Identification of Metabolism-Associated Prostate Cancer Subtypes and Construction of a Prognostic Risk Model

Yanlong Zhang[1,2†], Ruiqiao Zhang[1,2], Fangzhi Liang[1,2], Liyun Zhang[3*‡] and Xuezhi Liang[1*‡]

[1] Department of Urology, First Hospital of Shanxi Medical University, Taiyuan, China, [2] First Clinical Medical College, Shanxi Medical University, Taiyuan, China, [3] Department of Rheumatology, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Taiyuan, China

Background: Despite being the second most common tumor in men worldwide, the tumor metabolism-associated mechanisms of prostate cancer (PCa) remain unclear. Herein, this study aimed to investigate the metabolism-associated characteristics of PCa and to develop a metabolism-associated prognostic risk model for patients with PCa.

Methods: The activity levels of PCa metabolic pathways were determined using mRNA expression profiling of The Cancer Genome Atlas Prostate Adenocarcinoma cohort *via* single-sample gene set enrichment analysis (ssGSEA). The analyzed samples were divided into three subtypes based on the partitioning around medication algorithm. Tumor characteristics of the subsets were then investigated using t-distributed stochastic neighbor embedding (t-SNE) analysis, differential analysis, Kaplan–Meier survival analysis, and GSEA. Finally, we developed and validated a metabolism-associated prognostic risk model using weighted gene co-expression network analysis, univariate Cox analysis, least absolute shrinkage and selection operator, and multivariate Cox analysis. Other cohorts (GSE54460, GSE70768, genotype-tissue expression, and International Cancer Genome Consortium) were utilized for external validation. Drug sensibility analysis was performed on Genomics of Drug Sensitivity in Cancer and GSE78220 datasets. In total, 1,039 samples and six cell lines were concluded in our work.

Results: Three metabolism-associated clusters with significantly different characteristics in disease-free survival (DFS), clinical stage, stemness index, tumor microenvironment including stromal and immune cells, DNA mutation (*TP53* and *SPOP*), copy number variation, and microsatellite instability were identified in PCa. Eighty-four of the metabolism-associated module genes were narrowed to a six-gene signature associated with DFS, *CACNG4*, *SLC2A4*, *EPHX2*, *CA14*, *NUDT7*, and *ADH5* (p <0.05). A risk model was developed, and external validation revealed the strong robustness our risk model possessed in diagnosis and prognosis as well as the association with the cancer feature of drug sensitivity.

Conclusions: The identified metabolism-associated subtypes reflected the pathogenesis, essential features, and heterogeneity of PCa tumors. Our metabolism-associated risk model may provide clinicians with predictive values for diagnosis, prognosis, and treatment guidance in patients with PCa.

Keywords: prostate cancer, metabolism-associated subtype, risk model, tumor heterogeneity, immunotherapy

INTRODUCTION

Prostate cancer (PCa) is the second most frequent urinary system-associated type of cancer, accounting for 13% of all malignant tumors in men (1). Radical prostatectomy (RP) has been used to cure PCa patients by removing the malignant prostate. However, the recurrence rates after the surgery are high. Recurrent cancer has risks of developing into castration-resistant PCa, which will either continue progressing the pre-existing PCa or spreading cancer to other parts of the body (2). Therefore, exploring the tumor characteristic and finding a new therapy for PCa remains crucial. Furthermore, identifying biomarkers for disease-free survival (DFS) is needed to improve patients' prognosis with PCa.

Due to the unrestricted multiplicative nature of cancer cells, tumors exhibit different metabolic statues from normal tissue, thus provide a possible way to identify tumors through the difference in metabolism. Recent studies have proven that some metabolisms, such as citrate and choline metabolism, are closely related to PCa (3). Studies have also shown that based on the variance in metabolites, such as increased urea cycle metabolites, PCa can be characterized (4).

Classification analysis based on a large number of samples that can better reflect tumor features and heterogeneity becomes possible with the advent of high-throughput sequencing. HIgh-throughput sequencing has been successfully applied to classify subtypes in different cancers. Subtypes are then used to either guide immune therapy, portray multiple dimensions of tumor characteristics, or assist patient prognosis prediction (5, 6). Although many genome-wide analyses have been performed in regards to PCa, there has been a lack of hierarchical cluster analyses of the PCa transcriptome to exploring tumor metabolic features. Meanwhile, almost all previous studies were based on PCa tumor metabolism concentrated on individual tumor cells rather than mixed tissue, including tumor cells, stromal cells, and immune cells. These studies and therefore do not reflect the metabolic characteristics of PCa *in vivo* (7). So a hierarchical cluster analysis of the PCa transcriptome from a metabolic view to exploring tumor heterogeneity is therefore crucial.

Based on the information above, we performed unsupervised clustering to explore the potential metabolism-associated subtypes and explored the correlations between the subtypes and tumor heterogeneity. Biomarkers associated with subtypes were also selected. Finally, a risk model to predict PCa patients' prognosis was constructed. We hypothesize the metabolism-associated characteristics of PCa to understand the PCa metabolic mechanism better and further identify tumors. The risk model will be able to guide the PCa diagnosis, prognosis, and treatment.

MATERIALS AND METHODS

Data Collection

Gene expression files, DNA mutation data, and copy number variation (CNV) of prostate adenocarcinoma (PRAD) tissues were downloaded from TCGA (https://portal.gdc.cancer.gov/). Gene expression data were acquired using the Illumina HiSeq RNA Sequencing platform and expressed as fragments per kilobase of transcript per million fragments (FPKM). The cBioPortal for Cancer Genomics (https://www.cbioportal.org/) provided clinical data of the PRAD patients (8). RNA sequencing (RNA-seq) data of normal prostate tissues from testing cohorts for diagnosis were obtained from the Genotype-Tissue Expression (GTEx) (https://www.gtexportal.org/) and tumor tissue from the ICGC (https://icgc.org/). RNA-seq and microarray data of PRAD tissues and clinical information from testing cohorts for prognosis were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/gds/). The GEO search strategy of the GSE datasets was as follows: 1) Include "prostate cancer" and dataset types of RNA-seq or micro-array; 2) Include more than one hundred PRAD samples with survival data; and 3) Include expression information of six risk model genes. Two datasets that met these requirements were identified, GSE54460 and GSE70768. Microarray data of cell lines (including 22RV1, DU-145, LNCaP-Clone-FGC, PC-3, PWR-1E, and VCaP), and RNA-seq data of melanoma samples were downloaded from the Genomics of Drug Sensitivity in Cancer (GDSC) (https://www.cancerrxgene.org/) and GSE78220. We also acquired the immunohistochemistry (IHC) data for PRAD and normal prostate tissues from the Human Protein Atlas (HPA) data portal (https://www.proteinatlas.org/). Immune infiltrate data for PRAD tissues were downloaded from the Cistrome Project (http://www.cistrome.org/) using the Tumor IMmune Estimation Resource version 2.0 (TIMER2.0) (9).

ssGSEA Assessment of Metabolism-Associated Pathways Expression Levels

Data for 41 metabolism pathway gene sets were acquired from Molecular Signatures Database (MSigDB; https://www.gsea-msigdb.org/) (10) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (11) and the PCa related activity levels were calculated using ssGSEA and the gene set variance analysis (GSVA) R package 1.34.0 (10). The metabolism-associated signatures used included galactose metabolism, ascorbate and aldarate metabolism, fatty acid metabolism, purine metabolism, pyrimidine metabolism, alanine aspartate and glutamate metabolism, glycine serine and threonine metabolism, cysteine and methionine

metabolism, arginine and proline metabolism, histidine metabolism, tyrosine metabolism, phenylalanine metabolism, tryptophan metabolism, beta alanine metabolism, taurine and hypotaurine metabolism, selenoamino acid metabolism, glutathione metabolism, starch and sucrose metabolism, amino sugar and nucleotide sugar metabolism, glycerolipid metabolism, inositol phosphate metabolism, glycerophospholipid metabolism, ether lipid metabolism, arachidonic acid metabolism, linoleic acid metabolism, alpha linolenic acid metabolism, sphingolipid metabolism, pyruvate_metabolism, glyoxylate and dicarboxylate metabolism, propanoate metabolism, butanoate metabolism, riboflavin metabolism, nicotinate and nicotinamide metabolism, retinol metabolism, porphyrin and chlorophyll metabolism, nitrogen metabolism, sulfur metabolism, metabolism of xenobiotics by cytochrome P450, drug metabolism cytochrome P450, and drug metabolism other enzymes (**Table S1**).

Identification of PRAD Subtypes by Partitioning Around Medication (PAM) and T-Distributed Stochastic Neighbor Embedding (t-SNE) Analyses

Unsupervised clustering analysis using the PAM algorithm was performed based on the ssGSEA score of each sample using the R package ConsensusClusterPlus function (12). The samples were then divided into three subtypes. The t-SNE analysis of the ssGSEA scores using R package Rtsne identified three clusters. Kaplan–Meier (K–M) survival analysis of the three metabolism-associated subtypes was performed using R package survival.

Stemness Index Calculation and Immune Infiltration Estimation of PRAD Tumors

To evaluate the tumor stemness index, we downloaded the mRNA expression-based stemness index (mRNAsi) calculated by machine learning in previous studies (13). The stem cell gene set was obtained in a previous study (14), and the ssGSEA stemness index (ssGSEAsi) was calculated using the GSVA R package 1.34.0 (10). Tumor purity was calculated using R package ESTIMATE 1.0.13 (15) and then used to correct the stemness index. The immune scores, stromal scores, and ESTIMATION scores calculated using the R package ESTIMATE 1.0.13 (15) were used to evaluate immune cell and stromal cell abundance in the PRAD tumors.

Metabolism-Associated Module Genes Filtered by WGCNA and Functional Enrichment Analysis

After selecting the metabolism-associated genes, we generated an adjacency matrix (AM) and topological overlap matrix (TOM) using the gradient method based on power values ranging from 1 to 20. When the correlation between the average degree of connectivity (k) and p (k) reached 0.88, we obtained the optimal power value and constructed a scale-free topology network. Network connectivity of the genes was measured using a TOM transformed from an AM (16). Modules were calculated using a divided cluster tree (17). Finally, we linked the module eigengenes (MEs) with the subtypes in the current study related to metabolism-associated status (C1, C2, and C3) and for the next analysis, selected the module with the highest correlation based on module-trait correlation coefficients and gene significance (GS) with C1 and C3 (|cor| >0.3). These genes were considered metabolism-associated module genes. To annotate the molecular functions of the genes, Gene Ontology (GO) and KEGG functional enrichment analyses of the metabolism-associated module genes were performed using the clusterProfiler R package 3.42.0 (18).

Biomarker Selection and Risk Model Construction

The correlation between metabolism-associated genes and DFS of PRAD was calculated and analyzed using univariate Cox analysis with R package survival 3.1.8, and candidate biomarkers were screened at p values <0.05. LASSO regression using R package glmnet 3.0.2 (19, 20) was then applied to resolve any multilinear problem that may have existed in the regression analysis, and the biomarkers were filtered. Multivariate Cox was used next to build a risk model and to obtain estimated regression coefficients. Finally, we calculated the risk score for each sample to quantify the prognosis risk of each patient with PRAD. Survival data were analyzed as K–M survival curves *via* R package survival 3.1.8. To evaluate the precision of the risk model and nomogram, time-dependent receiver operating characteristic (ROC) analysis was applied using the R package survival ROC 1.0.3. An area under the ROC curve (AUC) >0.60. indicated the prediction ability of the model was meaningful, and an AUC >0.70 indicated an outstanding predictive value of the model. To investigate the function of risk model genes, we performed GSEA of the TCGA cohort according to the high-risk group and low-risk group divided by the risk score medium value. The correlation between clinical variates and the DFS of PRAD was analyzed and calculated using univariate and multivariate Cox with the R package survival v3.1.8. The nomogram was obtained with the R package survival v3.1.8. The C-index analysis was performed with the R package pec v2019.11.3 (21).

HPA Analysis

Protein levels of six risk model genes expressed in PRAD and normal prostate samples were analyzed using IHC staining data obtained from the HPA database. Four categories of high, medium, low, and not detected were used to evaluate expression levels. These categories included a scoring system based on the proportion of positive-stained cells (>75, 25–75, or <25%) and staining intensity (strong, moderate, weak, or negative).

Statistical Analysis

All statistical analyses were conducted using R software (version 3.6.1). The Mann–Whitney U-test was used to compare two groups with a non-normal distribution of variables. For comparisons of three groups, Kruskal–Wallis tests of variance were used as nonparametric methods. Correlation analysis was

performed using the Person coefficient. All statistical tests were two-sided, and p-values < than 0.05 were considered statistically significant.
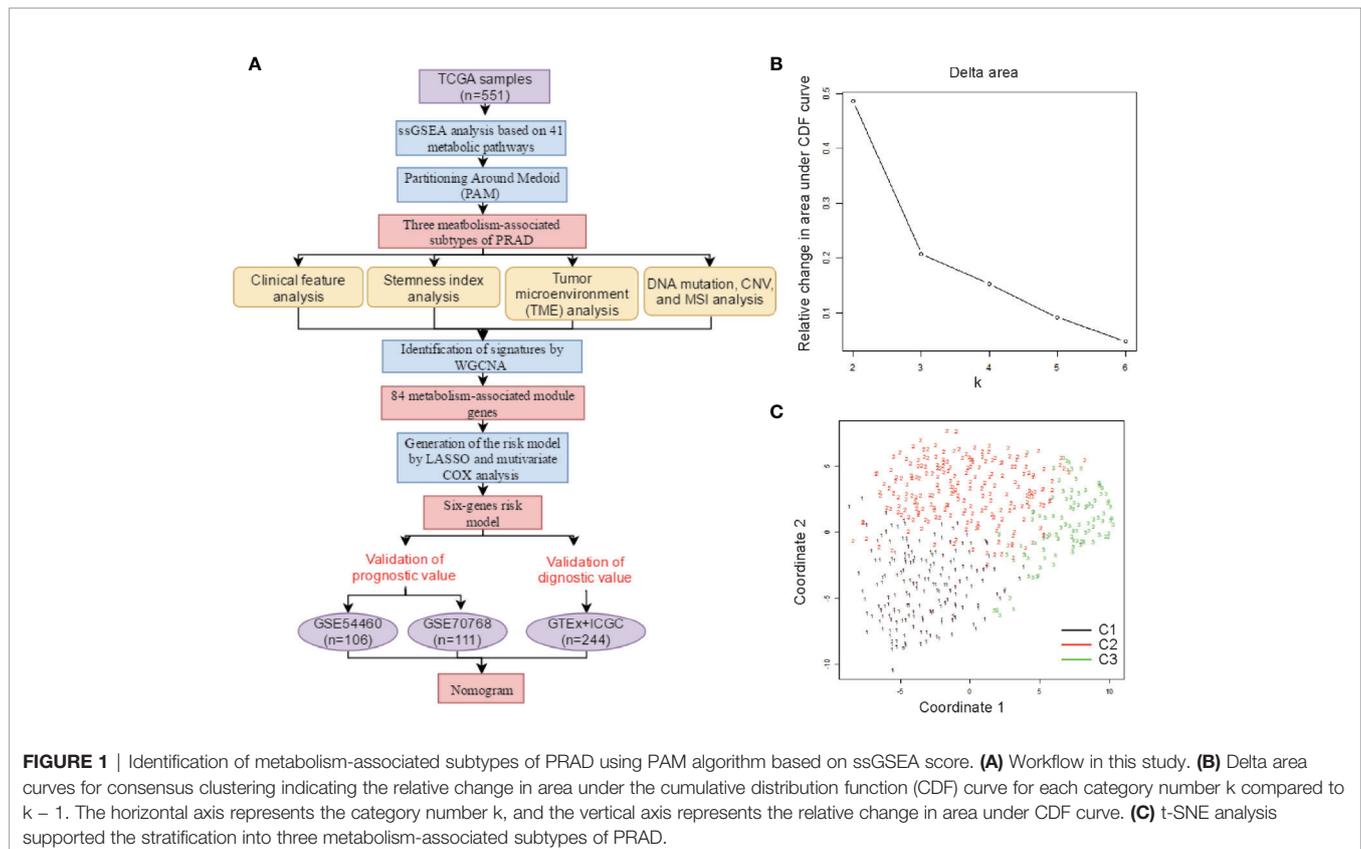
RESULTS

Metabolism-Associated Subtypes Identified by ssGSEA and PAM Analysis

A schematic of our research workflow is shown in **Figure 1A**, and the clinical information regarding the TCGA PRAD cased included in our study is summarized in **Table 1**. To determine the level of activity of the metabolic pathways in each PRAD sample, we calculated the enrichment scores of 41 metabolism-associated gene sets using ssGSEA (**Table S1**). We then performed PAM analysis of ssGSEA scores for 499 PRAD samples and determined the matrix heatmap of the ssGSEA scores retained sharp and clear sides when k = 3, which indicated there were three different metabolism-associated clusters, C1, C2, and C3 (**Figure 1B** and **Figures S1A–F**). To verify the subtype distribution, we performed t-SNE to dimensionally reduce the ssGSEA scores and found the subclass assignment was approximately accordant with the t-SNE coordinates designation (**Figure 1C**).

To explore the characterization of each subtype, we described the clustering hot map of the metabolic pathway scores (**Figure 2A**).

Compared to that of cluster C1, C2 had higher enrichment scores in most metabolic pathways, indicating that tumors from C2 exhibited higher metabolic activity than that of tumors from C1. Concomitantly, the highest specific metabolic pathways scores were observed for cluster C3 and included retinol metabolism, metabolism of xenobiotics by cytochrome P450, drug metabolism cytochrome_P450, drug metabolism other enzymes, starch and sucrose metabolism, ascorbate and aldarate metabolism, and porphyrin and chlorophyll metabolism. The other pathway scores of C3 were higher compared to those of C1, but lower than those of C2. This indicated that tumors from C3 might have had a medium metabolic status at levels between those from C1 and C2 and, at the same time, exhibited some unique metabolic characteristics.

Clinical analysis comparing the different subtypes revealed patients from C1 and C3 had higher primary tumor (T) stage and Gleason scores than patients from C3, but the age and regional lymph nodes (N) stage differences of the patients among these subtypes were not significant (**Figures 2B, C** and **Figures S2A, B**). Prostate-specific antigen (PSA) is the most common index used in the diagnosis and prediction of prognosis for PCa (22). Differences in PSA levels among the three subtypes indicated the subtypes were independent of PSA without any detectable connection (**Figure S2C**). We then performed a K–M survival analysis of patients with PARD from each subtype. The results suggested there were considerable differences in DFS among the three subtypes (p < 0.05; **Figure 2D**). The patients from C1 had the shortest DFS compared to those from the other subtypes.

FIGURE 1 | Identification of metabolism-associated subtypes of PRAD using PAM algorithm based on ssGSEA score. **(A)** Workflow in this study. **(B)** Delta area curves for consensus clustering indicating the relative change in area under the cumulative distribution function (CDF) curve for each category number k compared to k – 1. The horizontal axis represents the category number k, and the vertical axis represents the relative change in area under CDF curve. **(C)** t-SNE analysis supported the stratification into three metabolism-associated subtypes of PRAD.

TABLE 1 | Clinical information from the 545 PCa patient of TCGA.

Clinical parameters	Variable	N (total = 545)	Percentages (%)
Age (years)	<=60	242	44.40%
	>60	303	55.60%
T/N grade	T2	188	34.50%
	T3	295	54.13%
	T4	10	1.83%
	unknow (T stage)	52	9.54%
	N0	348	63.85%
	N1	79	14.50%
	unknow (N stage)	118	21.65%
Gleason score	6	48	8.81%
	7	285	52.29%
	8	66	12.11%
	9 & 10	146	26.78%
Survival status	Dead	10	1.83%
	Alive	482	88.44%
	unknown	53	9.72%

This result indicated that the metabolism-associated subtypes would be associated with different prognoses, and tumors from the different subtypes exhibited considerable differences in their metabolic status.

Correlation of PRAD Subtypes With Cancer Stem Cell Characteristics

In previous studies, cancer stem cell characteristics have been shown to represent the capability of tumor proliferation and are associated with the development and progression of PRAD (23, 24). To determine the heterogeneity of the current study subtypes, we compared the stemness index of each subtype that was calculated using one-class logistic regression (OCLR) machine learning and ssGSEA. We initially obtained two stemness indices, mRNAsi, and ssGSEAsi (**Table S2**).

Differential analysis of mRNAsi indicated there were significant differences among the three subtypes (p <0.05; **Figure 3A**). C1 had the highest stemness index, whereas C2 had the lowest. Moreover, ssGSEAsi analysis indicated that C3 had the highest stemness index (p <0.05; **Figure S3A**). To compensate for the impact of tumor purity on the stemness index, we recalculated the indices using corrected mRNAsi and ssGSEAsi values by dividing them by their respective tumor purity values and then re-performing the differential analysis. The results for the two corrected stemness indices were in approximate accordance with the original results (**Figure 3B** and **Figure S3B**). This suggested the tumors from clusters C1

FIGURE 2 | Association between clinical characteristics and the metabolism-associated subtypes. **(A)** Heatmap of the ssGSEA score calculated by metabolic pathways gene sets and specific metabolic pathways of C3 in the red frame. Gleason score **(B)** and Primary Tumor (T) stage **(C)** for each metabolism-associated subtype in the TCGA cohort. The P values are labeled above each boxplot with asterisks (ns represents no significance, *P < 0.05, **P < 0.01, ***P < 0.001). **(D)** Survival curves for each metabolism-associated subtype in the TCGA cohort. The horizontal axis represents survival time (year), and the vertical axis represents the probability of survival. The log-rank test was used to assess the statistical significance of the differences between the three subtypes.

and C3 had a stronger capacity for invasion, proliferation, and self-renewal compared to that for those from C2.

Because C3 demonstrated the highest ssGSEAsi and COssGSEAsi and had specific metabolic pathways, we performed a correlation analysis between COssGSEAsi/ssGSEAsi and the specific metabolic pathways. The results indicated the starch and sucrose metabolism and porphyrin and chlorophyll metabolism pathways were highly correlated with the PRAD stemness index (cor >0.3; **Figure S3C**).

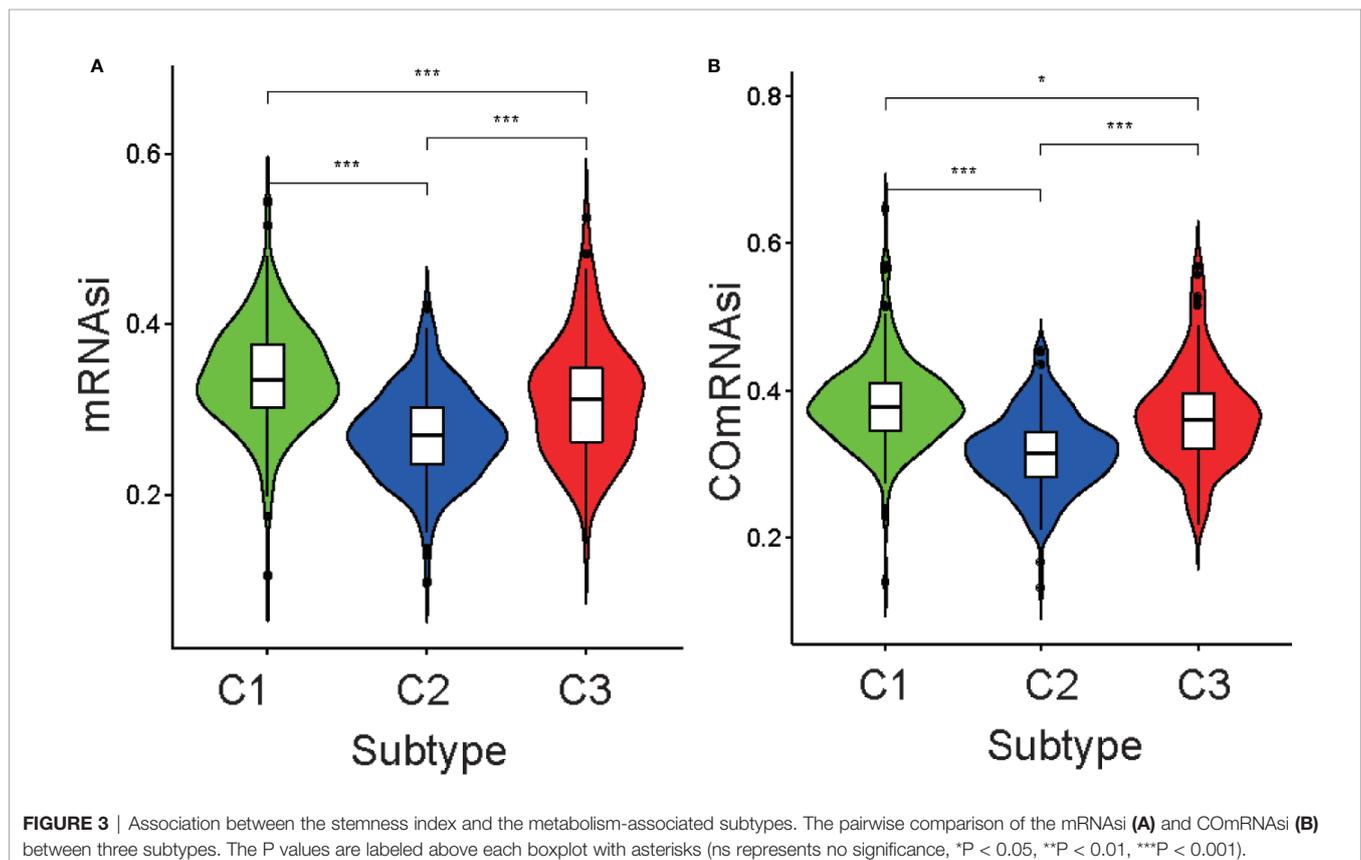## Relationship Between PRAD Subtypes and TME

To further investigate PRAD tumor heterogeneity, we compared the TME among the metabolism-associated subtypes. In previous studies, TME compounded by both stromal and immune cells played a crucial role in the occurrence and progression of PRAD (25, 26). Moreover, TME may reflect a tumor's sensitivity to immunotherapies (27, 28). Accordingly, we obtained stromal scores and immune scores for the PRAD tumors in the current study using the ESTIMATE algorithm and then performed a differential analysis of the three subtypes. The results showed that C1 tumors had lower stromal and immune scores compared to those C2 and C3 tumors (**Figures 4A–C**). This suggested that tumor tissue from C1 had higher tumor purity and lower immune infiltration compared to tumor tissues from C2 and C3. To further investigate the differences in stromal cells among the three clusters, we calculated ssGSEA
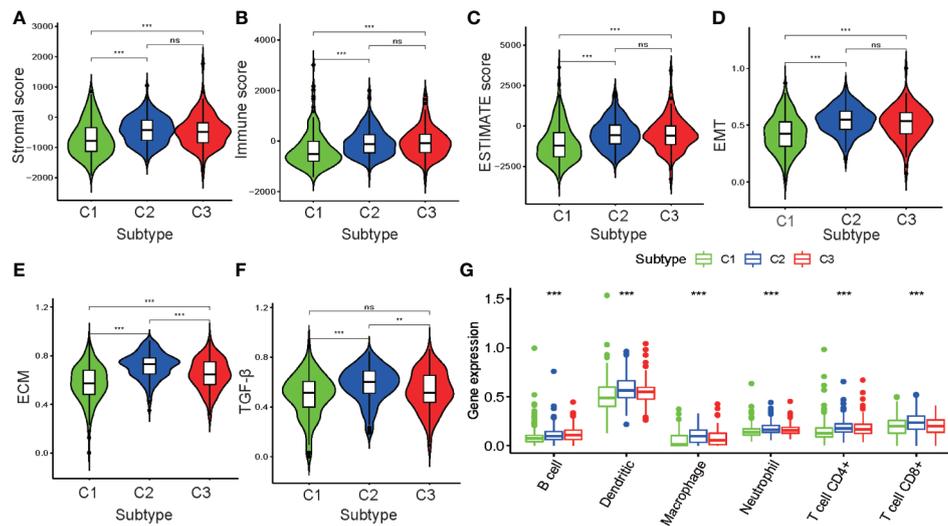
scores for epithelial-mesenchymal transition (EMT), extracellular matrix (ECM), and transforming growth factor-beta (TGF-β) using the corresponding gene sets downloaded from the Molecular Signatures Database (**Table S2**). Differential analysis of these ssGSEA scores suggested C1 tumors had the lowest scores for all three gene sets, which was consistent with the results from the comparison of the stromal scores for the three subtypes (**Figures 4D–F**).

Because of the significant difference in immune scores between PRAD subtypes, we explored immune infiltration to identify their respective immunologic landscapes. The abundance of six immune-related cell types, B cell, dendritic, macrophage, neutrophil, CD4+ T cell, and CD8+ T cell, was download from TIMER2.0. Significant differences for all six immune cell types were verified among the cluster subtypes, with the tumors from C2 having the highest abundance of all immune cells, except B cells. The cluster with the highest abundance of B cells was C3, whereas C1 tumors had the lowest abundance of all the immune cell types evaluated (**Figure 4G**). Our results indicated that the metabolism-associated subtypes of PRAD exhibited remarkably distinct characteristics with respect to immune infiltration.

PRAD Subtype Relationship With DNA Mutations, CNV, and MSI

To determine the reason for cluster subtype heterogeneity, we investigated whether differences existed among the three

FIGURE 3 | Association between the stemness index and the metabolism-associated subtypes. The pairwise comparison of the mRNAsi **(A)** and COmRNAsi **(B)** between three subtypes. The P values are labeled above each boxplot with asterisks (ns represents no significance, *P < 0.05, **P < 0.01, ***P < 0.001).

FIGURE 4 | Association between the tumor microenvironment and the metabolism-associated subtypes. The pairwise comparison of the stromal score **(A)**, immune score **(B)**, ESTIMATE score **(C)**, ssGSEA score of EMT **(D)**, ssGSEA score of ECM **(E)**, and ssGSEA score of TGF-β **(F)** between three subtypes. **(G)** The differential analysis of the abundance of immune cells between three subtypes. The P values are labeled above each boxplot with asterisks (ns represents no significance, *P < 0.05, **P < 0.01, ***P < 0.001).

subtypes in DNA mutation burdens and patterns of somatic mutation rates. By displaying the 15 genes determined to have the highest frequency of DNA mutations in PRAD in a waterfall plot, we observed remarkably different landscapes for each of the PRAD subtype (**Figure 5A**). Mutation of *TP53* was the most frequent DNA mutation in cluster C1 and mutation of *SPOP* was the most frequent in cluster C2. These results indicated that the *TP53* mutation was a characteristic mutation of C1 tumors, and the *SPOP* mutation was a characteristic mutation of C2 tumors. The C3 cluster had high mutation rates of both *SPOP* and *TP53*. This may explain why tumors from C3 exhibited a status between those of C1 and C2, regardless of the stemness index or TME analysis.

Tumor mutation burden (TMB) is considered to reflect the sensitivity of tumors to targeted drug therapies (29). To further investigate the features of DNA mutations and clinical treatment options for PARD subtypes, we compared the differences in the number of DNA mutations among the subtypes. We found that tumors of subtypes C1 and C3 had higher mutation counts than those of subtype C2 (**Figure 5B**). This further indicated that tumors from C1 and C3 exhibited greater heterogeneity compared to those from C2.

CNV occurring upstream of genes regulates gene expression and influence tumor occurrence and development (30). To further explore whether this DNA element may lead to increased heterogeneity among the metabolism-associated subtypes, we downloaded a list of metabolism-relevant genes (31) and analyzed the number of amplifications and deletions regarding the CNV in these genes. We found that the number of CNV amplifications and deletions was highest in subset C1, followed by that in subset C3, with the fewest being observed in
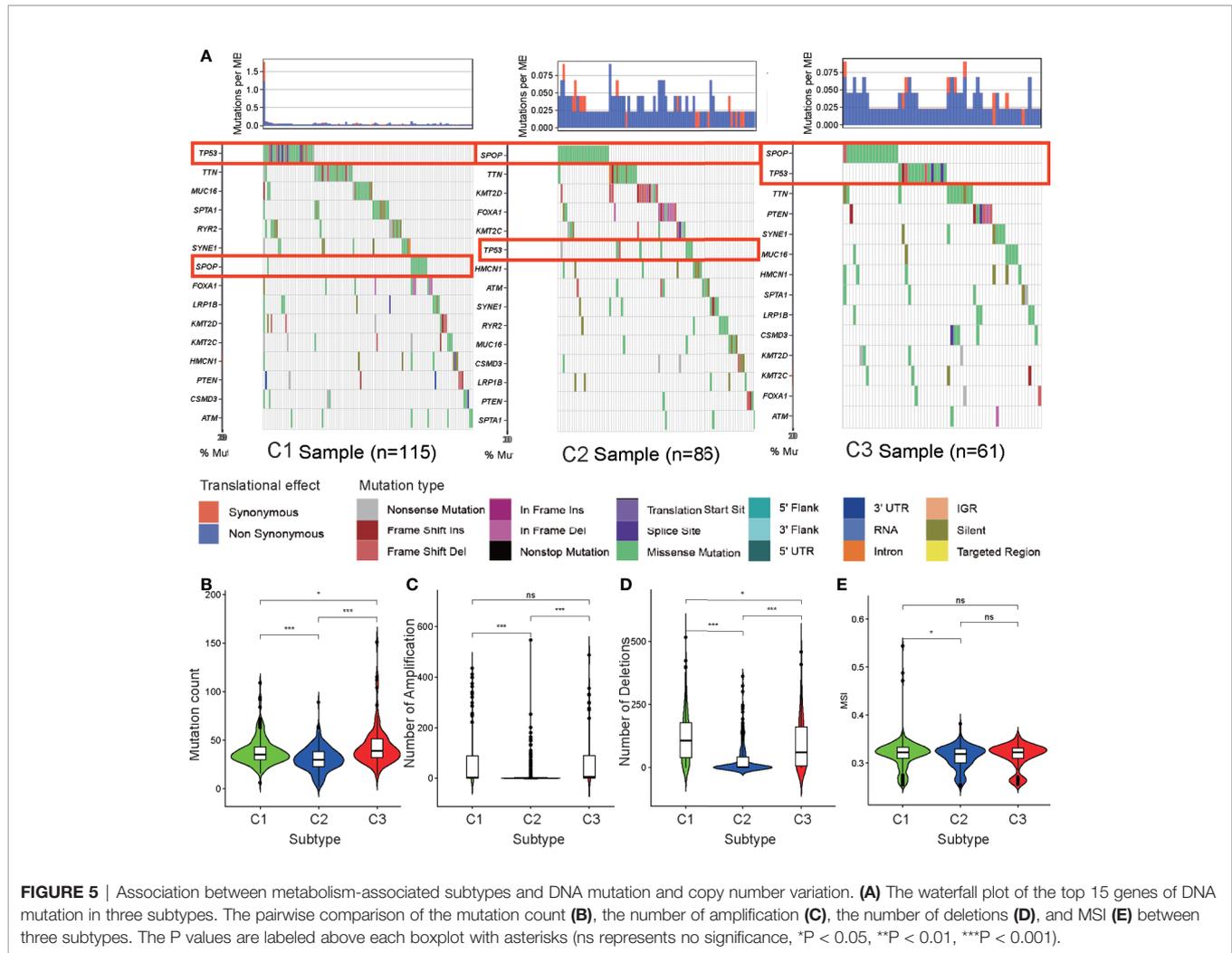
subset C2 (**Figures 5C, D**). These results suggest that CNV results in significant heterogeneity among the three subtypes.

Previous studies have shown that MSI is a crucial indicator of genome instability and is associated with many genetic diseases (32). In our studies, we obtained level data from MSI of each PRAD sample calculated in a previous study (33) and performed differential analysis. The results indicate that C1 has a higher level of MSI than C2 (p <0.05) (**Figure 5E**), suggesting that MSI may be the resource of tumor heterogeneity in C1.

Identification of Metabolism-Associated Signatures

We selected 2,029 metabolism-associated genes among the TCGA PRAD cohort samples and constructed a co-expression network through co-expression analysis (31). Average linkage hierarchical clustering identified five modules. To realize the scale-free co-expression network, a power of β = 3 was used (**Figure S4A**). We then adopted the dynamic hybrid tree cut method to combine highly similar modules using a cutoff value = 0.25 and module size = 50 (**Figure S4B**). Although we failed to identify a module associated with the C3 subtype, the green and blue modules showed a strong association with C1 and C2 subtypes (cor > 0.3 or <−0.3; **Figures 6A** and **S4C**). Ultimately, 489 associated genes were identified, including 388 genes in the blue module and 101 genes in the green module. Of the 489 genes, 84 (cor of GS with C1 and C2 >0.3 or <−0.3) were determined as metabolism-associated module genes (**Figure 6B** and **Table S3**).

To determine the biochemical functions of the metabolism-relevant module genes, we performed GO and KEGG function enrichment analyses. Some metabolic pathways, such as alcohol

FIGURE 5 | Association between metabolism-associated subtypes and DNA mutation and copy number variation. **(A)** The waterfall plot of the top 15 genes of DNA mutation in three subtypes. The pairwise comparison of the mutation count **(B)**, the number of amplification **(C)**, the number of deletions **(D)**, and MSI **(E)** between three subtypes. The P values are labeled above each boxplot with asterisks (ns represents no significance, *P < 0.05, **P < 0.01, ***P < 0.001).

metabolic process, sulfur compound metabolic process, cellular modified amino acid metabolic process, phenylalanine metabolism, drug metabolism cytochrome P450, and glutathione metabolism were significantly enriched by these genes (**Figures 6C, D**; **Table S4**). This further confirmed that the functions of the selected genes were closely associated with PRAD metabolism.
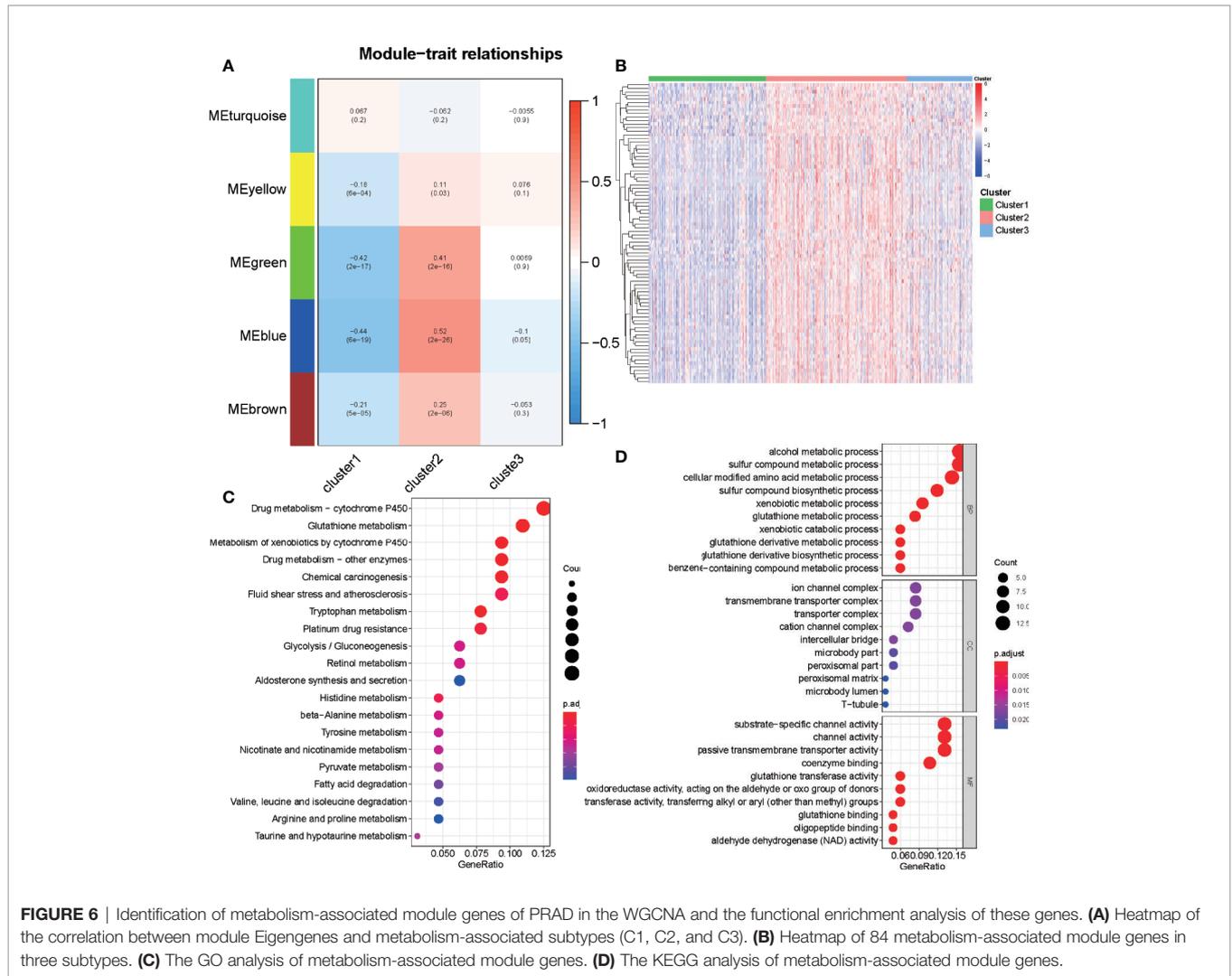
Metabolism-Associated Risk Model Development and Validation

Patients with DFS <20 d and those without available DFS information were excluded. A total of 489 patients in the TCGA database were included in the training cohort of this study (**Table 2**). We performed a univariate Cox regression analysis of the 84 previously selected variables to identify potential optimal prognostic targets. A total of 23 genes that met the prognostic criteria were identified (p <0.05; **Table S5**). To avoid overfitting of the model. The prognostic biomarkers that highly correlated with one another were first removed using LASSO regression, resulting in six candidate prognostic genes (**Figures S5A, B**). These six genes were then analyzed using the

multivariate Cox proportional hazards regression method. Finally, metabolism-associated module genes related to DFS of PRAD were identified, including *CACNG4*, *SLC2A4*, *EPHX2*, *CA14*, *NUDT7*, and *ADH5* (**Table 3**). The formula used for calculating the risk score was as follows:

$$
\begin{aligned}
\text{Risk score} = &(-0.083 \times \text{FPKM of } CACNG4) \\
&+ (-0.0980 \times \text{FPKM of } SLC2A4) \\
&+ (-0.0161 \times \text{FPKM of } EPHX2) \\
&+ (-0.2182 \times \text{FPKM of } CA14) \\
&+ (-0.2055 \times \text{FPKM of } NUDT7) \\
&+ (-0.0213 \times \text{FPKM of } ADH5)
\end{aligned}
$$

To verify the robustness of the risk model, two external cohorts available in the GEO repository, datasets GSE54460 and GSE70768, were obtained and used as validation cohorts. Each cohort was separated into two groups according to the median value of each risk score. To evaluate the differences in prognosis between the high-risk and low-risk groups, a K–M

FIGURE 6 | Identification of metabolism-associated module genes of PRAD in the WGCNA and the functional enrichment analysis of these genes. **(A)** Heatmap of the correlation between module Eigengenes and metabolism-associated subtypes (C1, C2, and C3). **(B)** Heatmap of 84 metabolism-associated module genes in three subtypes. **(C)** The GO analysis of metabolism-associated module genes. **(D)** The KEGG analysis of metabolism-associated module genes.

TABLE 2 | Grouping of PCa patients for survival analysis.

Clinical parameter	Variable	TCGA	GSE 54460	GSE 70768
Recurrence or no	Recurrence	91 (18.61%)	51 (48.11%)	19 (17.11%)
	No recurrence	398 (81.39%)	55 (51.89%)	92 (82.88%)

survival curve was constructed based on the log-rank test. Patients in the high-risk group of the TCGA cohort GSE54460 dataset exhibited poorer outcomes compared to those in the low-risk group (p < 0.05; **Figures 7A, B**). As for the GSE70768 cohort, the arrangement characteristic of microarray data differed from RNA-seq data. Therefore, we divided the cohort according to most cutoff value, which was calculated using X-tile, and found there was a significantly different prognosis between the high-risk and low-risk groups (**Figure 7C**). We used a time-dependent ROC curve to investigate the predictive accuracy of our model and determined the AUC of the prognostic model using the TCGA training cohort was 0.769 at one year, 0.702 at three years, and 0.705 at five years (**Figure 7D**). For the testing cohorts, the AUC of the prognostic model for the GSE54460
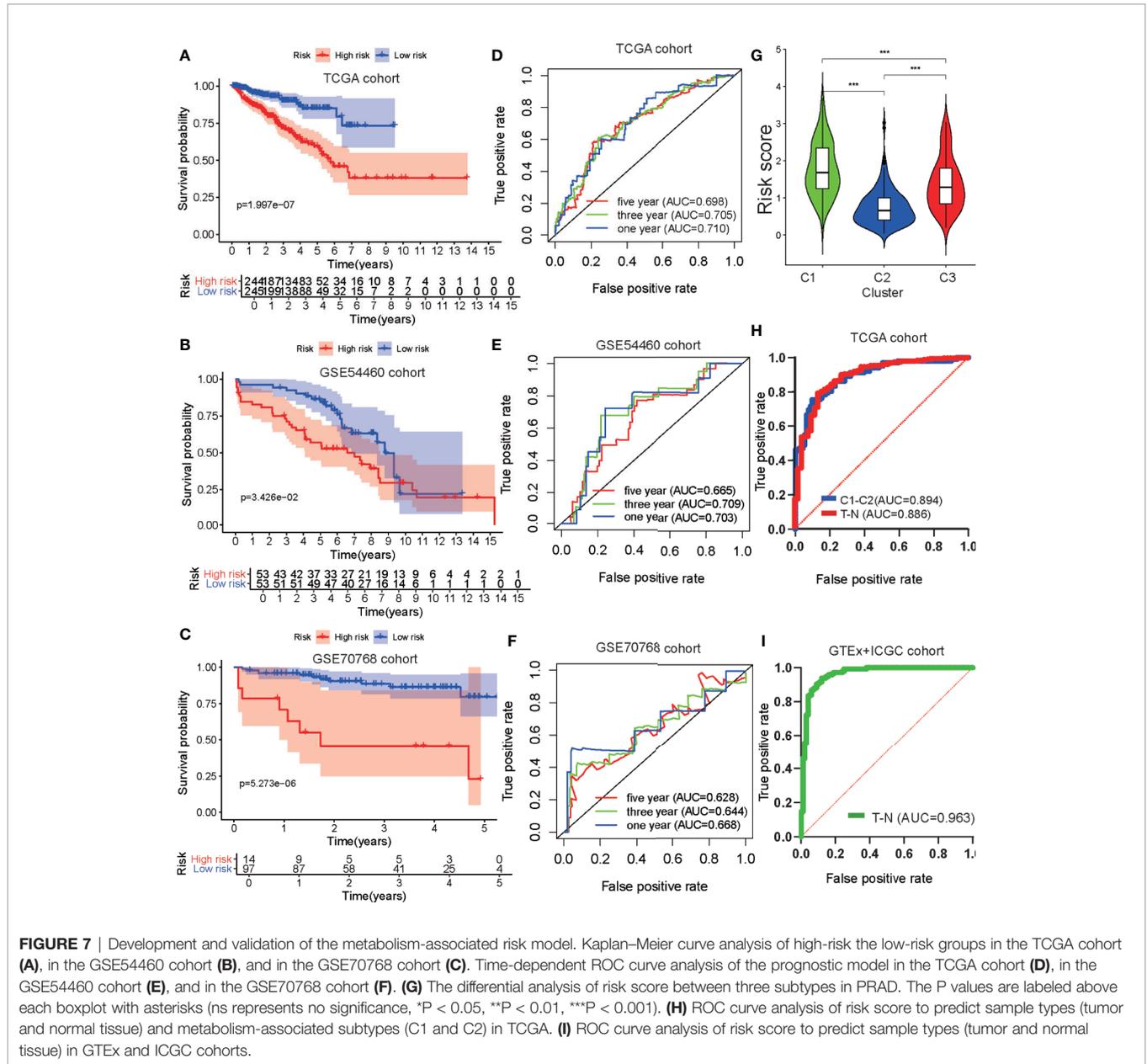
cohort was 0.703 at one year, 0.709 at three years, and 0.665 at five years and 0.668 at one year, 0.644 at three years, and 0.628 at five years for the GSE70768 cohort (**Figures 7E, F**).

Identification of Risk Model Biomarkers Biological Functions and Construction of Nomogram

GSEA analysis was performed to explore the biological functions of the risk model genes. The results indicated the genes had a significant relationship with cell cycle, DNA replication, homologous recombination, RNA degradation, and spliceosome pathways in the high-risk group and with beta alanine metabolism, dilated cardiomyopathy, drug metabolism cytochrome P450, metabolism of xenobiotics by cytochrome_P450, and vascular

TABLE 3 | Risk genes in the prognostic risk model.

GENE	Coef	HR	HR.95L	HR.95H	P value
CACNG4	−0.00837	0.991664	0.96811	1.015792	0.49494
SLC2A4	−0.09798	0.906668	0.768444	1.069754	0.245648
EPHX2	−0.01612	0.984005	0.960761	1.007812	0.18618
CA14	−0.21817	0.803992	0.352487	1.833836	0.604063
NUDT7	−0.20554	0.814211	0.501774	1.321193	0.405298
ADH5	−0.02131	0.978912	0.95889	0.999352	0.043232

FIGURE 7 | Development and validation of the metabolism-associated risk model. Kaplan–Meier curve analysis of high-risk the low-risk groups in the TCGA cohort **(A)**, in the GSE54460 cohort **(B)**, and in the GSE70768 cohort **(C)**. Time-dependent ROC curve analysis of the prognostic model in the TCGA cohort **(D)**, in the GSE54460 cohort **(E)**, and in the GSE70768 cohort **(F)**. **(G)** The differential analysis of risk score between three subtypes in PRAD. The P values are labeled above each boxplot with asterisks (ns represents no significance, *P < 0.05, **P < 0.01, ***P < 0.001). **(H)** ROC curve analysis of risk score to predict sample types (tumor and normal tissue) and metabolism-associated subtypes (C1 and C2) in TCGA. **(I)** ROC curve analysis of risk score to predict sample types (tumor and normal tissue) in GTEx and ICGC cohorts.

smooth muscle contraction pathways in the low-risk group (**Figures S5C, D**; **Table S6**). This suggested our risk model genes influenced these pathways and that these pathways may impact PRAD DFS.

We then explored the relationship between clinical/ metabolism-associated subtypes and risk scores and found there was a close relationship between risk score and age/ Gleason score/T stage/N stage/metabolism-associated subtypes

(**Figures S5E–H** and **Figure 7G**). This suggested that our risk model had predictive value, not only in PRAD DFS, but also in tumor size, lymphatic node metastasis, and metabolism-associated subtype. Meanwhile, when we used ROC to determine whether our risk model could predict sample type (tumor vs. normal tissue) and metabolism-associated subtypes (C1 vs. C2), we observed that the AUC values of the risk model were 0.886 and 0.894, respectively, in TCGA cohort (**Figure 7H**). To determine the diagnostic value of our model, we used ROC to evaluate samples from the external group (GTEx and ICGC) and found that there was an outstanding predictor value for PCa diagnosis with our model (AUC = 0.963) (**Figure 7I** and **Table 4**).

Finally, we selected clinical variates with independent prognostic value to obtain a nomogram through univariate and multivariate Cox analyses (**Figure S5I** and **Table S7**). ROC analysis and C-index calculation assessed the clinical meaning of the nomogram and suggested that the clinical nomogram had a better net benefit than clinical variate or risk score only models (**Figures S5J, K**).

To determine the diagnostic and prognostic value of a single risk model gene, we performed differential analyses of six risk model genes between different types, Gleason score, T- and N stage of samples, and K–M analyses of each risk model gene in the TCGA cohort. The findings indicated that six risk model genes have differing expression levels between normal and tumor tissue, and different Gleason score samples. Simultaneously, there was a significant difference in prognosis between PARD patients with high-risk and low-risk model gene expression (**Figures 8A–D** and **Figures S6A–F**).

To confirm whether there was a differential abundance of proteins associated with the selected genes between normal prostate tissues and tumor tissues of patients with PRAD, we downloaded IHC micrographs from the HPA database. Three of the six risk model genes were found to exhibit differential staining between normal prostate tissue and PRAD tissue, those being *EPHX2*, *NUDT7*, and *ADH5* (**Figures S7A–C**). The results suggested that the expression of these proteins was decreased in PRAD tissues. This was in accordance with the differential analysis of expression for the six risk model genes of the TCGA cohort and further indicated these genes might play crucial roles in the occurrence and development of PRAD.

Drug Sensibility Analysis With Metabolism-Associated Subtypes and Risk Model

Anti-androgen treatment is the first non-surgical treatment for PRAD (34). It has been shown that the level of the androgen receptor (AR) gene expression in tumor tissue is closely related to anti-androgen treatment sensitivity (35). Therefore, we also compared the expression levels of AR in the three subtypes and

TABLE 4 | Grouping of PCa patients for diagnostic analysis.

| Clinical parameter	Variable	TCGA	GTEx + ICGC
Normal or tumor tissue	Normal	52 (10.60%)	100 (40.98%)
	Tumor	499 (89.40%)	144 (59.02%)
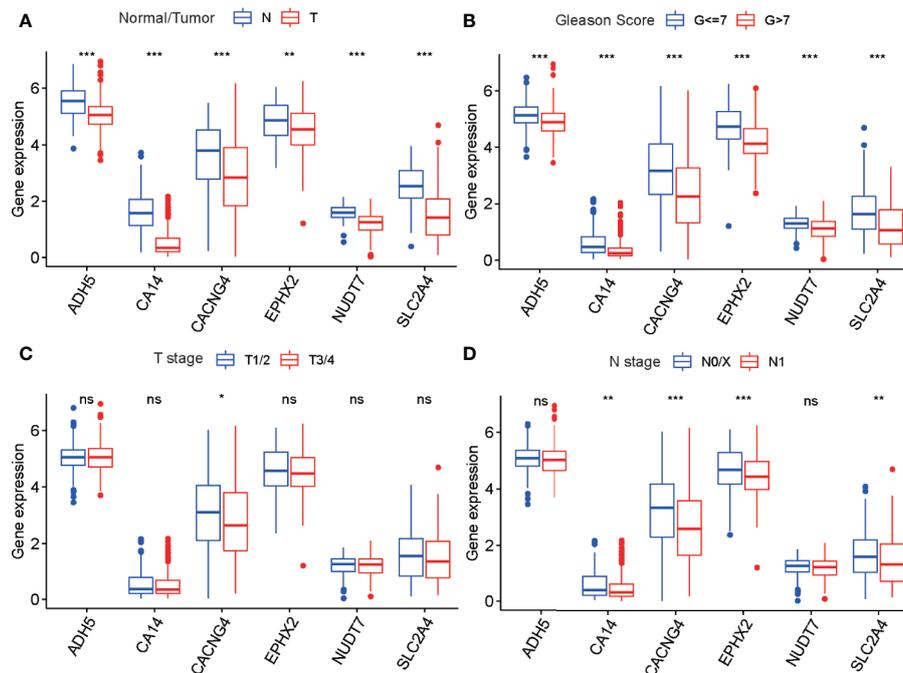
FIGURE 8 | The differential analysis of six risk model genes between PRAD tissue and normal prostate tissue **(A)**, different Gleason score PRAD samples **(B)**, different Primary Tumor (T) stage samples **(C)**, and different Lymph Nodes (N) stage samples **(D)** in TCGA cohort. The P values are labeled above each boxplot with asterisks (ns represents no significance, *P < 0.05, **P < 0.01, ***P < 0.001).

found that tumors from C1 and C3 had higher AR expression levels than those from C2 (**Figure 9A**). This indicates that patients from C1 and C3 may be more sensitive to anti-androgen treatment that those from C2, and that there existed considerable tumor heterogeneity among the subtypes.

For castration-resistant prostate cancer, immunotherapy and chemotherapy are the preferred treatments. To explore the sensitivity of each cluster subtype to immunotherapies, we collected data for 14 immune checkpoints and compared the gene expression levels of these proteins. Significant differences in gene expression of the checkpoint proteins among the three

subtypes were observed, except for CTLA4 and IL1A. Cluster C1 exhibited lower expression of most the checkpoint genes relative to that cluster C2 and C3 (**Figure 9B**), which is in accordance with the immune infiltration status described above. Therefore, we were able to determine that tumors of the C1 subtype demonstrated lower immune infiltration and may, therefore, garner fewer benefits from treatment with immunotherapies.

Then, to evaluate the association between drug sensitivity and the metabolism-associated risk model, correlation analysis was performed using half-inhibitory concentration (IC50) of chemotherapeutic drugs, risk score, and gene expression data
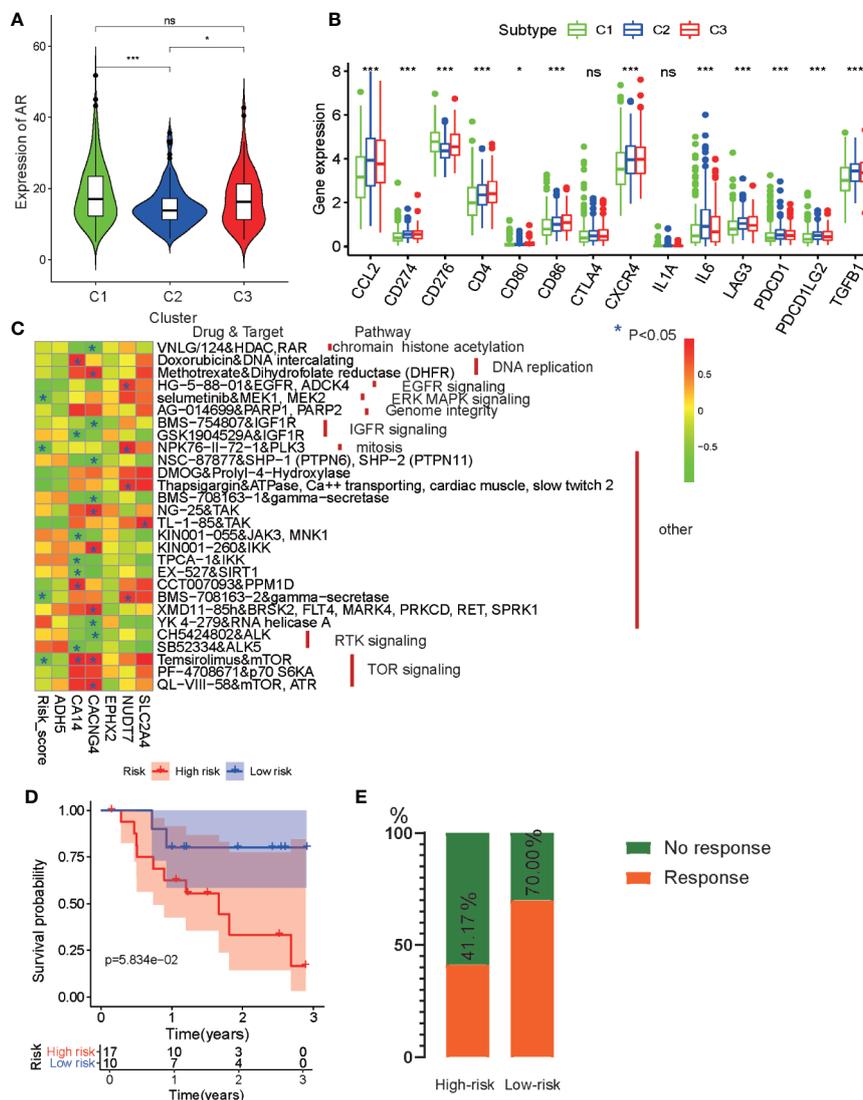
FIGURE 9 | Drug sensibility analysis with metabolism-associated subtypes and risk model. **(A)** The pairwise comparison of the androgen receptor (AR) expression between three subtypes. **(B)** The differential analysis of the expression of 14 checkpoints between three subtypes. The P values are labeled above each boxplot with asterisks (ns represents no significance, *P < 0.05, **P < 0.01, ***P < 0.001). **(C)** Heatmap for correlation between drug sensitivity and expression levels of six risk model genes. *p < 0.05. **(D)** Survival analyses for low-risk and high-risk scores patient groups in the anti-PD1 immunotherapy cohort using Kaplan–Meier curves. **(E)** The proportion of patients with response to PD-1 blockade immunotherapy in low-risk or high-risk scores groups.

from six PCa cell lines obtained from the GDSC database. As a result, four anti-tumor drugs (selumetinib, NPK76-11-72-1, BMS-708163, and temsirolimus) were negatively correlated with risk score (**Figure 9C** and **Table S8**). Among the six genes, CA14, CACNG4, and NUDT7 had associations with more than four drugs, indicating that these genes and risk scores may guide chemotherapy in drug choice.

Finally, immunotherapies represented by PD-L1 and PD-1 blockades have undoubtedly emerged as a breakthrough in cancer therapy (36, 37). We explored whether our risk model could predict patient responses to immune checkpoint blockade therapy based on the anti-PD-1 cohort (GSE78220). First, we performed K-M survival analysis and found that melanoma patients with high-risk scores had shorter OS than patients with low-risk scores, although no significant difference was observed (p = 0.058) (**Figure 9D**). The proportion of beneficiaries with anti-PD-1 treatment in the high-risk cohort (41.17%) was lower than that of the low-risk cohort (70%) (**Figure 9E**). These results implied that patients with low-risk scores would get more benefits from immunotherapy than patients with high-risk scores.

DISCUSSION

With the development of RNA-seq technology, many classical analyses based on gene expression data have been reported for most cancers (6, 38–40). However, few cluster studies have been performed on PCa to explore the tumor metabolic characteristics. Thus, we identified in our current work a metabolism-associated PRAD classification based on ssGSEA and 41 metabolic pathway gene sets. The PRAD cases included in the study were divided into three subtypes. The metabolic features, clinical characteristics, prognosis, TME, stemness index, DNA mutation, CNV, and MSI were then investigated in the different subtypes. Subtype C1 exhibited low metabolic levels and was similar to high-grade PCa with high tumor purity and low immune infiltration. Furthermore, patients from C1 had worst prognosis and the shortest DFS among the patients with PCa. In comparison, patients from C2 displayed high metabolic levels in most pathways and were similar to low-grade PCa with low tumor purity, low stemness index, and high immune cell infiltration. Patients with tumors from C2 had the best prognosis and longest DFS among the patients with PCa. Patients from C3 represented a medium state between the findings for those from C1 and C2 and demonstrated similar medium-grade PCa. Specifically, C3 had highly metabolic pathways activity and the highest stemness index. Accordingly, we believe there may be some connections between starch/sucrose/porphyrin/chlorophyll metabolism and a high stemness index.

Cancer stem cells play important roles in therapeutic responses and the progression of cancer (41). To further explore the reason C1 subset has the worst prognosis, we continued investigating the stem index between three subtypes. C1 and C3 had higher stemness indices, which indicates more

malignant tumors from these subsets compared to those from C2. This may partly explain why patients with PCa from subsets C1 and C3 presented with shorter DFS and worse prognosis compared to that of patients from subset C2. These results indicate a remarkable tumor heterogeneity among PRAD metabolism-associated subtypes. However, the reason for C3 tumors having the highest ssGSEAsi is unclear. In previous studies, starch and sucrose metabolism is associated with the progression of colon cancer (42). Therefore, we believe that specific metabolic pathways, such as those for starch/sucrose/porphyrin/chlorophyll metabolism, have crucial roles in tumor progression. This hypothesis needs to be tested using *in vitro* experiments and single-cell sequencing.

TME is a remarkable factor impacting the occurrence and development of PCa. Many cancer-promoting factors play a role in the EMT pathway (43). For instance, the expression of PDL1 can affect the prognosis of adrenocortical carcinoma (44). Regardless of stromal or immune cells, the C2 and C3 subtypes displayed more characteristics of TME. Stromal scores indicated that C2 and C3 had greater stromal cell content than C1. ECM functions as cell scaffoldings and can induce EMT in stromal cells, with the TGF-β pathway having a strong connection with this process (45). To further explore the features of stromal cells in PCa, ECM, EMT, and TGF-β, ssGSEA scores were calculated for each gene set. Differential analysis suggests that C2 had the most significant activity in this process. EMT often has a close relationship with cell cancer and poor prognosis (46–48); however, C2 had the best prognosis among the three subtypes. Therefore, we believe that EMT does not have an obvious cancer-promoting function in PCa. As for the increase of EMT in C2, we believe this phenomenon was the result of increased stromal cell content in C2. The immune system is the most important anti-cancer system in the body (49). In a previous study, immune cells were found to be strongly lethal in fighting tumor cells. Natural killer cells can kill lung cancer tumor cells and are regulated by TME (50). Additional support that the immune system is important is demonstrated by T cells being able to be used in clinical settings for the treatment of cancer (51). Herein, our study shows that the levels of all immune cells (B cells, dendritic cells, macrophages, neutrophils, CD4+ T cells, and CD8+ T cells) were increased in C2, indicating that tumors from C2 were in a state of immune activation. We believe this is the reason C2 had the best prognosis among the three subtypes.

TMB is presumed to have a close relationship with tumor heterogeneity (52). *TP53* is the most prominent gene in pan-cancer investigations. For instance, *TP53* mutations lead to high-grade cancer and tumor heterogeneity of ovarian granulosa (53). At the same time, mutations in *TP53* were shown to be strongly associated with the occurrence and progression of PCa (54). *SPOP* mutations have also been considered for their impact on castration sensitivity in PCa (55). Gene mutation spectra were significantly different among the three metabolism-associated subtypes in our current study. For instance, in C1, the *TP53* mutation rate was higher than that of others, whereas C2 had a high mutation rate of *SPOP*. This indicates that the tumor features of C1 and C2 partly result from *TP53* and *SPOP*

mutations, respectively. C3 had high mutation rates of both *TP53* and *SPOP*, which further supported that C3 exhibited an intermediate state between C1 and C2. Meanwhile, many studies have investigated CNV in PCa, with the results indicating that CNV can affect tumor features and heterogeneity (56). MSI also has been regarded as a vital factor in DNA mismatches and can improve tumor heterogeneity in many types of cancer. Herein, we show that there is a significant difference in CNV of metabolic genes and MSI levels among the three subtypes. Our current work revealed that there was a significant difference in CNV of metabolic genes among the three subtypes. C1 had the highest number of amplifications and deletions. This indicates that the *TP53* mutation, *SPOP* mutation, and increased CNV and MSI were key factors contributing to the tumor heterogeneity observed among the subtypes.

A general opinion regarding tumor cells with high activity levels of one or more specific metabolic pathways is that they have a stronger capacity for invasion, proliferation, and self-renewal compared to cells with low metabolic activity. For instance, higher sulfur amino acid metabolic levels in liver cancer can accelerate the EMT process and cancer cell migration (57). Aldehyde oxidase 1 decreases the metabolic level and displays tumor inhibition activity in bladder cancer (58), whereas long intergenic non-coding RNA-nucleotide metabolism regulator upregulates nucleotide metabolism and increase the proliferation of tumor cells (59). Our results partly contradict these views in that we found tumors from C1 with the lowest metabolic activity level had the worst prognosis in patients with PCa. The samples from patients with PCa were mixed tissues, including tumor cells, normal prostate cells, stromal cells, and immune cells. Tumor malignancy was determined based on several factors, including tumor purity, tumor proliferation ability, and the TME *in vivo*. Thus, our results differed in part from those concluded *in vitro*, where tumor malignancy was determined by the tumor cells only. Normal prostate cells are smooth muscle cells that exhibit high metabolic levels, whereas tumor cells may show lower levels of metabolic activity compared to normal prostate cells. Meanwhile, ssGSEA scores based on mRNA-seq data from second-generation sequencing reflects the metabolic level of the whole sample, rather than only tumor cells. Therefore, in this study, we determined a lower metabolic level, and that greater tumor heterogeneity of the PCa samples was associated with a worse prognosis for the patient. This was in accordance with that reported for liver cancer (60). These specifications may partly explain why tumors from the C1 subset had the worst prognosis in PCa.

According to the above subtype analyses, we consider that C2 is the subtype that characterizes early PCa. In fact, during the early stage, the tumor metabolic status often resembles that of normal tissue. Besides, during the initial stage, the immune system exerts a strong anti-tumor response, and the tumor has low heterogeneity and stemness index. On the contrary, due to immune escape and the decrease of blood supply in the later stage of tumor progression, immune cells are not able to infiltrate the tumor. Thus, C1 has milder immune characteristics than C2.

Nevertheless, because C1 is an advanced tumor, it had the highest tumor heterogeneity and the lowest metabolic status among the three subtypes. As for C3, which is regarded as an intermediate between C1 and C2, we consider it to be the crucial status of PCa from early to advanced tumor stage. The unique metabolic pathways of C3 reportedly affect the malignant transformation of healthy tissue.

To predict prognosis, previous studies have developed prognostic risk models for PCa based on gene expression data (61–63). This indicates that risk models based on high-throughput data may accurately predict the prognosis of PCa. Therefore, we used WGCNA to identify the characteristic genes of C1 and C2. Using multiple algorithms, a six-gene risk model, including *CACNG4*, *SLC2A4*, *EPHX2*, *CA14*, *NUDT7*, and *ADH5*, was established using the TCGA cohort. Through the testing of four external datasets, our metabolism-associated risk model was demonstrated to have strong robustness. GSEA analysis provided further evidence that our risk genes are related to PCa metabolism. Finally, we combined risk score and clinical variates to obtain a nomogram to help clinicians predict the DFS for PCa patients.

Many studies have demonstrated that gene expression data can be used to predict drug treatment sensitivity. For instance, molecular profiling can be used to identify treatment-refractory metastatic castration-resistant prostate cancer (64). C1 had the highest AR expression. This indicates that C1 may exhibit high sensitivity to anti-androgenic therapy. With a high expression of immune checkpoints and significant features of immune cell infiltration, C2 tumors may benefit more from immune-targeted therapy, whereas C3, with an active status of specific drug metabolism pathways, may facilitate the development of tolerance to traditional chemotherapy (65). In the drug sensitivity analysis of the risk model, our study suggests that patients with low-risk scores may benefit more from anti-PD-1 treatment, and this is consistent with C2 being more suitable for immune target therapy. All of these conclusions need to be validated *in vitro*.

In contrast to previous research that focused on the metabolic level of single tumor cell types, we explored the metabolic features of mixed cancer samples. We first investigated the characteristics of metabolic pathways using cluster analysis and explored tumor heterogeneity in multiple dimensions employing multi-omics. Finally, our risk model of PCa was constructed and verified using a large number of samples and multiple datasets. However, our research also had limitations. First, the data we studied were from public databases rather than our database. Second, we did not perform *in vitro* or *in vivo* experiments to further investigate the mechanism of metabolism-associated genes in PCa. These are what we plan to do next.

CONCLUSIONS

Three metabolism-associated subtypes were first identified by unsupervised cluster and ssGSEA analyses in PCa. Differential analyses indicated these subtypes could reflect tumor

heterogeneity in the stemness index, tumor microenvironment, TMB, CNV, MSI, and clinical features. So our metabolism-associated subtypes can better represent the metabolic characteristics of PCa and can be beneficial in exploring the metabolic mechanism of occurrence and development of PCa. Meanwhile, a six-gene metabolism-associated risk score model by using four separate datasets and demonstrated strong robustness in the prediction of sample types (tumor and normal tissue), DFS, metabolism-associated subtypes, and anti-tumor therapeutic effect. Therefore our model can powerfully help clinicians evaluate the prognosis and develop personalized treatment for PCa patients. Although the six prognostic markers still require experimental verification, they may provide insight and a prospect for further investigation and clinical work regarding PCa.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/ (PRAD); https://www.ncbi.nlm.nih.gov/gds/ (GSE54460, 70768, 78220); https://www.cancerrxgene.org/ (22RV1, DU-145, LNCaP-Clone-FGC, PC-3, PWR-1E, and VCaP); https://www.gtexportal.org/ (Prostate); https://icgc.org/(PRAD).

AUTHOR CONTRIBUTIONS

YZ designed the study and analyzed the data. YZ and RZ revised the images. YZ, RZ, and FL performed the literature search and collected data for the manuscript. XL and LZ revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.598801/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | The PAM analysis of ssGSEA score based on metabolism-associated pathways. **(A)** Consensus among clusters for each category number K. Color-coded heatmap corresponding to the consensus matrix

for k = 2 **(B)**, K = 3 **(C)**, K = 4 **(D)**, K = 5 **(E)**, and K = 6 **(F)** obtained by applying consensus clustering.

SUPPLEMENTARY FIGURE 2 | Association between clinical characteristics and the metabolism-associated subtypes. Age **(A)** and Lymph Nodes (N) stage **(B)** for each metabolism-associated subtype in the TCGA cohort. **(C)** The pairwise comparison of the PSA between three subtypes. The P values are labeled above each boxplot with asterisks (ns represents no significance, *P <0.05, **P <0.01, ***P <0.001).

SUPPLEMENTARY FIGURE 3 | Association between the stemness index and the metabolism-associated subtypes. The pairwise comparison of the ssGSEAsi **(A)** and COssGSEAsi **(B)** between three subtypes. The P values are labeled above each boxplot with asterisks (ns represents no significance, *P <0.05, **P <0.01, ***P <0.001). **(C)** The correlation analysis between specific metabolic pathways of C3 and ssGSEAsi in the TCGA cohort.

SUPPLEMENTARY FIGURE 4 | The WCGNA analysis of metabolism-associated genes among three subtypes. **(A)** Analysis of the scale-free fit index and the mean connectivity for various soft-thresholding powers, and checking the scale-free topology when β = 3. K represents the logarithm of whole network connectivity, p(k) represents the logarithm of the corresponding frequency distribution. K is negatively correlated with p(k) (correlation coefficient = 0.88), which represents scale-free topology. **(B)** Identification of a co-expression module in PCa. The branches of the cluster dendrogram correspond to the five different gene modules. Each piece of the leaves on the cluster dendrogram corresponds to a gene. **(C)** Scatter plot of module eigengenes in the blue and green modules.

SUPPLEMENTARY FIGURE 5 | Identification of risk model functions and development of the nomogram. **(A**, **B)** prognostic genes selected through Lasso regression. **(C, D)** Enrichment plots of the top five KEGG pathways in the high-risk score and low-risk score groups in PRAD. The relationship between the age **(E)**/Gleason score **(F)**/N stage **(G)**/T stage **(H)** and risk score in the TCGA cohort. **(I)** Nomogram for predicting the probability of 1, 3, and 5 disease-free survival times for PCa patients. **(J)** ROC analysis for nomogram in 1, 3, and 5 years. **(K)** The C-index analysis for clinical variates, risk model, and nomogram.

SUPPLEMENTARY FIGURE 6 | The K-M survival analysis of six risk model genes. **(A)** ADH5. **(B)** CA14. **(C)** CACNG4. **(D)** EPHX2. **(E)** NUDT7. **(F)** SLC2A4.

SUPPLEMENTARY FIGURE 7 | The protein expression of three risk model genes in normal prostate tissues and PRAD tissues from the Human Protein Atlas (HPA) database. **(A)** ADAH5. **(B)** EPHX2T. **(C)** NUDT7N.

SUPPLEMENTARY TABLE 1 | The 41 metabolism pathway gene sets collected to evaluate the metabolic status of PRAD samples.

SUPPLEMENTARY TABLE 2 | The ssGSEA score of each sample of PRAD in TCGA based on metabolic pathways, stem cell, EMT, ECM, and TGF-β gene sets.

SUPPLEMENTARY TABLE 3 | The WGCNA analysis between the expression of metabolism-associated genes and metabolism-associated subtypes in PRAD.

SUPPLEMENTARY TABLE 4 | The GO and KEGG functional enrichment analysis of metabolism-associated module genes.

SUPPLEMENTARY TABLE 5 | The Univariable Cox analysis of metabolism-associated prognostic genes.

SUPPLEMENTARY TABLE 6 | The GSEA analysis of risk score in TCGA.

SUPPLEMENTARY TABLE 7 | The Univariable and Multivariable Cox analysis of clinical variates and risk score.

SUPPLEMENTARY TABLE 8 | Drug sensitivity analysis of risk model genes and risk score.

REFERENCES

1. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. *Global Cancer Observatory: Cancer Today*. Lyon, France: International Agency for Research on Cancer (2018). Available at: https://gco.iarc.fr/today.

2. Culp MB, Soerjomataram I, Efstathiou JA, Bray F, Jemal A. Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. *Eur Urol* (2020) 77:38–52. doi: 10.1016/j.eururo.2019.08.005

3. Nassar ZD, Mah CY, Dehairs J, Burvenich IJ, Irani S, Centenera MM, et al. Human DECR1 is an androgen-repressed survival factor that regulates PUFA oxidation to protect prostate tumor cells from ferroptosis. *Elife* (2020) 9:e54166. doi: 10.7554/eLife.54166

4. Franko A, Shao Y, Heni M, Hennenlotter J, Hoene M, Hu C, et al. Human Prostate Cancer is Characterized by an Increase in Urea Cycle Metabolites. *Cancers (Basel)* (2020) 12(7):1814. doi: 10.3390/cancers12071814

5. Song B-N. Identification of an immunotherapy-responsive molecular subtype of bladder cancer. *EBioMedicine* (2019) 50:238–45. doi: 10.1016/j.ebiom.2019.10.058

6. Chen W, Zhuang J, Wang PP, Jiang J, Lin C, Zeng P, et al. DNA methylation-based classification and identification of renal cell carcinoma prognosis-subgroups. *Cancer Cell Int* (2019) 19:185. doi: 10.1186/s12935-019-0900-4

7. Xiao H, Wang J, Yan W, Cui Y, Chen Z, Gao X, et al. GLUT1 regulates cell glycolysis and proliferation in prostate cancer. *Prostate* (2018) 78:86–94. doi: 10.1002/pros.23448

8. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* (2012) 2:401–404. doi: 10.1158/2159-8290.CD-12-0095

9. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol Res* (2014) 58:234–239. doi: 10.1007/s12026-014-8516-1

10. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* (2005) 102:15545–15550. doi: 10.1073/pnas.0506580102

11. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* (1999) 27:29–34. doi: 10.1093/nar/27.1.29

12. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* (2010) 26:1572–1573. doi: 10.1093/bioinformatics/btq170

13. 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) 173:338–54.e15. doi: 10.1016/j.cell.2018.03.034

14. Miranda A, Hamilton PT, Zhang AW, Pattnaik S, Becht E, Mezheyeuski A, et al. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. *Proc Natl Acad Sci USA* (2019) 116:9020–9. doi: 10.1073/pnas.1818210116

15. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612

16. Botía JA, Vandrovcova J, Forabosco P, Guelfi S, D'Sa KUnited Kingdom Brain Expression Consortium, et al. An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst Biol* (2017) 11:47. doi: 10.1186/s12918-017-0420-6

17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559

18. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

19. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models *via* Coordinate Descent. *J Stat Softw* (2010) 33:1–22.

20. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model *via* Coordinate Descent. *J Stat Softw* (2011) 39:1–13. doi: 10.18637/jss.v039.i05

21. Mogensen UB, Ishwaran H, Gerds TA. Evaluating Random Forests for Survival Analysis using Prediction Error Curves. *J Stat Softw* (2012) 50:1–23. doi: 10.18637/jss.v050.i11

22. Tikkinen KAO, Dahm P, Lytvyn L, Heen AF, Vernooij RWM, Siemieniuk RAC, et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a clinical practice guideline. *BMJ* (2018) 362:k3581. doi: 10.1136/bmj.k3581

23. Lytle NK, Barber AG, Reya T. Stem cell fate in cancer growth, progression and therapy resistance. *Nat Rev Cancer* (2018) 18:669–680. doi: 10.1038/s41568-018-0056-x

24. Wang L, Zi H, Luo Y, Liu T, Zheng H, Xie C, et al. Inhibition of Notch pathway enhances the anti-tumor effect of docetaxel in prostate cancer stem-like cells. *Stem Cell Res Ther* (2020) 11:258. doi: 10.1186/s13287-020-01773-w

25. Qian J, Olbrecht S, Boeckx B, Vos H, Laoui D, Etlioglu E, et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res* (2020) 30(9):745–62. doi: 10.1038/s41422-020-0355-0

26. Chen Q, Gu M, Cai Z-K, Zhao H, Sun S-C, Liu C, et al. TGF-β1 promotes epithelial-to-mesenchymal transition and stemness of prostate cancer cells by inducing PCBP1 degradation and alternative splicing of CD44. *Cell Mol Life Sci* (2020) doi: 10.1007/s00018-020-03544-5

27. Li N, Kang Y, Wang L, Huff S, Tang R, Hui H, et al. ALKBH5 regulates anti-PD-1 therapy response by modulating lactate and suppressive immune cell accumulation in tumor microenvironment. *Proc Natl Acad Sci USA* (2020) 117(33):20159–170. doi: 10.1073/pnas.1918986117

28. Goliwas KF, Deshane JS, Elmets CA, Athar M. Moving Immune Therapy Forward Targeting TME. *Physiol Rev* (2020). doi: 10.1152/physrev.00008.2020

29. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* (2015) 348:124–8. doi: 10.1126/science.aaa1348

30. Vagner T, Spinelli C, Minciacchi VR, Balaj L, Zandian M, Conley A, et al. Large extracellular vesicles carry most of the tumour DNA circulating in prostate cancer patient plasma. *J Extracell Vesicles* (2018) 7:1505403. doi: 10.1080/20013078.2018.1505403

31. Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* (2011) 476:346–50. doi: 10.1038/nature10350

32. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* (2014) 30:1015–6. doi: 10.1093/bioinformatics/btt755

33. Cortes-Ciriano I, Lee S, Park W-Y, Kim T-M, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun* (2017) 8:15180. doi: 10.1038/ncomms15180

34. Attard G. Anti-androgen monotherapy for metastatic prostate cancer. *Lancet Oncol* (2014) 15:543–4. doi: 10.1016/S1470-2045(14)70159-7

35. Kokal M, Mirzakhani K, Pungsrinont T, Baniahmad A. Mechanisms of Androgen Receptor Agonist- and Antagonist-Mediated Cellular Senescence in Prostate Cancer. *Cancers (Basel)* (2020) 12(7):1833. doi: 10.3390/cancers12071833

36. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* (2018) 554:544–8. doi: 10.1038/nature25501

37. Topalian SL, Taube JM, Anders RA, Pardoll DM. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat Rev Cancer* (2016) 16:275–87. doi: 10.1038/nrc.2016.36

38. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* (2018) 174:1033. doi: 10.1016/j.cell.2018.07.036

39. Tan TZ, Ye J, Yee CV, Lim D, Ngoi NYL, Tan DSP, et al. Analysis of gene expression signatures identifies prognostic and functionally distinct ovarian clear cell carcinoma subtypes. *EBioMedicine* (2019) 50:203–10. doi: 10.1016/j.ebiom.2019.11.017

40. Yang C, Zhang Y, Xu X, Li W. Molecular subtypes based on DNA methylation predict prognosis in colon adenocarcinoma patients. *aging* (2019) 11:11880–92. doi: 10.18632/aging.102492

41. Najafi M, Farhood B, Mortezaee K. Cancer stem cells (CSCs) in cancer progression and therapy. *J Cell Physiol* (2019) 234:8381–95. doi: 10.1002/jcp.27740

42. Zhang Z, Zhang S, Yang J, Yi P, Xu P, Yi M, et al. Integrated transcriptomic and metabolomic analyses to characterize the anti-cancer effects of (-)-epigallocatechin-3-gallate in human colon cancer cells. *Toxicol Appl Pharmacol* (2020) 401:115100. doi: 10.1016/j.taap.2020.115100

43. Rennier K, Shin WJ, Krug E, Virdi G, Pachynski RK. Chemerin Reactivates PTEN and Suppresses PD-L1 in Tumor Cells *via* Modulation of a Novel CMKLR1-mediated Signaling Cascade. *Clin Cancer Res* (2020) 26(18):5019–35. doi: 10.1158/1078-0432.CCR-19-4245

44. Billon E, Finetti P, Bertucci A, Niccoli P, Birnbaum D, Mamessier E, et al. PDL1 expression is associated with longer postoperative, survival in adrenocortical carcinoma. *Oncoimmunology* (2019) 8:e1655362. doi: 10.1080/2162402X.2019.1655362

45. Park J, Kim D-H, Shah SR, Kim H-N, Kshitiz, Kim P, et al. Switch-like enhancement of epithelial-mesenchymal transition by YAP through feedback regulation of WT1 and Rho-family GTPases. *Nat Commun* (2019) 10:2797. doi: 10.1038/s41467-019-10729-5

46. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M, et al. Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* (2020) 21:341–52. doi: 10.1038/s41580-020-0237-9

47. Su Y, Feng W, Shi J, Chen L, Huang J, Lin T. circRIP2 accelerates bladder cancer progression *via* miR-1305/Tgf-β2/smad3 pathway. *Mol Cancer* (2020) 19:23. doi: 10.1186/s12943-019-1129-5

48. Wu N, Jiang M, Liu H, Chu Y, Wang D, Cao J, et al. LINC00941 promotes CRC metastasis through preventing SMAD4 protein degradation and activating the TGF-β/SMAD2/3 signaling pathway. *Cell Death Differ* (2020). doi: 10.1038/s41418-020-0596-y

49. Bruni D, Angell HK, Galon J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat Rev Cancer* (2020) 20(11):662–80. doi: 10.1038/s41568-020-0285-7

50. Ren J, Nie Y, Lv M, Shen S, Tang R, Xu Y, et al. Estrogen upregulates MICA/B expression in human non-small cell lung cancer through the regulation of ADAM17. *Cell Mol Immunol* (2015) 12:768–76. doi: 10.1038/cmi.2014.101

51. Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol* (2020) 20(11):651–68. doi: 10.1038/s41577-020-0306-5

52. Salem ME, Bodor JN, Puccini A, Xiu J, Goldberg RM, Grothey A, et al. Relationship between MLH1, PMS2, MSH2 and MSH6 gene-specific alterations and tumor mutational burden in 1057 microsatellite instability-high solid tumors. *Int J Cancer* (2020) 147(10):2948–56. doi: 10.1002/ijc.33115

53. Roze J, Monroe G, Kutzera J, Groeneweg J, Stelloo E, Paijens S, et al. Whole Genome Analysis of Ovarian Granulosa Cell Tumors Reveals Tumor Heterogeneity and a High-Grade TP53-Specific Subgroup. *Cancers (Basel)* (2020) 12(5):1308. doi: 10.3390/cancers12051308

54. Nyquist MD, Corella A, Coleman I, De Sarkar N, Kaipainen A, Ha G, et al. Combined TP53 and RB1 Loss Promotes Prostate Cancer Resistance to a Spectrum of Therapeutics and Confers Vulnerability to Replication Stress. *Cell Rep* (2020) 31:107669. doi: 10.1016/j.celrep.2020.107669

55. Swami U, Isaacsson Velho P, Nussenzveig R, Chipman J, Sacristan Santos V, Erickson S, et al. Association of SPOP Mutations with Outcomes in Men with De Novo Metastatic Castration-sensitive Prostate Cancer. *Eur Urol* (2020) 78(5):652–6. doi: 10.1016/j.eururo.2020.06.033

56. Su F, Zhang W, Zhang D, Zhang Y, Pang C, Huang Y, et al. Spatial Intratumor Genomic Heterogeneity within Localized Prostate Cancer Revealed by Single-nucleus Sequencing. *Eur Urol* (2018) 74:551–9. doi: 10.1016/j.eururo.2018.06.005

57. Xu Q, Li Y, Gao X, Kang K, Williams JG, Tong L, et al. HNF4α regulates sulfur amino acid metabolism and confers sensitivity to methionine restriction in liver cancer. *Nat Commun* (2020) 11:3978. doi: 10.1038/s41467-020-17818-w

58. Vantaku V, Putluri V, Bader DA, Maity S, Ma J, Arnold JM, et al. Epigenetic loss of AOX1 expression *via* EZH2 leads to metabolic deregulations and promotes bladder cancer progression. *Oncogene* (2019) 39(40):6265–85. doi: 10.1038/s41388-019-0902-7

59. Gandhi M, Groß M, Holler JM, Coggins SA, Patil N, Leupold JH, et al. The lncRNA lincNMR regulates nucleotide metabolism *via* a YBX1 - RRM2 axis in cancer. *Nat Commun* (2020) 11:3214. doi: 10.1038/s41467-020-17007-9

60. Yang C, Huang X, Liu Z, Qin W, Wang C. Metabolism-associated molecular classification of hepatocellular carcinoma. *Mol Oncol* (2020) 14:896–913. doi: 10.1002/1878-0261.12639

61. Zhang E, Hou X, Hou B, Zhang M, Song Y. A risk prediction model of DNA methylation improves prognosis evaluation and indicates gene targets in prostate cancer. *Epigenomics* (2020) 12:333–52. doi: 10.2217/epi-2019-0349

62. Cao Z-X, Xiao G-A, Zhang W, Ji J, Ye C, Liu D, et al. Comprehensive investigation of alternative splicing and development of a prognostic risk score for prostate cancer based on six-gene signatures. *J Cancer* (2019) 10:5585–96. doi: 10.7150/jca.31725

63. Hu D, Jiang L, Luo S, Zhao X, Hu H, Zhao G, et al. Development of an autophagy-related gene expression signature for prognosis prediction in prostate cancer patients. *J Transl Med* (2020) 18:160. doi: 10.1186/s12967-020-02323-x

64. Labrecque MP, Coleman IM, Brown LG, True LD, Kollath L, Lakely B, et al. Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. *J Clin Invest* (2019) 129:4492–505. doi: 10.1172/JCI128212

65. Brecht K, Schäfer AM, Meyer Zu Schwabedissen HE. Uptake Transporters of the SLC21, SLC22A, and SLC15A Families in Anticancer Therapy-Modulators of Cellular Entry or Pharmacokinetics? *Cancers (Basel)* (2020) 12(8):2263. doi: 10.3390/cancers12082263