



OPEN Pan-cancer analysis identifies *PSMA7* as a targets for amplification at 20q13.33 in tumorigenesis

Guangying Sheng^{1,2,4}, Fuyu Li^{1,2,3,4}, Wen Jin¹ & Kankan Wang^{1,2,3}✉

The chromosome 20 long arm (20q) is one of the genomic hotspots where copy number alterations frequently occur in multiple types of tumors. However, it remains elusive which genes are implicated in 20q-related tumorigenesis. Here, by querying TCGA and GEO databases, we observed frequent copy number amplification at 20q and the chromosome subband 20q13.33 was amplified in multiple cancers. Among those genes at 20q13.33, *PSMA7* was found with the strongest correlation with cancers. Further analysis revealed that *PSMA7* amplification was the most frequent genetic alteration event conferring adverse prognosis in various cancers. Consistent with the strong positive correlation between *PSMA7* amplification and gene expression, elevated *PSMA7* expression was observed in 20 of 33 types of cancers with a close link to adverse outcomes in certain tumors. In addition, *PSMA7* was essential for the growth of almost 1095 cancer lines. Mechanistically, aberrant *PSMA7* most probably influenced the proteasome and protease-related pathways to promote tumorigenesis and might be antagonized by several compounds, e.g., Docetaxel in relevant cancers. The current in-depth pan-cancer analysis refines our understanding of the crucial oncogenic role of copy number amplifications at *PSMA7* loci at the novel chromosome amplicon 20q13.33 across different tumors.

Copy number alterations (CNAs) are one of the types of genomic alterations that additionally consist of mutation (single nucleotide variation), chromosomal inversion, and translocation among others¹. CNAs include amplifications (more copies, often focal), gains (a few additional copies, often broad), and deep and shallow deletions/losses (homozygous and heterozygous deletions)². CNAs can impair genomic stability and dysregulate the involved genes to exert pro- or anti-tumor effects². Previous scattered studies revealed the chromosome 20 long arm (20q) as one of the CNA hotspots harboring various cancer-related genes³. The 20q was frequently deleted or lost in hematological neoplasms⁴, while in ovarian cancer, upregulated genes and relevant aberrant DNA amplification were commonly concentrated at 20q⁵. Moreover, abnormal amplification of 20q was commonly reported in Barrett adenocarcinoma (BA+) tumor⁶ and colorectal⁷, breast⁸, prostate⁹, pancreatic¹⁰, and bladder¹¹ cancers. Notably, Murty et al.¹² found 20q11.2 and 20q13.13 as the novel chromosomal amplification loci carrying tumor-related genes in cervical cancer. Although these findings suggested critical roles of amplification at 20q in various tumors, it remained unclear which genes played a predominant role at 20q.

The proteasome subunit alpha-type 7 (*PSMA7*) encodes the alpha-type subunit of the 20S proteasome core complex from the 26S proteasome of the ubiquitin–proteasome system (UPS). By forming a regulatory structure “gateway” with other proteasome subunits, *PSMA7* can modulate protein degradation¹³. Moreover, *PSMA7* interacts with multiple important proteins including c-ABL¹⁴, hypoxia-inducible factor-1 α (HIF-1 α)¹⁵, hepatitis B virus (HBV) X (HBX) protein¹⁶, Rab7¹⁷, and endothelial monocyte-activating polypeptide 2 (EMAP2)¹⁸. Thanks to these interactions, *PSMA7* is implicated in diverse cancer-related cellular activities, including transcription control, immunologic reaction, stress responses, cell cycle regulation, cellular differentiation, proliferation, and apoptosis. Additionally, c-ABL-mediated upregulation of *PSMA7* led to the formation of alternative proteasome isoforms of *PSMA7*-*PSMA7*, eventually influencing cancer cell fitness^{19,20}.

¹State Key Laboratory of Medical Genomics, Ruijin Hospital Shanghai Institute of Hematology, National Research Center for Translational Medicine at Shanghai, Shanghai Jiao Tong University School of Medicine, 197 Ruijin Er Rd, Shanghai 200025, China. ²Ruijin Hospital, Sino-French Research Center for Life Sciences and Genomics, Shanghai Jiao Tong University School of Medicine, Shanghai, China. ³School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. ⁴These authors contributed equally: Guangying Sheng and Fuyu Li. ✉email: kankanwang@shsmu.edu.cn

PSMA7 overexpression has been observed in gastric^{21,22}, cervical^{12,23}, and testicular²⁴ cancers and is shown to be closely linked to metastasis²⁵, progression²⁶, and reduced life span^{27,28} in certain cancers. Mechanistically, deficiency of *PSMA7* could inhibit cellular proliferation, migration, and invasion in colorectal²⁹, gastric²², and breast³⁰ cancers, and high-grade serous ovarian carcinoma (HGSOC)³¹. Besides, *PSMA7* depletion blunted the G₀/G₁ cell cycle and led to apoptosis of cervical cancer cells²³. It was evidenced that the *PSMA7*-mediated proteasome degradation was essential to maintain cell survival of HGSOC³¹, cervical²³, and castration-recurrent prostate cells³². Additionally, *PSMA7* was found to contribute to tumorigenesis by activating the nuclear factor- κ -gene binding (NF- κ B) pathway³³ and the mitogen-activated protein kinase (MAPK) signaling pathway²², and interactions with miRNAs³⁰. All these discoveries point to an oncogenic role of *PSMA7* in certain tumors. However, the precise roles and relevant mechanisms of *PSMA7* across human cancers remained less well understood.

Hereby, combining the cancer genome atlas (TCGA) and gene expression omnibus (GEO) data, we carried out a pan-cancer exploration of *PSMA7* at 20q concerning genetic alteration, survival state, gene expression, cellular pathway, molecular function, and potential drug screening, aiming to uncover the mechanisms of *PSMA7* in tumorigenesis and clinical outcomes across diverse human cancers.

Results

The amplification of chromosome 20q at the pan-cancer level

To depict the genomic landscape of chr20 in human tumors, we employed the revised Genomic Identification of Significant Targets in Cancer (GISTIC2.0) pipeline^{34–36} to integrate the copy number raw data from 33 types of TCGA tumors (Fig. 1A, B, Supplementary Tabs. S1–S3). Overall, the statistical data for each chromosome arm revealed that the broad genomic amplification was more significant in the 20q than in the short arm of chr20 (20p) at the pan-cancer level (Fig. 1A, Supplementary Tab. S1). Consistently, this discovery was confirmed when the online SNP6 copy number analysis (GISTIC2) from Broad Firehose³⁵ was run by each type of TCGA cancers in turn (Supplementary Fig. S1A). Besides, the GISTIC2.0-based pan-cancer integrated analysis detected only 3 focal events (amplification: 2; deletion: 1) located at chr20 (Fig. 1B, Supplementary Tabs. S2, S3), suggesting that amplification at chr20 was mostly in a broad level.

After downloading the data for CNA genes from the cBioPortal^{37,38}, we queried the median frequency of amplification of mapped genes in tumor samples and further illustrated the amplification frequency in the corresponding chr20 subband (Table 1). Excluding chromosome subbands with less than 10 genes, the overall amplification frequency of 20q was 2.25%, which was significantly higher than that of 20p (0.8%). This finding highlighted the predominance of amplification in 20q (Fig. 1A, Supplementary Fig. S1A). Notably, 20q13.33, which involves 85 genes, displayed the highest alteration frequency of 2.7%, highlighting this region as a novel chromosome amplicon. Although amplification at this region was identified as a broad CNA event by pan-cancer integrated copy number analysis (Fig. 1B, Supplementary Tabs. S2, S3), the focal amplification of 20q13.33 was further found in 7 (of 30) tumor types of liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), ovarian serous cystadenocarcinoma (OV), rectum adenocarcinoma (READ), sarcoma (SARC), skin cutaneous melanoma (SKCM), and uterine corpus endometrial carcinoma (UCEC), according to detailed amplification plots generated by the online tool GISTIC2 from Broad Firehose (Supplementary Fig. S1B). Other relatively frequent chromosome subbands included 20q13.32, 20q13.2, and 20q13.31, with an amplification frequency of 2.6%, 2.6%, and 2.5%, respectively. 20q13.13 was another candidate amplicon with a relatively higher amplification frequency of 2.3% (Table 1), which had been identified in cervical cancer previously¹². These findings revealed that amplification was a quite common genetic event for 20q, and 20q13.33 was likely to be an uncharted amplicon in human cancers.

To investigate which pathways were mainly involved in the 20q13.33-amplified cancers, we conducted an enrichment analysis of the 85 genes at 20q13.33 (Fig. 1C, D). The Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO) data manifested that most genes were significantly enriched in multiple cancer-related cellular and molecular functions, especially of translocations between the nucleus, nucleoplasm, and chromatin, cell proliferation, and differentiation (Fig. 1D). However, it remained undetermined which gene primarily mediated the effects of amplified 20q13.33 on cancer initiation and development. Thus, we used the Phenolyzer³⁹ tool to evaluate the potential association of the 85 genes at 20q13.33 with cancer development. Notably, *PSMA7*, one of the proteasome family members, achieved the highest correlation score of 0.452 with cancer (Fig. 1E), suggesting the crucial oncogenic role of *PSMA7* in 20q13.33. We thus focused on *PSMA7* for further study.

PSMA7 gene amplification occurs in multiple types of tumors.

To comprehensively explore the genetic alterations of the *PSMA7* gene locus rather than CNA^{1,2}, we conducted a genetic alteration survey of *PSMA7* across 32 types of TCGA cancers via the cBioPortal approach. The genetic alterations of *PSMA7* were found in a total of 25 types of tumors, and amplification was the most frequent molecular alteration observed in *PSMA7* (Fig. 2A). The uterine carcinosarcoma (UCS) cases carried the utmost amplification incidence of *PSMA7* (~8.8%), followed by colon adenocarcinoma (COAD) (~6.7%) and OV (~6.2%). Finally, we discovered that cohorts with thymoma (THYM) harbored deep deletion of *PSMA7* with the highest frequency (0.81%) but represented a minority of cancers.

Consistent with the above results, according to the Gene Set Cancer Analysis (GSCA) (a TCGA-based tool)⁴⁰, the CNA of *PSMA7* was the major genetic alteration event observed in 33 types of TCGA tumors (Fig. 2B). The highest CNA frequency appeared in nearly 95% of patients with READ. In addition, the amplified CNA of *PSMA7* predominated in each type of tumor with a frequency of up to 80%, while the copy number deletion was most frequently observed in less than 15% of patients with lung squamous cell carcinoma (LUSC). Furthermore, both amplification and deletion of CNA occurred in heterozygous status, with the highest frequency of 80% in cases

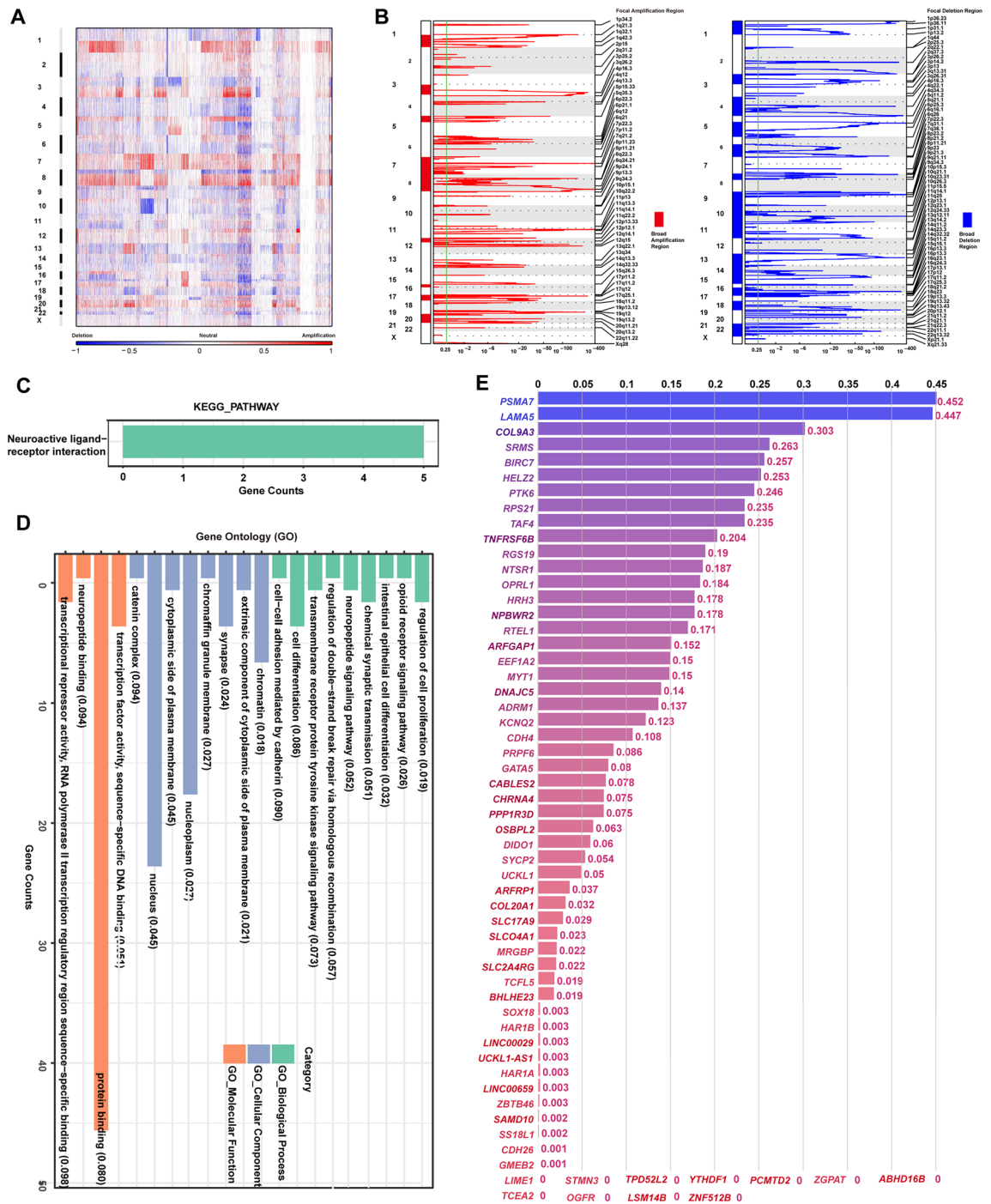


Figure 1. Copy number alterations (CNAs) of chromosome 20q across TCGA tumors. Using the revised Genomic Identification of Significant Targets in Cancer (GISTIC 2.0) pipeline, (A) the total segmented copy-number profile for each chromosome arm and (B) focal amplification and deletion events across the whole genome in 33 types of TCGA tumors were obtained and shown in a heatmap and detailed plots, respectively. Additionally, the broad CNA regions were marked as bars in the plots. Based on the gene set at 20q13.33, enrichment analysis of KEGG pathway (C) and GO (D) was conducted and shown as a bar chart with P values for each gene set attached in brackets. (E) The Phenolyzer was employed to evaluate the potential connection of “cancer” with each of the 85 genes at 20q13.33. A gene list ranked by scores was generated and visualized as bar plots.

with READ (Fig. 2C). In contrast, the homozygous CNA was most commonly observed in only 21% of patients with UCS (Fig. 2C). These data indicated that heterozygous copy number amplification of *PSMA7* represented the major genetic variation event across the TCGA tumors.

Cytband	Gene count	Median and range of frequency (%)	P value
20q	415 ^a	2.25 (1.4–3.4) ^b	<0.0001 ^c
20q11	1	3.1	NA
20q13.33	85	2.7 (2.0–2.9)	Reference
20q13.32	19	2.6 (2.3–2.7)	0.0063
20q13.2	19	2.6 (2.4–3.2)	0.8792
20q13.31	16	2.5 (2.2–2.7)	<0.0001
20q13.2-q13.31	1	2.5	NA
20q13.32-q13.33	1	2.4	NA
20q11.22-q11.23	1	2.4	NA
20q13.13	38	2.3 (1.8–2.6)	<0.0001
20q11.21-q11.22	1	2.3	NA
20q11.22	43	2.2 (1.9–2.6)	<0.0001
20q13.12	87	2.1 (1.9–2.5)	<0.0001
20q11.21	49	2.0 (2.1–3.4)	0.0220
20q12-q13.11	1	2	NA
20q13.11	2	2	NA
20q11.23	49	1.7 (1.5–2.1)	<0.0001
20q11.23-q12	1	1.5	NA
20q12	10	1.45 (1.4–1.5)	<0.0001
20q11.1	1	1.15 (0.7–2.6)	NA
20p	216 ^a	0.8 (0.6–1.5) ^b	<0.0001 ^c
20p12.3-p12.2	1	1	NA
20p13	94	0.9 (0.8–1.1)	<0.0001
20p12.3	23	0.9 (0.8–1.1)	<0.0001
20p11.21	36	0.8 (0.7–1.0)	<0.0001
20p12.1	21	0.8 (0.6–1.5)	<0.0001
20p12.2	13	0.8 (0.8–0.9)	<0.0001
20p11.1	5	0.8 (0.8–1.1)	NA
20p13-p12.3	1	0.8	NA
20p11.23	29	0.7 (0.6–0.8)	<0.0001
20p11.22	8	0.7 (0.7–0.8)	NA

Table 1. The median frequency of copy number amplification of each chr20 subband. (a) Total count of the mapped genes; (b) The median and range of amplification frequency of the mapped genes represented the amplification frequency of the corresponding subband at the long (q) or short arm (p) of chromosome 20. The chromosome subbands with a gene count below 10 were excluded. (c) chr20p vs. chr20q. NA, not available. *P* value less than 0.05 is statistically significant.

***PSMA7* amplification exhibits a positive link to gene expression**

Emerging evidence supported that genomic alterations could alter the transcript levels of mapped genes to promote cancer development and progression². Thus, we asked whether copy number amplification of *PSMA7* affected its gene expression levels. Based on the GISTIC database, the *PSMA7* expression level in the amplification of putative CNA was significantly higher than that in other types of CNA (Fig. 2D). Besides, spearman correlation analysis via the GSCA web also documented that the *PSMA7* CNA exhibited a positive link to gene expression in 23 of 33 types of TCGA tumors with false discovery rate (FDR) ≤ 0.05 (Fig. 2E). The correlations in detailed cancers were displayed in Supplementary Fig. S2, especially 0.8 in READ and 0.78 in COAD. Collectively, these data suggested that copy number amplification of *PSMA7* might lead to gene overexpression in human cancers.

Differentially expressed *PSMA7* between cancer and normal tissues

Having confirmed a close positive association of copy number amplification of *PSMA7* with gene expression (Fig. 2D, E, Supplementary Fig. S2), we further verified whether *PSMA7* was overexpressed in diverse TCGA cancers, contrasting with healthy volunteer samples via the tumor immune estimation resource 2 (TIMER2)⁴¹. In doing so, elevated *PSMA7* expression was confirmed in the cancers of bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cholangiocarcinoma (CHOL), COAD, esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), LIHC, LUAD, LUSC, READ, stomach adenocarcinoma (STAD), UCEC, and cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), prostate adenocarcinoma (PRAD), thyroid carcinoma (THCA), as compared with the matched control tissues (Fig. 3A).

Furthermore, adding the healthy samples of the genotype-tissue expression (GTEx) dataset to the current control group, we observed significantly overexpressed *PSMA7* in tumor samples of diffuse large B-cell lymphoma

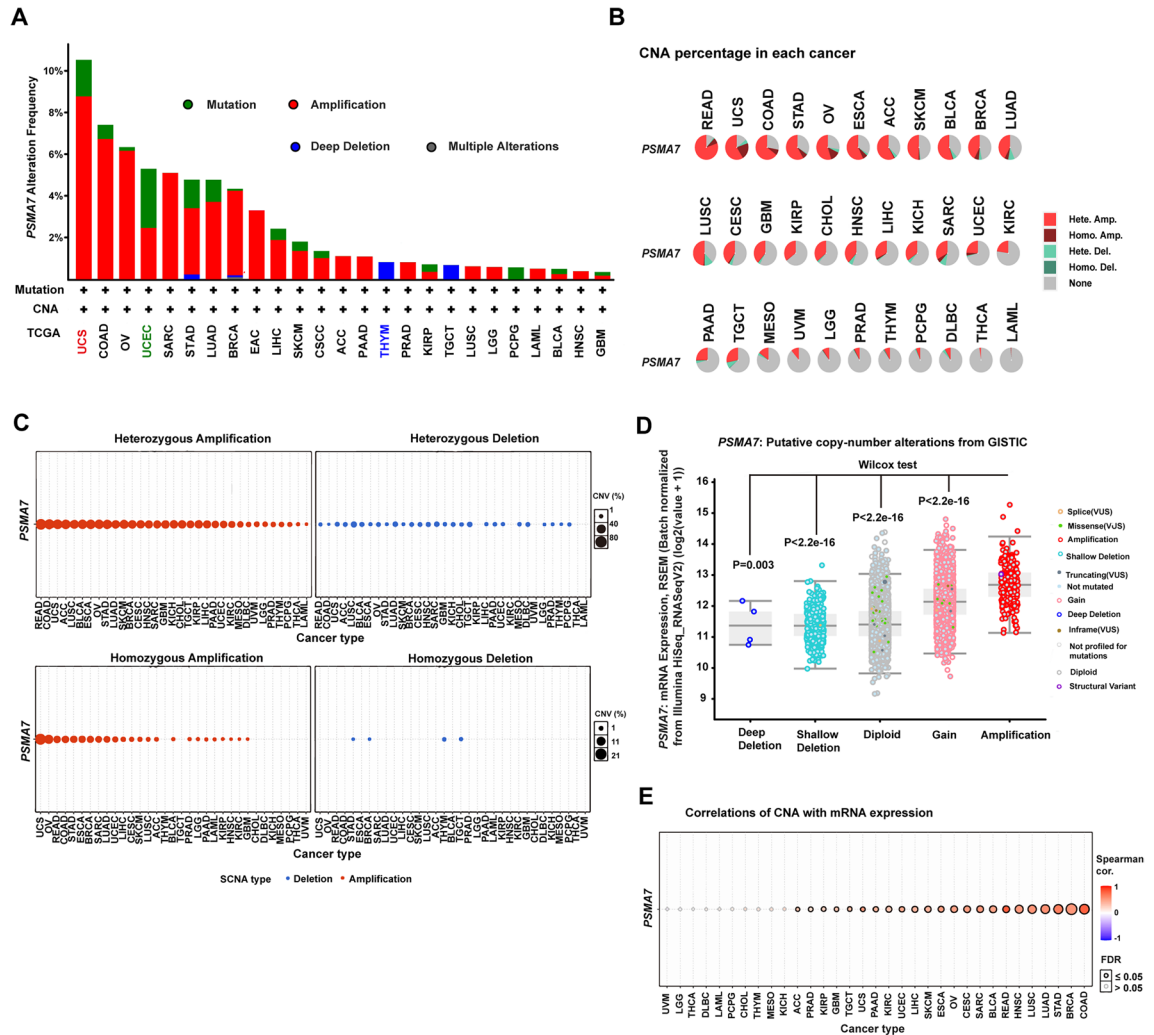


Figure 2. Copy number alteration of *PSMA7* across diverse TCGA cancers. (A) Via the cBioPortal web, the genetic variation incidence and mutation type of *PSMA7* were investigated across different tumors of TCGA. The percentage (B) and genotype status (C) of *PSMA7* CNA were also explored by the TCGA-based GSCA approach. (D) Based on the copy number data of GISTIC databases, the cBioPortal web was employed to evaluate the *PSMA7* expression level represented by different types of putative CNA. (E) Through the GSCA tool, the potential link of *PSMA7* CNA to gene expression was measured across TCGA tumors.

(DLBC), GBM, pancreatic adenocarcinoma (PAAD), and THYM (Fig. 3B), but not in other tumors, including adrenocortical carcinoma (ACC), CESC, brain lower-grade glioma (LGG), OV, SKCM, testicular germ cell tumors (TGCT), or UCS (Supplementary Fig. S3A) by the gene expression profiling interactive analysis 2 (GEPIA2)⁴². Moreover, *PSMA7* expression showed significant differences across the pathological stages of BRCA and LIHC while not including other tumors (Fig. 3C, Supplementary Fig. S3B). Taken together, these data showed that *PSMA7* overexpression frequently occurred in most human cancers.

PSMA7 amplification is an adverse prognostic factor in most tumors.

The prognostic significance of *PSMA7* CNAs (amplification, deletion, and wild type) was evaluated by log-rank survival analysis using the TCGA dataset (Fig. 4A, Supplementary Fig. S4). The kidney renal clear cell carcinoma (KIRC) patients harboring CNA amplification achieved shorter progression-free survival (PFS) and disease-specific survival (DSS) than those without CNA. Similarly, a reduced median time of DSS was observed in patients with THCA and uveal melanoma (UVM) with amplification. Nevertheless, the highly amplified CNA of *PSMA7* was linked to better PFS and disease-free interval (DFI) for ACC. In contrast, the amplified CNA of *PSMA7* was associated with worse PFS, DSS, DFI, and overall survival (OS) prognosis for UCEC. Patients with LGG with amplification had a shorter median time of OS, PFS, and DSS, and those with SARC with amplified CNA suffered much shorter OS and PFS times when compared with cases without variations. The amplified CNA of *PSMA7* was related to worse DFI for BLCA and lymphoid neoplasm DLBC, but similar OS for CESC. The above results showed that copy number amplification of *PSMA7* was a predominant genetic alteration event with adverse outcome effects on most TCGA tumors.

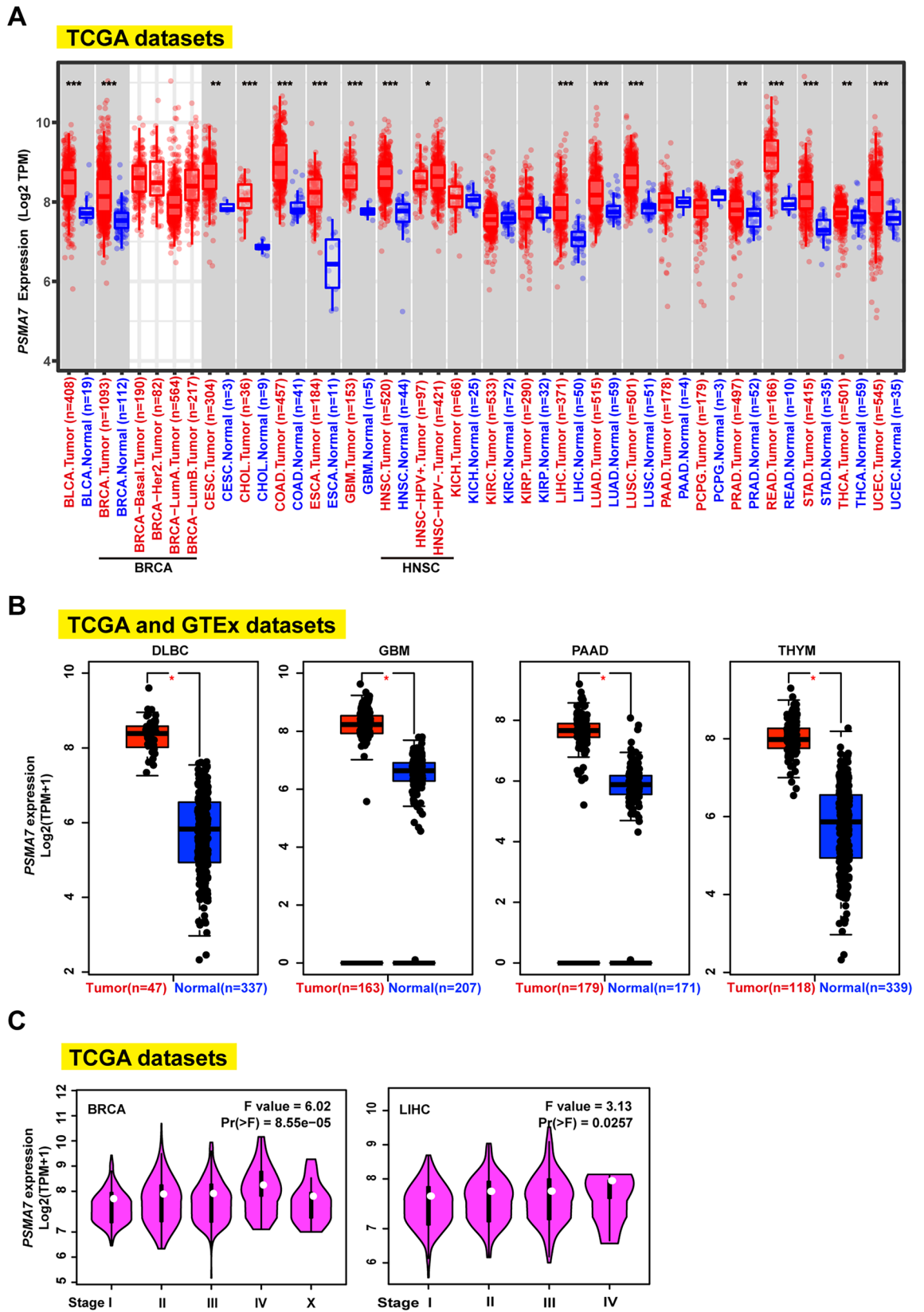


Figure 3. Expression level of *PSMA7* in various cancers. (A) Via the TIMER2 web, the *PSMA7* expression level was evaluated across diverse tumors or certain subtypes with matching normal samples from the TCGA dataset. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$. (B) The GEPIA2 was employed to generate the box plots data for the expression level of *PSMA7* across DLBC, GBM, PAAD, and THYM, with the matched healthy samples from the GTEx dataset being added into the control cohorts. * $P \leq 0.01$. (C) Based on the logarithmic scale of $\text{Log}_2(\text{TPM} + 1)$, *PSMA7* expression in each pathological stage of BRCA and LIHC was evaluated by GEPIA2.

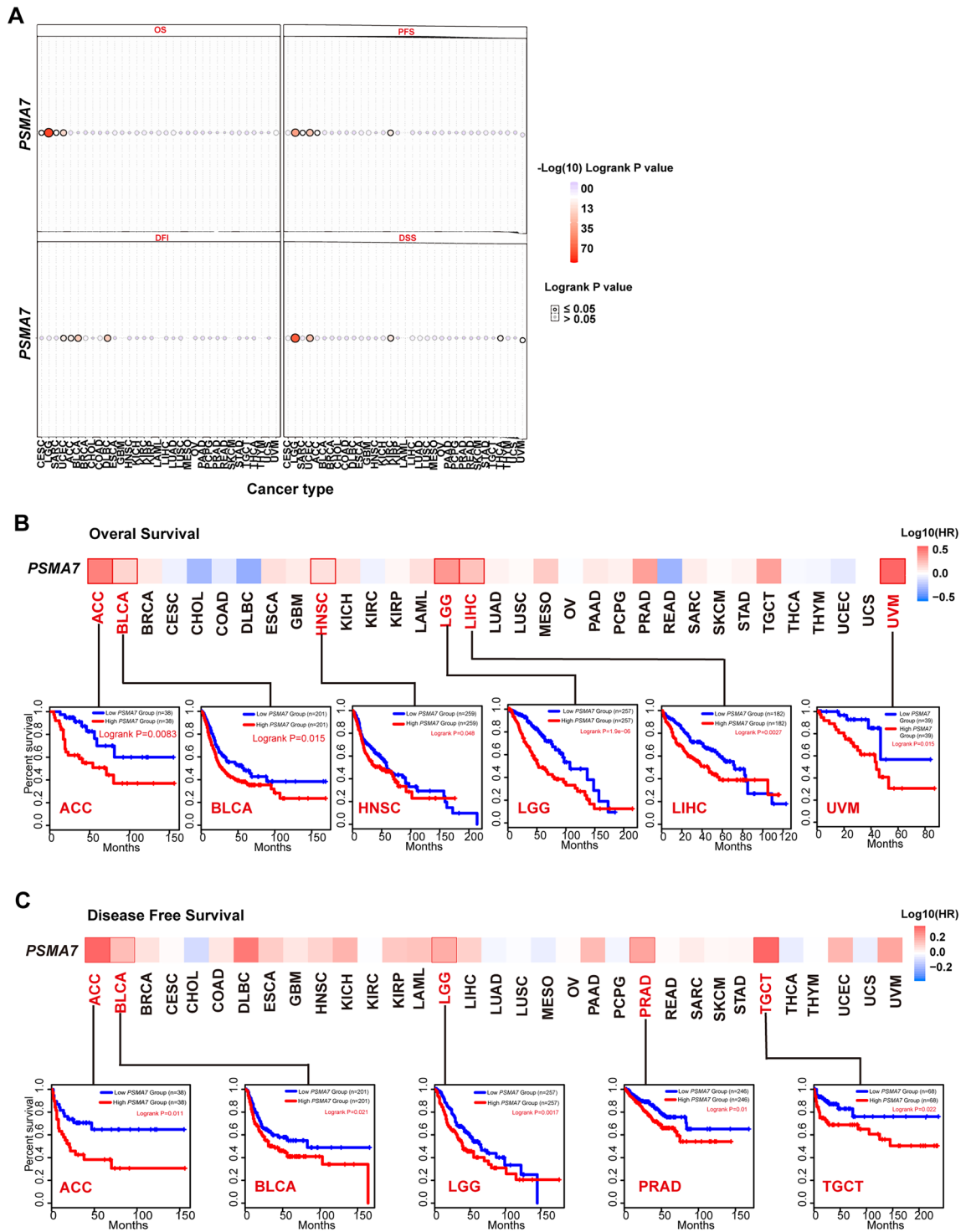


Figure 4. The prognostic significance of *PSMA7* CNA and expression on diverse TCGA tumors. **(A)** By the log-rank test, the prognostic significance of *PSMA7* CNA on TCGA tumors was estimated for OS, PFS, DSS, and DFI. **(B, C)** The potential link of *PSMA7* expression to OS and DFS prognosis for numerous TCGA cancers was investigated through the approach GEPIA2. The positive results were displayed as survival maps and K-M plots.

To evaluate whether the amplification of *PSMA7* represented an independent prognostic factor in the above tumors, we introduced a multivariate Cox regression model to adjust for established clinical variables such as patient age, tumor stage, and histological grade (Supplementary Tabs. S4-S8). Using survival data from Genomic Data Commons⁴³, we found that amplification of *PSMA7* achieved a relatively higher hazard ratio (HR) with all P values below 0.01 in OS, progression-free interval (PFI), and DSS of LGG, respectively, compared to diploid. However, this statistical trend remained significant merely for OS of LGG after isocitrate dehydrogenase (*IDH*) mutation and 1p/19q co-deletion were enrolled. Similarly, the amplified *PSMA7* exhibited significant HRs

exceeding 1 for OS and DSS in patients with SARC, contrasting to those with normal diploids. Notably, the OS of THCA showed the highest HR of 44.04 for amplified *PSMA7* with statistical significance when patient age was a continuous variable. In addition, we also observed an HR below 1 with a P value less than 0.05 for amplification of *PSMA7* in PFI of KIRC, but no positive data for common survival indicators of UCEC, BLCA, and ACC (data not shown). These findings further suggested that amplification of *PSMA7* might be an independent adverse biomarker for survival in certain types of tumors.

The overexpressed *PSMA7* is linked to adverse clinical outcomes in tumor patients

Next, we employed GEPIA2 to investigate the prognostic significance of *PSMA7* overexpression across numerous cancers according to the TCGA and GEO projects. The overexpressed *PSMA7* displayed a considerable link to the bad OS of ACC, BLCA, HNSC, LGG, LIHC, and UVM in the TCGA cohorts (Fig. 4B). Additionally, elevated *PSMA7* expression showed an association with reduced disease-free survival (DFS) in the TCGA samples of ACC, BLCA, LGG, PRAD, and TGCT (Fig. 4C).

Furthermore, the Kaplan–Meier (K–M)⁴⁴ plotter-based survival analysis was carried out to reveal the influence of high *PSMA7* expression on the clinical prognosis of breast, ovarian, lung, gastric, and liver cancers (Supplementary Fig. S5A–E). As expected, breast cancer patients with elevated *PSMA7* expression exhibited poor OS, relapse-free survival (RFS), and DFS. Similarly, overexpressed *PSMA7* showed a connection with adverse OS but not PFS or post-progression survival (PPS) for ovarian cancer, whereas poor OS and PPS but good first progression (FP) for lung carcinoma. In gastric cancer, elevated *PSMA7* expression merely exhibited a significant link with inferior FP. Moreover, the upregulated transcript level of *PSMA7* had a close link to poor OS and RFS while not PFS or DFS for liver cancer. The meta-analysis of these survival data further confirmed that *PSMA7* overexpression was an adverse factor for breast, gastric, and liver cancers but not ovarian or lung cancers (Supplementary Fig. S6). Additionally, consistent results were shown in survival analyses by subgroup when detailed clinical factors were considered (Supplementary Tabs. S9–S13). These aforementioned discoveries suggested that highly expressed *PSMA7* represented a poor prognostic biomarker for most human tumors.

Abnormal *PSMA7* adversely influences genomic DNA stability

As an initiating step of tumorigenesis, genomic instability (GIN) can induce various cancer-related pathological processes such as DNA damage reaction defects, cell cycle checkpoint dysfunction, and oncogene activation⁴⁵. The GIN can be measured by two biological indicators, microsatellite instability (MSI)⁴⁶ and tumor mutational burden (TMB)⁴⁷. In particular, the TMB has become a reliable biomarker to predict the response to programmed death 1 and ligand 1 (*PD-1/PD-L1*) therapy for cancers⁴⁷. In this regard, we attempted to explore the possible link of *PSMA7* expression to TMB and MSI across TCGA tumors. Consistent with the expected result, *PSMA7* expression exhibited a positive relationship with MSI of KIRC, DLBC, TGCT, kidney renal papillary cell carcinoma (KIRP), LIHC, UCEC, and BLCA (Supplementary Fig. S7A), pointing to compromised DNA mismatch repair function in these tumors. Besides, *PSMA7* expression also harbored a positive association with TMB in most types of tumors, including SARC, PAAD, BRCA, LGG, LUAD, STAD, SKCM, ACC, OV, LIHC, LUSC, PRAD, and BLCA (Supplementary Fig. S7B). Collectively, we postulated that aberrant high expression of *PSMA7* might result in decreased genomic stability in a general way and serve as a predictive immunotherapy biomarker.

Next, we wondered why *PSMA7* alterations imposed such a great negative influence on genomic stability. Since amplification of *PSMA7* accounted for predominance in genetic alteration types, we compared the alteration frequency of genes in 276 cases with CNA-amplified *PSMA7* and that in 10,684 cases without the corresponding *PSMA7* abnormality in TCGA tumors (Supplementary Fig. S8A). Interestingly, among 373 differential genes with an alteration frequency over a 5.5 log ratio, 122 (32.71%) genes preferentially were centralized at chromosome 20 (Supplementary Fig. S8B). Further analysis revealed that 84 of 122 (68.25%) genes were mostly enriched in the region of chromosome 20q13.33 where *PSMA7* was exactly located (Supplementary Fig. S8C). The top 30 genes ranked by the frequency at 20q13.33 were visually displayed in Supplementary Fig. S8D. In contrast to genes in samples without amplified *PSMA7*, most of those in cases with *PSMA7* amplification experienced high alteration burdens, with frequencies of nearly 80%. These discoveries suggested again that 20q13.33 harboring *PSMA7* among others was plausibly a novel amplicon that frequently existed in human cancers. In addition to 20q13.33, we also found other chromosomal amplification sites, such as 20q13.32, 20q13.31, and 20q13.2 (Supplementary Fig. S8B), corresponding to the potential chr20 amplicons ranked by amplification frequency in Table 1. The results manifested a co-occurrence amplification pattern of *PSMA7* and other 20q-related genes that together contributed to GIN.

The underlying molecular mechanism of aberrant *PSMA7* in tumorigenesis

Given that the aberrant *PSMA7* expression possibly adversely influenced the clinical outcome of various cancers (Fig. 4B, C, Supplementary Fig. S5A–E), we then explored the possible molecular mechanism linking high expression of *PSMA7* to tumorigenesis. With this regard, the dependency of tumor cells on *PSMA7* was evaluated by the genome-scale CRISPR-Cas9 knockout screens in the 1095 cancer cell lines extracted from the Depmap⁴⁸ database. The *PSMA7* depletion resulted in significantly negative screen scores (concentrate range: – 2.0 to – 3.0) in the majority of types of cancers (Fig. 5A), highlighting that *PSMA7* was required for tumor cell growth as previously reported^{23,31,32}.

Furthermore, via the STRING tool, the specific interaction network between 50 *PSMA7*-binding proteins supported by experimental evidence was visualized in Fig. 5B. In parallel with this, the top 100 genes with strong correlation with *PSMA7* expression were generated by the GEPIA2 online tool. Notably, *PSMA7* expression was positively correlated with selected 18 genes (all $P \leq 0.001$) (details in Fig. 5C). In good accordance with this, validation by TIMER2 revealed the positive link of *PSMA7* to each of the aforesaid 18 genes across most TCGA

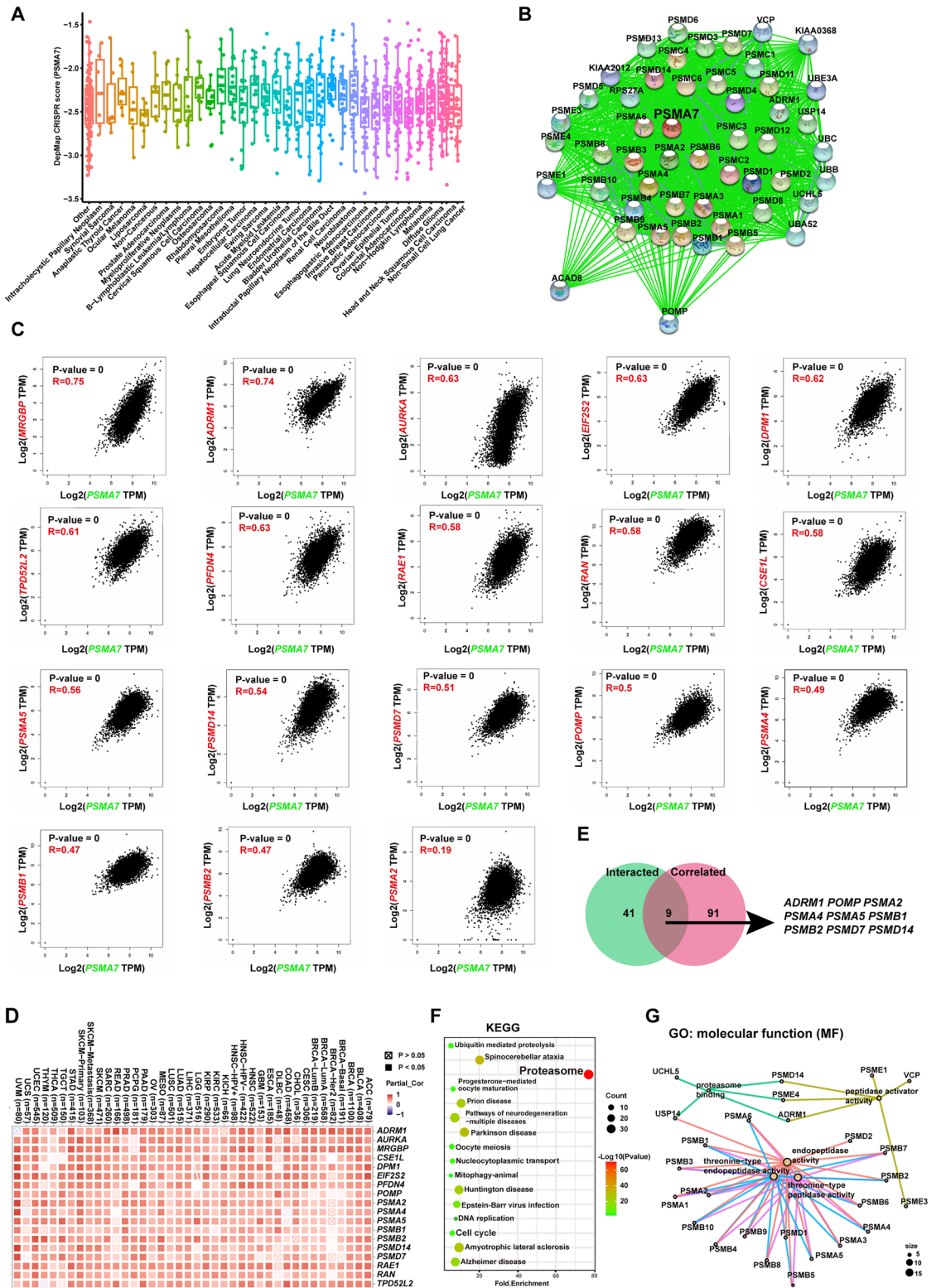


Figure 5. The underlying molecular mechanism behind the aberrant *PSMA7* in tumorigenesis. **(A)** The dependency of tumor cell viability on *PSMA7* was evaluated by the whole-wide CRISPR knockout screens in the 1095 cancer cell lines from the Depmap database. **(B)** Via STRING, we acquired proteins with experimental validation to interact with *PSMA7*. **(C)** According to the TCGA tumor dataset, we found the first 100 genes linked to *PSMA7* expression through the GEPIA2 approach. Further, the scatter plot displayed the expression correlations of *PSMA7* with 18 selected genes of *MRGBP*, *ADRM1*, *AURKA*, *EIF2S2*, *DPM1*, *TPD52L2*, *PFN4*, *RAE1*, *RAN*, *CSE1L*, *PSMA5*, *PSMD14*, *PSMD7*, *POMP*, *PSMA4*, *PSMB1*, *PSMB2*, and *PSMA2*. **(D)** The corresponding validation data across diverse TCGA cancers were generated into a heatmap by TIMER2.0. **(E)** The shared proteins between the *PSMA7*-interacting and expression-related genes were discovered by the interactive Venn diagram viewer Jvenn. **(F-G)** The KEGG and GO analysis of the aforementioned two gene sets (**B** and **C**) was conducted and presented as a bubble chart and cnetplot, respectively.

cancers (Fig. 5D). By overlapping the above two groups via the Jvenn⁴⁹, nine common genes were identified, including *ADRM1* (*ADRM1* 26S Proteasome Ubiquitin Receptor), *POMP* (Proteasome Maturation Protein), *PSMA2* (Proteasome 20S Subunit Alpha 2), *PSMA4* (Proteasome 20S Subunit Alpha 4), *PSMA5* (Proteasome 20S Subunit Alpha 5), *PSMB1* (Proteasome 20S Subunit Beta 1), *PSMB2* (Proteasome 20S Subunit Beta 2), *PSMD7* (Proteasome 26S Subunit, Non-ATPase 7) and *PSMD14* (Proteasome 26S Subunit, Non-ATPase 14) (Fig. 5E).

Next, we subjected those *PSMA7*-interacting proteins and genes whose expression harbored connections with *PSMA7* to the enrichment analysis. As foreseen, the KEGG-based findings manifested that the proteasome, cell cycle, ubiquitin-mediated proteolysis, and nucleocytoplasmic transport might participate in *PSMA7*-mediated tumorigenesis (Fig. 5F). Besides, GO data revealed that most genes were concentrated on protease activity-related pathways and cellular processes, especially proteasome binding or complex, threonine-type peptidase and endopeptidase activity, endopeptidase activity or complex, and peptidase complex, to guide the biological processes such as anaphase—promoting complex—dependent catabolism, RNA polymerase II transcriptional regulation, and control of cell cycle progression (Fig. 5G, Supplementary Fig. S9A, B). These results documented that aberrant *PSMA7* expression greatly influenced UPS status and thus mediated various downstream biological processes, eventually promoting tumor cell growth.

Candidate *PSMA7*-resistant and sensitive compounds

Elevated *PSMA7* expression was considered a prognostic molecular biomarker that might become a new treated target for different types of tumors. However, previous studies suggested the potential involvement of increased *PSMA7* protein levels in drug resistance. To uncover the drug-response landscape of cancers expressing upregulated *PSMA7*, we analyzed the potential connection of drug sensitivity with this gene mRNA expression across pan-cancers based on the drug sensitivity in cancer (GDSC) and cancer therapeutics response portal (CTRP) databases. Overall, positive correlation (R) values with $FDR \leq 1.0e-5$ were obtained in compounds of 89 (of 481) and 18 (of 251) in CTRP and GDSC, respectively, which were considered *PSMA7*-resistant agents (Fig. 6A). Figure 6B displays the top 30 ranked resistant compounds in each drug library. To further find the candidate most *PSMA7*-resistant agents, we overlapped the data and identified four shared compounds in both CTRP and GDSC (Fig. 6A, Table 2). Since *PSMA7* overexpression was validated due to amplification in clinical samples with different types of cancers, we further asked whether *PSMA7* amplification showed poor response to these resistant agents, for instance, the first-ranking compound TPCA-1⁵⁰. As expected, strong positive correlations existed between *PSMA7* expression and amplification in cell lines with TPCA-1 treatment based on the data from GDSC and CTRP⁵¹ (Fig. 6C). Using the genomic data from the catalogue of somatic mutations in cancer (COSMIC)⁵² database, we found that COSMIC-annotated *PSMA7* copy number gain and overexpression occurred in a total of nine cancer cell lines treated with TPCA-1 in GDSC and CTRP (Supplementary Tab. S14). Besides, almost all of these cell lines had high half maximal inhibitory concentrations (IC50) and poor area-under-the-curve (AUC) sensitivity scores for TPCA-1, suggesting that amplification of *PSMA7* might protect cancer cell lines from TPCA-1 treatment. These results manifested that *PSMA7*-related cancers might not benefit from these potentially resistant agents, which required much more validation in future research.

Although *PSMA7*-resistant compounds accounted for the majority, we found six compounds with negative R-values and $FDR \leq 0.05$ in GDSC that might be sensitive to cancers with elevated *PSMA7* expression (Fig. 6A, Table 2). The first-ranking agent was Docetaxel, an efficient microtubule depolymerization inhibitor⁵³ which had been approved to treat multiple types of tumors such as advanced breast cancer⁵⁴, gastric cancer⁵⁵, and melanoma⁵⁶. Consistent with the aforementioned finding, *PSMA7* expression in the cell lines treated with Docetaxel was also positively correlated with amplification in both GDSC and CTRP (Fig. 6D). Furthermore, we observed that most of the above nine cell lines with *PSMA7* copy number gain and elevated expression achieved low IC50 to Docetaxel treatment (Supplementary Tab. S14), highlighting the promising therapeutic value of this agent for *PSMA7*-related cancers. The data above generally described the drug-response pattern of abnormal *PSMA7* in cancer cell lines that need further exploration in future work.

Discussion

CNA is a common phenomenon regulating target gene expression in a wide variety of cancers^{1,2}. Recent advances manifested that 20q was widely amplified and gained in solid tumors^{3,12}. Most of the 20q CNAs and some potential cancer-driver genes conferred poor prognosis and therapeutic vulnerability by influencing immortalization, genomic instability, apoptosis, and proliferation⁵⁷. In myeloid hematological malignancies, however, deletion or loss of 20q and the involved common-deletion region (CDR) candidate genes correlated with a relatively good prognosis as a de novo chromosomal event⁵⁸. These findings emphasized the heterogeneous role of 20q CNAs in human cancers. In this study, we for the first time revealed that 20q13.33 was the most frequently amplified genomic loci at a pan-cancer level. Notably, aberrant high expression of the *PSMA7* gene at 20q13.33 due to amplification conferred poor prognosis and therapeutic vulnerability in various cancers^{12,25}. In this regard, by querying GDSC and CTRP databases, we proposed several candidate compounds, holding the promise to develop new therapeutics for relevant cancers.

In the past years, the correlation between elevated *PSMA7* expression with adverse prognosis was reported by sporadic research^{12,21–28}. Recently, a TCGA-based pan-cancer study mentioned briefly the prognostic value of 49 proteasome family member genes including *PSMA7*⁵⁹, which indeed serves as a significant precursor to our own research. During the explorations of CNA of chr20 using similar genomic data from TCGA tumors, we identified *PSMA7* as the most potential target at the newly found amplicon of 20q13.33. Then we at a pan-cancer level systematically explored the status of amplification and expression of the *PSMA7* gene and their clinical significances in terms of prognosis and potential therapeutic opportunity. Distinguishing from the previous pan-cancer research focusing on the prognostic values of *PSMA7*⁵⁹, we further investigated the underlying

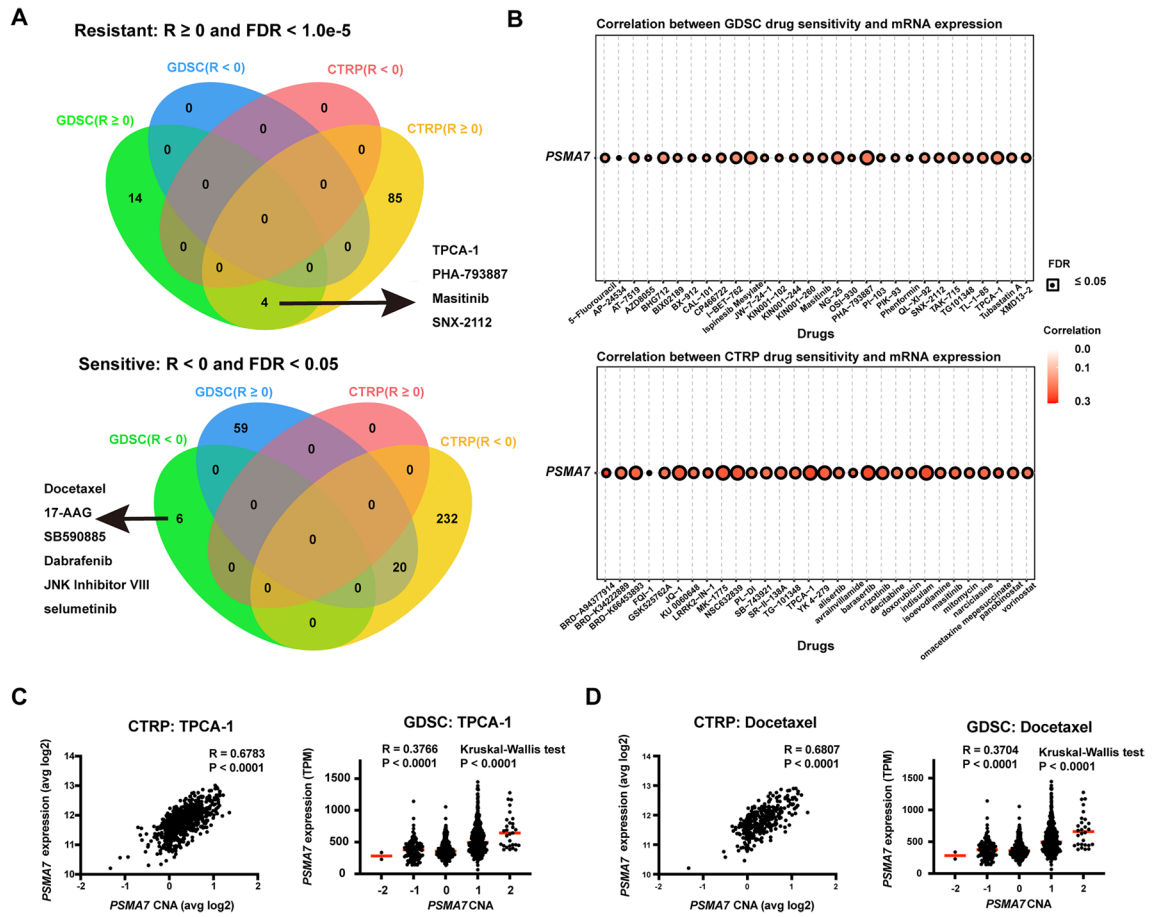


Figure 6. Candidate *PSMA7*-resistant and sensitive compounds for TCGA cancers. The GSCA tool includes genomic data of diverse TCGA cancers and compounds from the genomics of drug sensitivity in cancer (GDSC) and cancer therapeutics response Portal (CTRP) databases. Using GSCA, we screened the potential agent that strongly correlated with *PSMA7* expression. (A) An intersection analysis via Jvrenn was performed to find the potential *PSMA7*-resistant and sensitive compounds in GDSC and CTRP. (B) The top 30 ranked agents that positively correlated with *PSMA7* expression in GDSC and CTRP were also supplied, respectively. Based on the data from both drug databases, the relationship of *PSMA7* expression with amplification was evaluated in cancer cell lines treated with (C) TPCA-1 and (D) Docetaxel, respectively.

Compounds	Ranking-GDSC	R-GDSC	Ranking-CTRP	R-CTRP	R-GDSC+R-CTRP	Ranking
Resistant: $R \geq 0$ and $FDR < 1.0e-5$						
TPCA-1	3	0.18775856	2	0.260356754	0.448115314	1
PHA-793887	1	0.206831292	42	0.203181836	0.410013128	2
Masitinib	13	0.161537014	17	0.227204315	0.388741329	3
SNX-2112	12	0.161659721	77	0.177990594	0.339650315	4
Sensitive: $R < 0$ and $FDR < 0.05$						
Docetaxel	1	-0.118107832	/	/	/	/
17-AAG	2	-0.114885903	/	/	/	/
SB590885	3	-0.106534666	/	/	/	/
Dabrafenib	4	-0.09261363	/	/	/	/
JNK Inhibitor VIII	5	-0.083405322	/	/	/	/
Selumetinib	6	-0.075319267	/	/	/	/

Table 2. Candidate *PSMA7*-resistant and sensitive compounds in GDSC and CTRP databases. GDSC, genomics drug sensitivity in cancer; CTRP, cancer therapeutics response portal; R, correlation coefficient; FDR, false discovery rate; JNK, c-Jun N-terminal kinase.

mechanism linking high *PSMA7* expression to poor prognosis. Analysis of proteins interacting or genes correlating with *PSMA7* revealed broad functions of *PSMA7* in multiple cellular processes¹³, especially networks with proteasome binding or complex. Moreover, we found that *PSMA7* was significantly essential for tumor cell viability in the CRISPR-mediated loss of function screens of 1095 cancer cell lines. These findings suggested that the *PSMA7*-mediated proteasome activity might necessarily support tumor cell survival, as reported previously^{23,31,32}. Moreover, Hochstrasser et al. found that elevated cellular *PSMA7* levels increased preferentially the assembly of *PSMA7*-*PSMA7* proteasome alternative isoform and protected tumor cells from the environmental pressure induced by heavy metals²⁰. However, more experiment-based evidence is warranted to prove this hypothesis.

Notably, we discovered that *PSMA7* expression displayed a positive link to TMB and MSI in diverse tumors, figuring out the potential role of *PSMA7* in DNA damage repair for tumor cells⁴⁵. According to the two-hit hypothesis of tumorigenesis⁶⁰, amplification of 20q13.33 might serve as the first “hit”, and aberrant high expression of *PSMA7* might contribute to the second “hit” required for overt tumorigenesis by introducing more genetic lesions. Interestingly, *PSMA7* can be used as a plausible immunotherapy predictive biomarker, given its correlation with TMB⁴⁷. Besides, we further identified a cohort of other differentially overexpressed genes between samples with and without *PSMA7* amplification, revealing a co-occurrence amplification pattern of *PSMA7* and other genes across whole chromosomes, especially 20q. One noteworthy co-amplified gene was *LAMA5* located at 20q13.33, a member of the laminin family that encoded the α chain of the extracellular matrix protein laminin⁵⁶¹. This gene was found to promote migrations of some tumors^{62,63} and thus achieved a comparable score with *PSMA7* in cancer in the current study (Fig. 1D). The amplification of *LAMA5* which harbored the highest gene expression than any other CNA types was indeed observed in 23 (of 32) types of TCGA tumors in cBioPortal (Supplementary Fig. S10 A, B). However, genome-scale CRISPR-Cas9 knockout screens showed that *LAMA5* depletion resulted in nearly zero in almost all types of cancers, suggesting *LAMA5* was not required for tumor cell survival (Supplementary Fig. S10C). Therefore, we inferred that the chromosome amplicon 20q13.33 might be the work unit in which *PSMA7* but at least not *LAMA5* functioned as a cancer-promoting factor. Given its critical role in tumorigenesis, therapeutic compounds targeting the *PSMA7* pathway were screened using CTRP and GDSC databases, with six of them showing therapeutic potential. Further validation to test the anti-tumor effects of these candidate compounds will provide more evidence to support this.

Taken together, our study supported *PSMA7* as the crucial gene at the 20q amplification hotspot and a potential new biomarker for the diagnostic testing, prognostic evaluation, and molecular therapeutic development of human cancers. Identification of several chromosome 20q amplicons, especially the novel amplicon 20q13.33 consisting of the amplified *PSMA7* and other overexpressed genes, presents a more comprehensive landscape of genomic amplification of 20q, and the oncogenic role and underlying molecular mechanism of *PSMA7* at a pan-cancer level.

Methods

Genetic variation analysis

The GISTIC2.0 pipeline^{34–36} was applied to analyze the copy-number raw data from 33 types of TCGA tumors and achieve the total segmented copy-number profile and the broad and focal CNA regions across the whole genome. The results were visualized as a heatmap and detailed chromosome frequency plots. The online tool GISTIC2 from Broad Firehose³⁵ was employed to obtain statistical data for each chromosome arm and detailed amplification frequency plots across the whole genome in 30 types of TCGA tumors. The “TCGA Pan Cancer Atlas Studies” of the “Quick select” module of the cBioPortal approach^{37,38} was operated with “Explore Selected Studies” to download the data for “CNA genes” from the “Summary” module. We summarized the amplification frequency of each chr20 subband and performed a Mann–Whitney U test in which a P value less than 0.05 was considered statistically significant. The enrichment analysis of genes at the chromosome subband with the highest amplification frequency was performed by the KEGG and GO. Additionally, we employed another tool Phenolyzer³⁹ to explore the potential link of genes to cancer.

For *PSMA7*, the “Cancer Types Summary” section on this web was queried to quickly acquire the genomic variations including frequency, mutation type, and CNAs of this gene in TCGA tumors. The *PSMA7* expression level in various types of putative CNA was investigated in the “PLOT” module. Besides, the “Genomic Alterations” module in the “Comparison/Survival” section was employed to get the difference in co-occurrence pattern between other altered genes and amplified *PSMA7* in TCGA patients with or without *PSMA7* copy number amplification. We focused on the differential genes with statistically significant alteration frequencies above the 5.5 log ratio and observed their chromosomal and subband distributions.

According to the GSCA database⁴⁰, the “Copy Number Alteration (CNA)” module in the “Mutation” section was employed to reveal the genetic alteration of *PSMA7* in TCGA tumors. The CNA frequency, genotype status, and their association with survival prognosis were observed in the “CNA summary”, and “CNA & Survival” modules, respectively. The log-rank and Cox regression tests were each applied for the prognosis estimation of CNA on OS, PFS, DFI, and DSS. The multivariate Cox regression analysis was further performed to evaluate whether the amplification of *PSMA7* represented an independent factor by adjusting established clinical and molecular variables. The relationship of *PSMA7* CNA with gene expression was evaluated in the “CNA & Expression” module. In addition, the “Pan-Cancer” section of the Assistant for Clinical Bioinformatics web was employed to estimate the possible link of *PSMA7* expression to TMB and MSI.

Differentially expressed gene analysis

The *PSMA7* expression across TCGA tumors was compared with that in normal samples through the “Gene_DE” section of the TIMER2 online⁴¹. Notably, for tumors with rare numbers of normal samples, the adjacent healthy volunteer samples of the GTEx dataset were added. Then *PSMA7* expression in those tumors was reevaluated by

the “Expression Analysis-Box Plots” section of the GEPIA2⁴². Furthermore, the “Pathological Stage Plot” section was operated with *PSMA7* to investigate gene expression at each pathological stage of diverse TCGA cancers. We presented these results in violin plots.

Survival analysis

With the setting of the median expression level of *PSMA7* as the cutoff point, we operated the GEPIA2 to reveal the clinical outcome effect of gene expression on OS and DFS across TCGA cancers. The data were generated by the “Survival Map” and “Survival Analysis” modules and displayed as survival maps and log-rank survival plots, respectively. In parallel, according to the K-M plotter tool⁴⁴, the prognosis implication of *PSMA7* expression was also estimated using the log-rank test via the “mRNA genechip” module for breast, ovarian, lung, and gastric cancers (Affy ID for 201114_x_at), as well as the “mRNA RNA-seq” module for liver cancer (RNAseq ID for 5688), across the GEO cohorts. The “Auto select best cutoff” was regarded as the cutoff point. The survival indexes included OS, distant metastasis-free survival (DMFS), PFS, FP, RFS, PPS, and DSS. P values below 0.05 were recognized with statistical differences.

CRISPR-Cas9-mediated *PSMA7* loss of function analysis

The Depmap database⁴⁸ captured the data of the whole genome CRISPR knockout screens in exceeding 1000 cancer cell lines. With inputting the gene “*PSMA7*”, the “Perturbation Effects” module displayed the influence of *PSMA7* depletion on cancer cell viability using the dataset of “CRISPR (DepMap Public 23Q2+ Score, Chronos)”. The data for the DepMap CRISPR dependency scores were visualized as box plots. A lower score indicates a higher likelihood that *PSMA7* is essential in a given cell line. A score of 0 and -1 represents that *PSMA7* is dispensable or comparable to the median of all pan-essential genes, respectively.

Enrichment analysis of *PSMA7*-related genes

By inputting “*PSMA7*”, we identified the *PSMA7*-interacting proteins with experimental evidence in the organism “*Homo sapiens*” via STRING online. The major filtered conditions were listed in the following: the definition of network boundaries limited by “evidence”, active interaction sources originated from “experiments”, “Low confidence (0.150)” selected for the smallest required interplay score, and “no more than 50 interactors” in 1st shell for showing maximum interactors.

Besides, the first 100 genes whose expressions were linked to *PSMA7* were acquired in the “Similar Gene Detection” section of GEPIA2 by combining data from TCGA tumors and healthy volunteer samples. We then explored the potential link of *PSMA7* expression to these chosen genes through the Pearson correlation test. The data were shown as dot plots with log₂ TPM. In addition, via the Spearman rank correlation test, we further validated the association of these selected genes with *PSMA7* expression in the “Gene_Corr” section of TIMER2. Finally, we interpreted the result into a heat diagram.

Moreover, we overlapped the aforementioned two gene sets by Jvenn⁴⁹ to obtain the intersection genes. Subsequently, the union genes were subjected to the KEGG and GO enrichment analysis. Briefly, all genes were functionally annotated by DAVID (the database for annotation, visualization, and integrated discovery), under the filtered conditions of chosen identifier “OFFICIAL_GENE_SYMBOL” and species “*Homo sapiens*”. The final results for the KEGG enriched pathway and GO were presented as bubble charts and cnetplots with R languages, respectively. P values below 0.05 were recognized with statistical differences.

Potential *PSMA7*-resistant and sensitive drug screening

The GSCA web⁴⁰ integrates the tumor data from diverse TCGA cancers and compounds exceeding 750 from the GDSC and CTRP databases. The “Drug” module was applied to reveal the relationship of *PSMA7* expression with the IC₅₀ representing the drug susceptibility. For resistant (R-value: positive) and sensitive (R-value: negative) agent analysis, the FDR values below 1.0e-5 and 0.05 were considered statistically significant in each. Intersection analysis via the Jvenn online tool was conducted to obtain the candidate *PSMA7*-resistant and sensitive compounds in GDSC and CTRP. The common potential agents were re-ranked by the sum of the R-value in both drug libraries. Additionally, the top 30 ranked compounds positively correlated with *PSMA7* expression were visualized in GDSC and CTRP, respectively. The copy number data of cancer cell lines from COSMIC (Version 99, GRCh38)⁵² was applied to evaluate the potential association of *PSMA7* CNAs with the sensitivity of specific compounds in GDSC and CTRP. Finally, the gene expression and copy number profile data from both drug libraries⁵¹ were used to analyze the possible relationship between *PSMA7* expression and CNAs in cancer cell lines under detailed agent treatment by the correlation analysis and Kruskal-Wallis test.

Data availability

The data produced during the current research are attached in the issued main text and the additional files. All bioinformatics tools mentioned during the analysis process are public to access. Please contact shengesther@163.com if someone wants to request the data from this study.

Received: 7 August 2023; Accepted: 2 February 2024

Published online: 06 February 2024

References

1. Trost, B., Loureiro, L. O. & Scherer, S. W. Discovery of genomic variation across a generation. *Hum. Mol. Genet.* **30**, R174–R186. <https://doi.org/10.1093/hmg/ddab209> (2021).

2. Nakamura, Y. DNA variations in human and medical genetics: 25 years of my experience. *J. Hum. Genet.* **54**, 1–8. <https://doi.org/10.1038/jhg.2008.6> (2009).
3. Knuutila, S. *et al.* DNA copy number amplifications in human neoplasms: Review of comparative genomic hybridization studies. *Am. J. Pathol.* **152**, 1107–1123 (1998).
4. Huh, J. *et al.* Characterization of chromosome arm 20q abnormalities in myeloid malignancies using genome-wide single nucleotide polymorphism array analysis. *Genes Chromosomes Cancer* **49**, 390–399. <https://doi.org/10.1002/gcc.20748> (2010).
5. Israeli, O. *et al.* In silico chromosomal clustering of genes displaying altered expression patterns in ovarian cancer. *Cancer Genet. Cytogenet.* **160**, 35–42. <https://doi.org/10.1016/j.cancergencyto.2004.11.011> (2005).
6. El-Rifai, W. *et al.* Genetic differences between adenocarcinomas arising in Barrett's esophagus and gastric mucosa. *Gastroenterology* **121**, 592–598. <https://doi.org/10.1053/gast.2001.27215> (2001).
7. Schlegel, J. *et al.* Comparative genomic in situ hybridization of colon carcinomas with replication error. *Cancer Res.* **55**, 6002–6005 (1995).
8. Kallioniemi, A. *et al.* Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 2156–2160. <https://doi.org/10.1073/pnas.91.6.2156> (1994).
9. Cher, M. L. *et al.* Comparative genomic hybridization, allelic imbalance, and fluorescence in situ hybridization on chromosome 8 in prostate cancer. *Genes Chromosomes Cancer* **11**, 153–162. <https://doi.org/10.1002/gcc.2870110304> (1994).
10. Speel, E. J. *et al.* Genetic evidence for early divergence of small functioning and nonfunctioning endocrine pancreatic tumors: Gain of 9Q34 is an early event in insulinomas. *Cancer Res.* **61**, 5186–5192 (2001).
11. Knuutila, S., Autio, K. & Aalto, Y. Online access to CGH data of DNA sequence copy number changes. *Am. J. Pathol.* **157**, 689. [https://doi.org/10.1016/S0002-9440\(10\)64579-8](https://doi.org/10.1016/S0002-9440(10)64579-8) (2000).
12. Scotto, L. *et al.* Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: Potential role in progression. *Genes Chromosomes Cancer* **47**, 755–765. <https://doi.org/10.1002/gcc.20577> (2008).
13. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479. <https://doi.org/10.1146/annurev.biochem.67.1.425> (1998).
14. Iwafune, Y., Kawasaki, H. & Hirano, H. Electrophoretic analysis of phosphorylation of the yeast 20S proteasome. *Electrophoresis* **23**, 329–338. [https://doi.org/10.1002/1522-2683\(200202\)23:2%3c329::AID-ELPS329%3e3.0.CO;2-1](https://doi.org/10.1002/1522-2683(200202)23:2%3c329::AID-ELPS329%3e3.0.CO;2-1) (2002).
15. Pugh, C. W., O'Rourke, J. F., Nagao, M., Gleadle, J. M. & Ratcliffe, P. J. Activation of hypoxia-inducible factor-1; definition of regulatory domains within the alpha subunit. *J. Biol. Chem.* **272**, 11205–11214. <https://doi.org/10.1074/jbc.272.17.11205> (1997).
16. Huang, J., Kwong, J., Sun, E. C. & Liang, T. J. Proteasome complex as a potential cellular target of hepatitis B virus X protein. *J. Virol.* **70**, 5582–5591. <https://doi.org/10.1128/JVI.70.8.5582-5591.1996> (1996).
17. Dong, J., Chen, W., Welford, A. & Wandinger-Ness, A. The proteasome alpha-subunit XAPC7 interacts specifically with Rab7 and late endosomes. *J. Biol. Chem.* **279**, 21334–21342. <https://doi.org/10.1074/jbc.M401022200> (2004).
18. Tandle, A. T. *et al.* Endothelial monocyte activating polypeptide-II modulates endothelial cell responses by degrading hypoxia-inducible factor-1alpha through interaction with PSMA7, a component of the proteasome. *Exp. Cell Res.* **315**, 1850–1859. <https://doi.org/10.1016/j.yexcr.2009.03.021> (2009).
19. Kusmierczyk, A. R., Kunjappu, M. J., Funakoshi, M. & Hochstrasser, M. A multimeric assembly factor controls the formation of alternative 20S proteasomes. *Nat. Struct. Mol. Biol.* **15**, 237–244. <https://doi.org/10.1038/nsmb.1389> (2008).
20. Padmanabhan, A., Vuong, S. A. & Hochstrasser, M. Assembly of an evolutionarily conserved alternative proteasome isoform in human cells. *Cell Rep.* **14**, 2962–2974. <https://doi.org/10.1016/j.celrep.2016.02.068> (2016).
21. Xia, S. *et al.* Overexpression of PSMA7 predicts poor prognosis in patients with gastric cancer. *Oncol. Lett.* **18**, 5341–5349. <https://doi.org/10.3892/ol.2019.10879> (2019).
22. Xia, S. *et al.* Proteasome subunit alpha type 7 promotes proliferation and metastasis of gastric cancer through MAPK signaling pathway. *Dig. Dis. Sci.* **67**, 880–891. <https://doi.org/10.1007/s10620-021-06903-9> (2022).
23. Ren, C. C. *et al.* Effects of shRNA-mediated silencing of PSMA7 on cell proliferation and vascular endothelial growth factor expression via the ubiquitin-proteasome pathway in cervical cancer. *J. Cell Physiol.* **234**, 5851–5862. <https://doi.org/10.1002/jcp.26408> (2019).
24. Richardson, A. L. *et al.* X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell.* **9**, 121–132. <https://doi.org/10.1016/j.ccr.2006.01.013> (2006).
25. Hu, X. T. *et al.* The proteasome subunit PSMA7 located on the 20q13 amplicon is overexpressed and associated with liver metastasis in colorectal cancer. *Oncol. Rep.* **19**, 441–446 (2008).
26. Romanuik, T. L. *et al.* LNCaP Atlas: Gene expression associated with in vivo progression to castration-recurrent prostate cancer. *BMC Med. Genom.* **3**, 43. <https://doi.org/10.1186/1755-8794-3-43> (2010).
27. Tu, J. *et al.* Bioinformatics analysis of molecular genetic targets and key pathways for hepatocellular carcinoma. *Onco Targets Ther.* **12**, 5153–5162. <https://doi.org/10.2147/OTT.S198802> (2019).
28. Yang, B., Chen, Z., Huang, Y., Han, G. & Li, W. Identification of potential biomarkers and analysis of prognostic values in head and neck squamous cell carcinoma by bioinformatics analysis. *Onco Targets Ther.* **10**, 2315–2321. <https://doi.org/10.2147/OTT.S135514> (2017).
29. Hu, X. T. *et al.* Depletion of the proteasome subunit PSMA7 inhibits colorectal cancer cell tumorigenicity and migration. *Oncol. Rep.* **22**, 1247–1252. https://doi.org/10.3892/or_00000561 (2009).
30. Xiao, Y. *et al.* MiR-466 as a poor prognostic predictor suppresses cell proliferation and EMT in breast cancer cells by targeting PSMA7. *Eur. Rev. Med. Pharmacol. Sci.* **26**, 3799. https://doi.org/10.26355/eurrev_202206_28946 (2022).
31. Tangri, A. *et al.* Deubiquitinase UCHL1 maintains protein homeostasis through the PSMA7-APEH-proteasome axis in high-grade serous ovarian carcinoma. *Mol. Cancer Res.* **19**, 1168–1181. <https://doi.org/10.1158/1541-7786.MCR-20-0883> (2021).
32. Lin, H. K. *et al.* Proteasome activity is required for androgen receptor transcriptional activity via regulation of androgen receptor nuclear translocation and interaction with coregulators in prostate cancer cells. *J. Biol. Chem.* **277**, 36570–36576. <https://doi.org/10.1074/jbc.M204751200> (2002).
33. Yang, L. *et al.* PSMA7 directly interacts with NOD1 and regulates its function. *Cell Physiol. Biochem.* **31**, 952–959. <https://doi.org/10.1159/000350113> (2013).
34. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 20007–20012. <https://doi.org/10.1073/pnas.0710052104> (2007).
35. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, 41. <https://doi.org/10.1186/gb-2011-12-4-r41> (2011).
36. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905. <https://doi.org/10.1038/nature08822> (2010).
37. Cerami, E. *et al.* The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404. <https://doi.org/10.1158/2159-8290.CD-12-0095> (2012).
38. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1. <https://doi.org/10.1126/scisignal.2004088> (2013).
39. Yang, H., Robinson, P. N. & Wang, K. Phenolyzer: Phenotype-based prioritization of candidate genes for human diseases. *Nat. Methods.* **12**, 841–843. <https://doi.org/10.1038/nmeth.3484> (2015).

40. Liu, C. J. *et al.* GSCALite: A web server for gene set cancer analysis. *Bioinformatics* **34**, 3771–3772. <https://doi.org/10.1093/bioinformatics/bty411> (2018).
41. Li, T. *et al.* TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–W514. <https://doi.org/10.1093/nar/gkaa407> (2020).
42. Li, C., Tang, Z., Zhang, W., Ye, Z. & Liu, F. GEPIA2021: Integrating multiple deconvolution-based analysis into GEPIA. *Nucleic Acids Res.* **49**, W242–W246. <https://doi.org/10.1093/nar/gkab418> (2021).
43. Liu, J. *et al.* An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416. <https://doi.org/10.1016/j.cell.2018.02.052> (2018).
44. Lanczky, A. & Gyorffy, B. Web-based survival analysis tool tailored for medical research (KMplot): Development and implementation. *J. Med. Internet Res.* **23**, e27633. <https://doi.org/10.2196/27633> (2021).
45. Ferguson, L. R. *et al.* Genomic instability in human cancer: Molecular insights and opportunities for therapeutic attack and prevention through diet and nutrition. *Semin. Cancer Biol.* **35**(Suppl), S5–S24. <https://doi.org/10.1016/j.semcancer.2015.03.005> (2015).
46. Salmaninejad, A. *et al.* Genomic instability in cancer: Molecular mechanisms and therapeutic potentials. *Curr. Pharm. Des.* **27**, 3161–3169. <https://doi.org/10.2174/1381612827666210426100206> (2021).
47. Palmeri, M. *et al.* Real-world application of tumor mutational burden-high (TMB-high) and microsatellite instability (MSI) confirms their utility as immunotherapy biomarkers. *ESMO Open.* **7**, 100336. <https://doi.org/10.1016/j.esmoop.2021.100336> (2022).
48. Tsherniak, A. *et al.* Defining a cancer dependency map. *Cell* **170**, 564–576. <https://doi.org/10.1016/j.cell.2017.06.010> (2017).
49. Bardou, P., Mariette, J., Escudie, F., Djemiel, C. & Klopp, C. jvenn: An interactive Venn diagram viewer. *BMC Bioinform.* **15**, 293. <https://doi.org/10.1186/1471-2105-15-293> (2014).
50. Nan, J. *et al.* TPCA-1 is a direct dual inhibitor of STAT3 and NF-kappaB and regresses mutant EGFR-associated human non-small cell lung cancers. *Mol. Cancer Ther.* **13**, 617–629. <https://doi.org/10.1158/1535-7163.MCT-13-0464> (2014).
51. Rees, M. G. *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **12**, 109–116. <https://doi.org/10.1038/nchembio.1986> (2016).
52. Sondka, Z. *et al.* COSMIC: A curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res.* **52**, D1210–D1217. <https://doi.org/10.1093/nar/gkad986> (2024).
53. Wang, S. *et al.* The two novel DLL4-targeting antibody-drug conjugates MvM03 and MGD03 show potent anti-tumour activity in breast cancer xenograft models. *Cancer Lett.* **409**, 125–136. <https://doi.org/10.1016/j.canlet.2017.09.004> (2017).
54. Wang, J., Yu, Y., Lin, Q., Zhang, J. & Song, C. Efficacy and safety of first-line therapy in patients with HER2-positive advanced breast cancer: A network meta-analysis of randomized controlled trials. *J. Cancer Res. Clin. Oncol.* **150**, 21. <https://doi.org/10.1007/s00432-023-05530-3> (2024).
55. Kodaera, Y. *et al.* Addition of docetaxel to S-1 results in significantly superior 5-year survival outcomes in Stage III gastric cancer: A final report of the JACCRO GC-07 study. *Gastric Cancer* **26**, 1063–1068. <https://doi.org/10.1007/s10120-023-01419-9> (2023).
56. Mahgoub, T. M. *et al.* Evaluation of ABT-751, a novel anti-mitotic agent able to overcome multi-drug resistance, in melanoma cells. *Cancer Chemother. Pharmacol.* <https://doi.org/10.1007/s00280-023-04624-6> (2024).
57. Hodgson, J. G., Chin, K., Collins, C. & Gray, J. W. Genome amplification of chromosome 20 in breast cancer. *Breast Cancer Res. Treat.* **78**, 337–345. <https://doi.org/10.1023/a:1023085825042> (2003).
58. Braun, T. *et al.* Characteristics and outcome of myelodysplastic syndromes (MDS) with isolated 20q deletion: A report on 62 cases. *Leuk. Res.* **35**, 863–867. <https://doi.org/10.1016/j.leukres.2011.02.008> (2011).
59. Larsson, P. *et al.* Pan-cancer analysis of genomic and transcriptomic data reveals the prognostic relevance of human proteasome genes in different cancer types. *BMC Cancer* **22**, 993. <https://doi.org/10.1186/s12885-022-10079-4> (2022).
60. Knudson, A. G. Jr. Mutation and cancer: Statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 820–823. <https://doi.org/10.1073/pnas.68.4.820> (1971).
61. Rogers, H. A. *et al.* The role of the WNT/beta-catenin pathway in central nervous system primitive neuroectodermal tumours (CNS PNETs). *Br. J. Cancer* **108**, 2130–2141. <https://doi.org/10.1038/bjc.2013.170> (2013).
62. Diao, B., Sun, C., Yu, P., Zhao, Z. & Yang, P. LAMA5 promotes cell proliferation and migration in ovarian cancer by activating Notch signaling pathway. *FASEB J.* **37**, e23109. <https://doi.org/10.1096/fj.202300306R> (2023).
63. Qin, Y. *et al.* Laminin 521 enhances self-renewal via STAT3 activation and promotes tumor progression in colorectal cancer. *Cancer Lett.* **476**, 161–169. <https://doi.org/10.1016/j.canlet.2020.02.026> (2020).

Acknowledgements

Here we gratefully thanked Dr. Lou Jiacheng (Blogger, WeChat official accounts: Chris_Life_Science) provided extremely constructive suggestions for manuscript revisions. This research was funded by the National Key R&D Program of China (2023YFA1800401 and 2019YFA0905902) and the National Natural Science Foundation of China (82350710226, 82370178 and 81890994).

Author contributions

G.Y.S. raised and designed the research idea with K.K.W.'s guidance. G.Y.S. and F.Y.L. did the bioinformatic analysis by online tool databases and R language, respectively. The bioinformatics data was interpreted by G.Y.S. and F.Y.L., G.Y.S. drafted the manuscript. The manuscript was revised and proofread by K.K.W. and W.J. And the final text was read and allowed by the aforementioned authors to submit for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-53585-0>.

Correspondence and requests for materials should be addressed to K.W.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2024