPP1 is mainly expressed in macrophage while UBE2C is mainly expressed in pre-B cell $CD34^{-}$. Furthermore, based on the study of cell-cell communication relationships, the results demonstrated that the SPP1 pathway plays an important role in the interaction between different cells, and the macrophage can transmit signals to other cells through different LP pairs.

Protein ubiquitination modification is an important cellular mechanism used to target short-lived or abnormal proteins for degradation [40]. The ubiquitination process involves three classes of enzymes: ubiquitin-conjugating enzymes, ubiquitin-activating enzymes and ubiquitin-protein ligases [41]. UBE2C (ubiquitin conjugating enzyme E2 C) encodes a member of the E2 family of ubiquitin-conjugating enzymes, whose regulated proteins are used to disrupt mitotic cyclins and thus participate in cancer progression [42,43]. It was shown that high expression of UBE2C is associated with advanced histological grade and poorer prognosis of endometrial cancer, while UBE2C knockdown inhibits cell proliferation, migration, invasion and epithelial-mesenchymal transition (EMT) [44]. Zhao et al. found that UBE2C-mediated ubiquitination and degradation of SIRT1 promotes malignant progression of endometrial cancer through epigenetic inhibition of autophagy [45]. Similarly, it has been shown that UBE2C affects breast cancer proliferation through the AKT/mTOR signaling pathway [46]. Here in this study, we found that UBE2C was highly expressed in BCR group and the expression levels of UBE2C in CTCs were higher than other cells and tissues, indicated that UBE2C may affect the BCR event of PCa patients through CTCs. Our study provides a new direction for the specific mechanism of action of UBE2C in PCa.

Studies have shown that reduced EMID1 expression was significantly associated with lung adenocarcinoma progression and metastasis, while survival analysis indicated that patients with low EMID1 expression had a poorer prognosis [47]. In the present study, we found that BCR group had lower EMID1 expression levels, suggesting its potential role in predicting BCR of PCa. In this study, we analyzed the correlation between UBE2C and different biological functions. The results showed that the expression level of UBE2C was closely related to “cell cycle” and “proliferation”, suggested that UBE2C may be closely related to tumor cell proliferation and metastasis.

Another interesting finding in this study was that high expression of UBE2C had a significant better response to ICI immunotherapy. It has been demonstrated that UBE2C expression levels were associated with tumor mutational burden (TMB) in a variety of cancers, and immune correlation analysis showed that UBE2C may be involved in the immune response of cancers [48]. Therefore, we believe that UBE2C may be a potential diagnostic and therapeutic biomarker in PCa.

There are still some limitations should be considered in our research. Firstly, our study was mainly based on public databases and there was a selection bias in the design. Secondly, the IMvigor210 cohort is a study of ICI treatment in locally progressive or metastatic urothelial cancer, more data on immunotherapy in PCa patients are still needed to supplement. Finally, our study focused on the



(caption on next page)

Fig. 7. (A) EMID1 was highly expressed in non-BCR group (p = 7.8e-05); **(B)** SPP1 expression levels were not significantly different between the two groups (p = 0.5866), **(C)** UBE2C was highly expressed in BCR group (p = 1.6e-03); **(D–F)** The expression of EMID1, SPP1 and UBE2C at the protein level of PCa tissues in the HPA database, scale bars: 200 μm; **(G)** EMID1 expression levels were not significantly different between non-response and response groups for patients who received immunotherapy (p = 0.23); **(H)** SPP1 expression levels were not significantly different between non-response and response groups for patients who received immunotherapy (p = 0.65); **(I)** High expression of UBE2C (p = 0.00058) had a significant better response for patients who received immunotherapy; **(J)** The expression levels of UBE2C in CTCs were higher than primary tumor and WBC; **(K)** The expression levels of UBE2C in CTCs were higher than primary tumor and metastatic sites.

expression and BCR prediction value of CRGs, further experiments are needed to explore the potential roles especially the mechanism of UBE2C in PCa based on CTCs.

In the future, we will continue to conduct in vivo and in vitro experiments based on the explored CRGs, especially to study the interaction between CRGs and androgen receptors (AR). If possible, we will try to use the above CRGs as targets for molecular imaging of CTCs to analyze their chemotaxis and colonization abilities. Further studies on CRGs will provide new perspectives on the clinical treatment and pathogenesis of PCa, ultimately advancing cancer research.

5. Conclusion

It was the first time to investigate the relationship of CTCs and BCR in PCa. In contrast to other liquid biopsy techniques, CTCs are active, intact tumor cells that carry multi-omics information about the tumor cells. Key molecules in the evolution of PCa could be revealed through the study of CTCs. In this study, we constructed and validated a BCR prediction signature for PCa patients based on 3 differentially expressed CRGs (EMID1, SPP1 and UBE2C), and the signature was an independent factor to predict BCR for PCa. Further analyses suggested UBE2C was highly expressed in BCR group and high expression of UBE2C had a better response for patients who received immunotherapy. Moreover, the expression levels of UBE2C in CTCs were higher than other cells and tissues, indicated that UBE2C may affect the BCR event of PCa patients through CTCs. Based on the above CRGs, we will continue to study them in depth in the future, especially to explore the clinical translational value of CRGs as well as the potential mechanisms, with a view to providing new directions for PCa research.

Fundings

This study was supported by Capital's Funds for Health Improvement and Research (2020-2-1024), Xisike-leader Oncology Research Foundation (Y-2019AZZD-0369) and Beijing Bethune Charitable Foundation (mnzl202003).

Ethical approval

Not applicable.

Data availability statement

The total data of GSE106363, GSE147493, GSE116918, GSE143791 and GSE70770 dataset acquired from GEO (https://www.ncbi.nlm.nih.gov/geo/) database.

CRediT authorship contribution statement

Xuezhou Zhang: Data curation, Software, Validation, Writing – original draft. **Baoan Hong:** Conceptualization, Data curation, Resources, Writing – review & editing. **Zhipeng Sun:** Resources, Validation. **Jiahui Zhao:** Data curation, Resources, Validation. **Mingchuan Li:** Resources, Visualization. **Dechao Wei:** Resources. **Yongxing Wang:** Resources, Supervision. **Ning Zhang:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix ASupplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e22648.

References

[1] R.L. Siegel, K.D. Miller, N.S. Wagle, et al., Cancer statistics, 2023, CA Cancer J Clin 73 (1) (2023) 17–48.
[2] F. Bray, J. Ferlay, I. Soerjomataram, et al., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 68 (6) (2018) 394–424.
[3] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 71 (3) (2021) 209–249.
[4] T.J. Wilt, K.M. Jones, M.J. Barry, et al., Follow-up of prostatectomy versus Observation for early prostate cancer, N. Engl. J. Med. 377 (2) (2017) 132–142.
[5] E. Lalonde, A.S. Ishkanian, J. Sykes, et al., Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study, Lancet Oncol. 15 (13) (2014) 1521–1532.
[6] Y.H. Shao, K. Demissie, W. Shih, et al., Contemporary risk profile of prostate cancer in the United States, J Natl Cancer Inst 101 (18) (2009) 1280–1283.
[7] L.C. Hofbauer, A. Bozec, M. Rauner, et al., Novel approaches to target the microenvironment of bone metastasis, Nat. Rev. Clin. Oncol. 18 (8) (2021) 488–505.
[8] E. Schuster, R. Taftaf, C. Reduzzi, et al., Better together: circulating tumor cell clustering in metastatic cancer, Trends Cancer 7 (11) (2021) 1020–1032.
[9] R. Agashe, R. Kurzrock, Circulating tumor cells: from the laboratory to the cancer clinic, Cancers 12 (9) (2020) 2361.
[10] A. Ring, B.D. Nguyen-Sträuli, A. Wicki, et al., Biology, vulnerabilities and clinical applications of circulating tumour cells, Nat. Rev. Cancer 23 (2) (2023) 95–111.
[11] K. Ried, T. Tamanna, S. Matthews, et al., New screening test improves detection of prostate cancer using circulating tumor cells and prostate-specific markers, Front. Oncol. 10 (2020) 582.
[12] L.E. Lowes, M. Lock, G. Rodrigues, et al., The significance of circulating tumor cells in prostate cancer patients undergoing adjuvant or salvage radiation therapy, Prostate Cancer Prostatic Dis. 18 (4) (2015) 358–364.
[13] T. Kawata, K. Muramatsu, N. Shishito, et al., EMID1, a multifunctional molecule identified in a murine model for the invasion independent metastasis pathway, Sci. Rep. 11 (1) (2021), 16372.
[14] Y. Wu, L. Ren, Y. Tang, et al., Immunobiological signatures and the emerging role of SPP1 in predicting tumor heterogeneity, malignancy, and clinical outcomes in stomach adenocarcinoma, Aging (Albany NY) 15 (20) (2023) 11588–11610.
[15] X. Pang, J. Zhang, X. He, et al., SPP1 promotes Enzalutamide resistance and epithelial-mesenchymal-transition activation in castration-resistant prostate cancer via PI3K/AKT and ERK1/2 pathways, Oxid. Med. Cell. Longev. 2021 (2021), 5806602.
[16] S. Ma, Q. Chen, X. Li, et al., UBE2C serves as a prognosis biomarker of uterine corpus endometrial carcinoma via promoting tumor migration and invasion, Sci. Rep. 13 (1) (2023 Oct 6), 16899.
[17] J.E. Rosenberg, J. Hoffman-Censits, T. Powles, et al., Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial, Lancet 387 (10031) (2016) 1909–1920.
[18] S. Hnzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-Seq data, BMC Bioinf. 14 (1) (2013), 7-7.
[19] Z.R. Zhou, W.W. Wang, Y. Li, et al., In-depth mining of clinical data: the construction of clinical prediction model with R, Ann. Transl. Med. 7 (23) (2019) 796.
[20] A.J. Vickers, A.M. Cronin, E.B. Elkin, et al., Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers, BMC Med Inform Decis Mak 8 (2008) 53.
[21] D. Jovic, X. Liang, H. Zeng, et al., Single-cell RNA sequencing technologies and applications: a brief overview, Clin. Transl. Med. 12 (3) (2022) e694.
[22] B. Li, W. Zhang, C. Guo, et al., Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution, Nat. Methods 19 (6) (2022) 662–670.
[23] X. Zhao, S. Wu, N. Fang, et al., Evaluation of single-cell classifiers for single-cell RNA sequencing data sets, Brief Bioinform 21 (5) (2020) 1581–1595.
[24] S. Jin, C.F. Guerrero-Juarez, L. Zhang, et al., Inference and analysis of cell-cell communication using CellChat, Nat. Commun. 12 (1) (2021) 1088.
[25] W. Yang, J. Soares, P. Greninger, et al., Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells, Nucleic Acids Res. 41 (Database issue) (2013) D955–D961.
[26] J.D. Hebert, J.W. Neal, M.M. Winslow, Dissecting metastasis using preclinical models and methods, Nat. Rev. Cancer 23 (6) (2023) 391–407.
[27] K.E. de Visser, J.A. Joyce, The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth, Cancer Cell 41 (3) (2023) 374–403.
[28] T.R. Ashworth, A case of cancer in which cells similar to those in the tumours were seen in the blood after death, Aust Med J 14 (1869) 146.
[29] E.S. Barnett, N. Schultz, K.H. Stopsack, et al., Analysis of BRCA2 copy number loss and genomic instability in circulating tumor cells from patients with metastatic castration-resistant prostate cancer, Eur. Urol. 83 (2) (2023) 112–120.
[30] L. Groen, I. Kloots, D. Englert, et al., Transcriptome profiling of circulating tumor cells to predict clinical outcomes in metastatic castration-resistant prostate cancer, Int. J. Mol. Sci. 24 (10) (2023) 9002.
[31] C.R. Davies, T. Guo, E. Burke, et al., The potential of using circulating tumour cells and their gene expression to predict docetaxel response in metastatic prostate cancer, Front. Oncol. 12 (2023), 1060864.
[32] I. Casanova-Salas, A. Athie, P.C. Boutros, et al., Quantitative and qualitative analysis of blood-based liquid biopsies to inform clinical decision-making in prostate cancer, Eur. Urol. 79 (6) (2021) 762–771.
[33] K.M. Mahuron, Y. Fong, Applications of liquid biopsy for surgical patients with cancer: a review, JAMA Surg (2023). Published online November 01, 2023.
[34] H. Pan, R. Gray, J. Braybrooke, et al., 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years, N. Engl. J. Med. 377 (19) (2017) 1836–1846.
[35] M.L. Cheng, E. Pectasides, G.J. Hanna, et al., Circulating tumor DNA in advanced solid tumors: clinical relevance and future directions, CA: a cancer journal for clinicians 71 (2) (2021) 176–190.
[36] D. Kolenčík, S.N. Shishido, P. Pitule, et al., Liquid biopsy in colorectal carcinoma: clinical applications and challenges, Cancers 12 (6) (2020) 1376.
[37] R. Piñeiro, I. Martínez-Pena, R. López-López, Relevance of CTC clusters in breast cancer metastasis, Circulating Tumor Cells in Breast Cancer Metastatic Disease (2020) 93–115.
[38] P. Thanh Huong, S. Gurshaney, N. Thanh Binh, et al., Emerging role of circulating tumor cells in gastric cancer, Cancers 12 (3) (2020) 695.
[39] B. Van de Sande, J.S. Lee, E. Mutasa-Gottgens, et al., Applications of single-cell RNA sequencing in drug discovery and development, Nat. Rev. Drug Discov. (2023) 1–25.
[40] J. Faktor, M. Pjechová, L. Hernychová, et al., Protein ubiquitination research in Oncology, Fall, Klin. Onkol. 32 (Supplementum 3) (2019) 56–64. English.
[41] K.N. Swatek, D. Komander, Ubiquitin modifications, Cell Res. 26 (4) (2016 Apr) 399–422, https://doi.org/10.1038/cr.2016.39.
[42] P. Garg, D.F. Ceccarelli, A.F.A. Keszei, et al., Structural and functional analysis of ubiquitin-based inhibitors that target the backsides of E2 enzymes, J. Mol. Biol. 432 (4) (2020) 952–966, https://doi.org/10.1016/j.jmb.2019.09.024. Epub 2019 Oct 19. PMID: 31634471; PMCID: PMC7054140.
[43] A. Williamson, K.E. Wickliffe, B.G. Mellone, et al., Identification of a physiological E2 module for the human anaphase-promoting complex, Proc Natl Acad Sci U S A 106 (43) (2009) 18213–18218.
[44] Y. Liu, R. Zhao, S. Chi, et al., UBE2C is upregulated by estrogen and promotes epithelial-mesenchymal transition via p53 in endometrial cancer, Mol. Cancer Res. 18 (2) (2020) 204–215.
[45] R. Zhao, Y. Liu, Z. Wang, et al., UBE2C-mediated autophagy inhibition via ubiquitination of SIRT1 contributes to endometrial cancer progression, Mol. Cancer Res. 21 (6) (2023) 564–577.
[46] Z.N. Lu, J. Song, T.H. Sun, et al., UBE2C affects breast cancer proliferation through the AKT/mTOR signaling pathway, Chin Med J (Engl). 134 (20) (2021) 2465–2474.
[47] Y. Shao, Z. Zheng, S. Li, et al., Upregulation of EMID1 accelerates to a favorable prognosis and immune infiltration in lung adenocarcinoma, J Oncol 2022 (2022), 5185202.
[48] G.Z. Huang, Z.Q. Chen, J. Wu, et al., Pan-cancer analyses of the tumor microenvironment reveal that ubiquitin-conjugating enzyme E2C might Be a potential immunotherapy target, J Immunol Res 2021 (2021), 9250207.