



Research Article

Copyright© Xudong Fu

Pan-Cancer Analysis of UBX Domain-Containing Protein 4 (UBXN4) in Human Cancers

Yongjie Han, Zhou Jing, Zian Li, Bing Guo, Ziyang Jia, Qingyi Wang, Minghe Li and Xudong Fu*

Department of Neurosurgery, International Joint Laboratory of Glioma Metabolism and Microenvironment Research, The Fifth Affiliated Hospital of Zhengzhou University, China

*Corresponding author: Xudong Fu, Department of Neurosurgery, The Fifth Affiliated Hospital of Zhengzhou University, No.3 Kangfu-qian Street, Zhengzhou, Henan Province, 450052, China.

To Cite This Article: Yongjie Han, Zhou Jing, Zian Li, Bing Guo, Ziyang Jia, Qingyi Wang, Minghe Li and Xudong Fu*. TPan-Cancer Analysis of UBX Domain-Containing Protein 4 (UBXN4) in Human Cancers. *Am J Biomed Sci & Res.* 2023 18(1) AJBSR.MS.ID.002431, DOI: [10.34297/AJBSR.2023.18.002431](https://doi.org/10.34297/AJBSR.2023.18.002431)

Received: February 07, 2023; **Published:** February 22, 2023

Abstract

Background: UBX domain-containing protein 4 (UBXN4) was discovered for the first time because its UBX domain can interact with p97/VCP to participate in endoplasmic reticulum-associated protein degradation (ERAD). Although there are many reports that the UBX domain-containing protein family is involved in the occurrence and development of tumors, a pan-cancer analysis of UBXN4 is not yet available.

Material and Methods: We first explored the potential carcinogenic effects of UBXN4 in 33 tumors based on the datasets of TCGA (The Cancer Genome Atlas) and GEO (Gene Expression Omnibus).

Results: UBXN4 is highly expressed in many cancers and distinct associations exist between UBXN4 expression and tumor prognosis. We also observed that the promoter methylation level of UBXN4 was significantly reduced in most primary tumors compared to that in normal tissues, whereas the opposite was observed in kidney renal clear cell carcinoma (KIRC). UBXN4 expression correlates with cancer-related fibroblast infiltration in head and neck squamous cell carcinoma (HNSC), skin cutaneous melanoma (SKCM), and stomach adenocarcinoma (STAD). Glycoprotein endoplasmic reticulum-associated degradation and calnexin/calreticulin cycle function are involved in the functional mechanisms of UBXN4.

Conclusions: As the first study to perform pan-cancer analysis of UBXN4, the results of this study will improve the understanding of the carcinogenic role of UBXN4 in different tumors.

Keywords: UBXN4; Cancer; Prognosis; Methylation; Immune Infiltration

Introduction

Pan-cancer analysis is designed to examine the similarities and differences between genome and cell changes found in different tumor types. Pan-cancer expression analysis of a target gene is helpful for evaluating its relevance to clinical prognosis and potential molecular mechanisms [1]. The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) are currently the two most famous public databases that contain a variety of tumor functional genomics datasets and allow us to conduct pan-cancer analysis [2,3].

The UBX protein superfamily is comprised of a diverse group of UBX (ubiquitin-regulatory X) domain-containing proteins in mammalian cells. Members of this family contain a UBX domain that is generally located at the carboxy-terminus of the protein [4]. The UBX domain, with approximately 80 residues, has a three-dimensional structure that is similar to that of ubiquitin. In contrast to ubiquitin, there is no double glycine motif or suitably positioned lysine side chain in the UBX domain structure; therefore, the UBX domain cannot be conjugated to other proteins or become part of the mixed UBX-ubiquitin chain [5,6]. Therefore, all members of the UBX family can bind to the multifunctional AAA-ATPase p97/

VCP protein through the amino-terminal domain of p97, thereby exerting a variety of cellular functions including membrane fusion, protein degradation, and autophagy [7]. P97 (also known as the valosin-containing protein), a molecular chaperone of AAA-ATPase, is important in the endoplasmic reticulum-associated degradation (ERAD) pathway [8].

Here, we characterized a human protein containing the UBX domain, UBXN4 (also called UBXD2 or Erasin), which is highly conserved in mammals and the UBXN4 gene is localized on chromosome 2q21.3, according to the complete human genome sequence (Gene ID:23190, HGNC:14860). Biochemical fractionation, immunofluorescence, electron microscopy, and protease protection experiments show that UBXN4 is an integral membrane protein of the endoplasmic reticulum and nuclear envelope, with its N- and C-termini facing the cytoplasm or nucleoplasm, and immunoprecipitation and GST-pulldown experiments confirmed that UBXN4 binds p97/VCP through its UBX domain [9]. Upon binding to p97, which is associated with the lumen of the endoplasmic reticulum (ER), UBXN4 becomes a critical cofactor of the ER-associated degradation pathway [10,11].

Endoplasmic reticulum (ER) stress is a phenomenon in which ER function is disrupted by stimuli such as hypoxia or nutrient deficiency, which often disrupts the cellular microenvironment and leads to the accumulation of improperly folded proteins in the ER [12]. Although tumor cells are frequently exposed to ER stress microenvironments (such as chronic ischemia and hypoxia), they can maintain normal physiological functions and survival. Because most tumor cells have evolved a greater ability to cope with ER stress, they can restore ER homeostasis by facilitating the elimination of incorrectly folded proteins, thereby creating favorable conditions for their survival [13,14]. Normally, ER stress can initiate ERAD, which helps tumor cells eliminate misfolded proteins in the ER and maintain survival. Therefore, ERAD plays an important role in tumorigenesis [15].

UBXN4 overexpression was shown to enhance the degradation of classical ERAD substrates, whereas siRNA-mediated reduction in UBXN4 expression levels almost completely blocked ERAD and UBXN4 protein expression levels increased when cells underwent an endoplasmic reticulum stress response [9]. Although the involvement of UBX domain-containing protein families in tumorigenesis and development has been reported in the literature [16-19], the role of UBXN4 in tumors has rarely been reported, and related pan-cancer analyses are lacking. Considering the important role of ERAD in tumor development, and the fact that UBXN4 proteins positively regulate the ERAD pathway, we hypothesized that UBXN4 could be a new target for tumor therapy.

In this study, for the first time, we used the TCGA and GEO databases to conduct a pan-cancer analysis of UBXN4. We also incorporated a group of factors, such as gene expression, survival status, DNA methylation, genetic alterations, immune infiltration, and related cellular pathways to explore the potential molecular

mechanisms of UBXN4 in the pathogenesis and clinical prognosis of different cancers.

Materials and Methods

Gene expression analysis

We entered UBXN4 into the “Gene_DE” module of TIMER2 (Tumor Immune Estimation Resource, Version 2) website (<http://timer.comp-genomics.org/>) [20] and observed UBXN4 expression differences between tumors and adjacent normal tissues for different tumors or specific tumor subtypes in the TCGA project. For certain tumors with no or few normal tissue samples [for example, TCGA-ACC (adrenocortical carcinoma), TCGA-PAAD (pancreatic adenocarcinoma), etc.], we used the “Expression analysis-Box Plots” module of the GEPIA2 (Gene Expression Profiling Interactive Analysis, Version 2) website (<http://gepia2.cancer-pku.cn/#analysis>) [21] to obtain box plots of expression differences between these tumors and the corresponding normal tissues in the Genotype-Tissue Expression (GTEx) database, under the settings of P-value cutoff = 0.01, Log2 fold change cutoff = 1, and “Match TCGA normal and the GTEx data”. In addition, by using the “Pathological Stage Plot” module of GEPIA2, we obtained violin plots of UBXN4 expression in all TCGA tumors at different pathological stages (stage I, II, III, IV). Box plots and violin plots are represented in log2 [TPM (Transcripts Per Million) + 1] transformed expression data.

UALCAN is an interactive web portal (<http://ualcan.path.uab.edu/analysis-prot.html>) [22] based on PERL-CGI, JavaScript, and CSS for analyzing cancer omics data, including the Clinical Proteomic Tumor Analysis Consortium (CPTAC) and TCGA (The Cancer Genome Atlas) databases. Herein, we explored the total protein levels of UBXN4 in primary tumors and normal tissues using the CPTAC database, which has six tumor datasets: breast cancer, ovarian cancer, colon cancer, clear cell renal cell carcinoma, uterine corpus endometrial carcinoma (UCEC), and lung adenocarcinoma (LUAD). Subsequently, we revealed the differential expression of UBXN4 promoter methylation in primary tumors and normal tissues across the TCGA database.

Survival Prognostic Analysis

Using the Kaplan-Meier “survival map module in GEPIA2, the Overall Survival (OS) and disease-free survival (DFS) significance map data of UBXN4 were obtained for all TCGA tumors by setting the cutoff high (50%) and cutoff low (50%) values as the expression thresholds to segment UBXN4 high or low expression, and the log-rank test as the hypothesis test, respectively. We then obtained survival plots with the same settings using the Kaplan-Meier Survival Analysis” module of GEPIA2. In addition, we extracted TCGA data and plotted ROC (Receiver Operating Characteristics) curves using the “Survival ROC” software package, with sensitivity (true positive rate) as the vertical axis and 1-specificity (false positive rate) as the abscissa, and the larger the area under the curve (AUC), the higher the prognostic accuracy [23].

Genetic alteration analysis

On the cBioPortal website (<http://www.cbioportal.org/>) [24,25], we selected the “TCGA Pan-Cancer Atlas Studies” in the “Quick select” section and entered “UBXN4” for queries of the genetic variant characteristics of UBXN4. The results of alteration frequency, mutation type, structural variation, and copy number alteration (CNA) for all TCGA tumors can be observed in the “Cancer Types Summary” module. In the “Mutations” module, we can also observe detailed information about the UBXN4 mutation sites, which are displayed in the schematic diagram of the protein structure or the 3D (Three-dimensional) structure. Apart from that, we used the “Comparison/Survival” module to obtain the data on overall survival (OS), disease-specific survival, progression-free survival (PFS), and disease-free survival (DFS) differences with and without UBXN4 genetic alteration in the TCGA database and generated Kaplan-Meier plots with log-rank P-value.

The UniProt (Universal Protein) website (<https://www.uniprot.org/>) [26], a protein database with the most information and resources, allowed us to obtain an identity document for the UBXN4 protein [UniProtKB ID: Q92575 (UBXN4_HUMAN)]. SMART (Simple Modular Architecture Research Tool (SMART)) is a web resource (<http://smart.embl-heidelberg.de/>) [27] for the identification and annotation of protein domains and analysis of protein domain architectures. Finally, we input the ID of the UBXN4 protein into the SMART website to obtain a schematic diagram of the UBXN4 protein domain.

Immune infiltration analysis

We used the “Immune-Gene” module of the TIMER2 website to explore the relationship between UBXN4 expression and immune infiltration in TCGA tumors. Immune cells of cancer-associated fibroblasts were selected, and the EPIC, MCP-COUNTER, XCELL, and TIDE algorithms were used to estimate immune infiltration. P-values and partial correlation (cor) values were obtained using the Spearman’s rank correlation test with purity adjustment. We used these data to construct scatter plots and heatmaps. Thereafter we extracted RNA-seq data and clinical information in the level 3 HTSeq-FPKM format from the TCGA database, and used the Gene Set Variation Analysis (GSVA) package of R software to analyze the correlation between the UBXN4 gene and various immune cells and to make up a “lollipop” chart. We analyzed the following cell types: T helper (Th) cells, T central memory (Tcm) cells, Th17 cells, macrophages, T gamma delta (Tgd) cells, aDC (activated Dendritic Cells), neutrophils, T effector memory (Tem) cells, B cells, mast cells, Th2 cells, Eosinophils, T cells, T follicular helper (TFH) cells, Th1 cells, iDC (immature Dendritic Cells), Natural Killer (NK) cells, DC (Dendritic Cells), T Regulatory (TReg) cells, NK CD56bright cells, CD8+ T cells, pDC (plasmacytoid Dendritic Cells), cytotoxic cells, and NK CD56dim cells. P-values and partial correlation (cor) values were obtained using the Spearman’s rank correlation test.

UBXN4-related gene enrichment analysis

STRING (search tool for the retrieval of interacting genes/proteins) is an online database (<https://string-db.org/>) [28] designed to integrate all known and predicted associations between proteins, including physical interactions and functional associations. We selected the organism (“Homo sapiens”) and the single protein name (“UBXN4”) for the query on the STRING website. The following parameters were used: network type (“full STRING network”), meaning of network edges (“evidence”), active interaction sources (“experiments”), minimum required interaction score [“low confidence (0.150)”] and max number of interactors to show (“no more than 50 interactors in the first shell”). Finally, we screened available experimentally validated UBXN4-binding proteins.

We used the “Similar Genes Detection” module of GEPIA2 to obtain the top 100 UBXN4-correlated target genes based on the datasets of all TCGA tumors and normal tissues. Subsequently, we performed Pearson Correlation Analysis of the screened genes with UBXN4 respectively in the “Correlation Analysis” module of GEPIA2. Dot plots are presented as log2 TPM-transformed expression data with P-value and correlation coefficients (R). Finally, we used the “Gene_Corr” module of TIMER2 to provide heatmap data for the selected genes, which contained the partial correlation (cor) and P-value in Spearman’s rank correlation test after purity adjustment.

We used Venn diagram software (<http://bioinformatics.psb.ugent.be/webtools/Venn/>), a tool that visually depicts the merges, intersections, and differences between multiple datasets, to compare the UBXN4-binding and interacting genes. Moreover, we combined the two sets of data and uploaded the gene lists to the Metascape website (<http://metascape.org/gp/index.html#/main/step1>) [29] with species settings (“Homo sapiens (142)”) and analysis type (“Express Analysis”) to obtain the enriched pathway data.

Results

Gene Expression Analysis Data

In this study, we aimed to explore the oncogenic role of human UBXN4 (NM_014607.4 for mRNA or NP_055422.1 for protein). We first analyzed the expression status of UBXN4 in different types of TCGA tumors using the TIMER2 website. As shown in Figure 1A, the expression levels of UBXN4 in the tumor tissues of cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), HNSC (Head and Neck squamous cell carcinoma (HNSC), KICH (kidney chromophobe (KICH), LIHC (liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), LUSC (lung squamous cell carcinoma (LUSC), SKCM (Skin Cutaneous Melanoma (SKCM), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA) (P<0.001), BLCA (Bladder Urothelial Carcinoma (BLCA), and BRCA (breast invasive carcinoma (BRCA) (P<0.05) were higher than those in the corresponding control tissues.

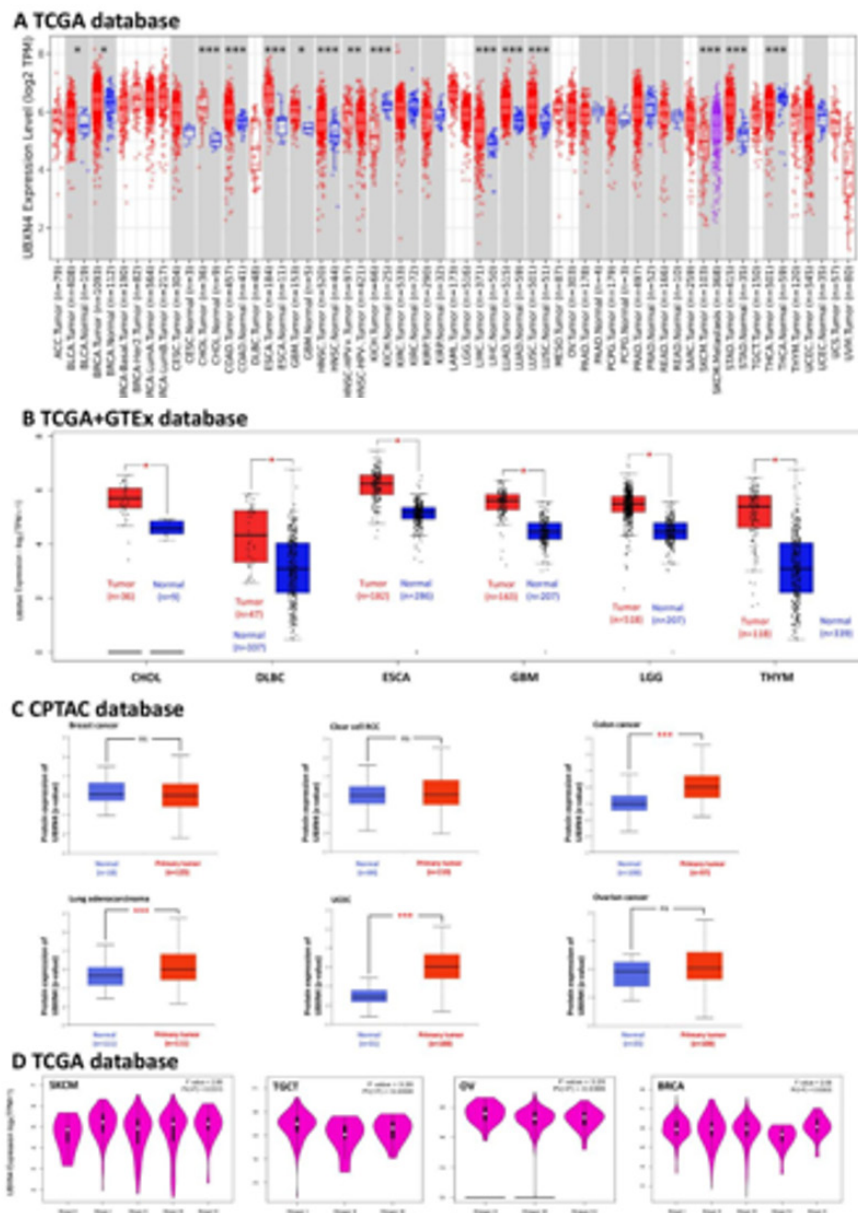


Figure 1: Expression level of UBXM4 gene in different tumors and pathological stages.

(A) Analysis of UBXM4 gene expression in different cancers or specific cancer subtypes by TIMER2 (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

(B) The types of CHOL, DLBC, ESCA, GBM, LGG and THYM in the TCGA project are shown with the corresponding normal tissues in the GTEx database as controls. The box plot data were provided (* $P < 0.05$).

(C) Based on the CPTAC database, we also analyzed the expression levels of UBXM4 total protein between primary tumors and normal tissues in breast cancer, clear cell RCC, colon cancer, lung adenocarcinoma, uterine corpus endometrial carcinoma (UCEC) and ovarian cancer (*** $P < 0.001$; ns: no statistical significance).

(D) Based on the TCGA data, the expression levels of UBXM4 gene were analyzed by the main pathological stages (stages I, II, III, IV) of SKCM, TGCT, OV and BRCA. Log2 (TPM+1) was applied for logarithmic scale.

Subsequently, considering the small number of normal tissue samples from some TCGA tumors [for example, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC) and thymoma (THYM)], we included normal tissues from the GTEx database as controls and further evaluated the expression difference of UBXM4 between tumors and normal tissues in CHOL, DLBC, ESCA, glioblastoma multiforme (GBM), brain lower-grade glioma (LGG), and THYM (Figure 1B) $P < 0.05$. For other tumors such as ACC (Adrenocortical carcinoma), CESC (Cervical squamous cell

carcinoma and endocervical adenocarcinoma), PAAD (Pancreatic adenocarcinoma), TGCT (Testicular Germ Cell Tumors), OV (Ovarian serous cystadenocarcinoma), we did not get significant differences.

We used the CPTAC database to analyze the differences in the expression of the UBXM4 protein in primary tumors and corresponding normal tissues. We found higher expression of UBXM4 total protein in the primary tumors of colon cancer, lung adenocarcinoma (LUAD), and uterine corpus endometrial

carcinoma (UCEC) than in normal tissues (Figure 1C) $P < 0.001$, whereas in breast cancer, ovarian cancer, and clear cell renal cell carcinoma, the difference was not statistically significant compared to that in normal tissues.

We also observed a correlation between UBXN4 expression and the pathological stages of cancers through the "Pathological Stage Plot" module of GEPIA2, including SKCM, TGCT, OV, and BRCA (Figure 1D) $P < 0.05$, but not others.

Survival Analysis Data

Based on the expression level of UBXN4, we divided cancer cases into high- and low-expression groups and investigated the

correlation of UBXN4 expression with prognosis in different tumor patients, mainly using TCGA database. As shown in Figure 2A, low expression of UBXN4 in the TCGA project was associated with poor Overall Survival (OS) prognosis for kidney renal clear cell carcinoma (KIRC) ($P = 0.026$), whereas high expression of UBXN4 was associated with poor OS prognosis for LAML ($P = 0.042$). Disease-free survival (DFS) analysis (Figure 2B) showed a correlation between high UBXN4 expression and poor prognosis in TCGA cases of ACC ($P = 0.0073$) and LGG ($P = 0.032$). As the Figure 2C shows that the AUC (Area under the curve) values for CHOL (AUC=0.941), STAD (AUC=0.907), LAML (AUC=0.906), LUAD (AUC=0.895), KICH (AUC=0.893) and LUSC (0.866) all exceeded 0.85, indicating that UBXN4 is a highly reliable clinical prognostic predictive molecule.

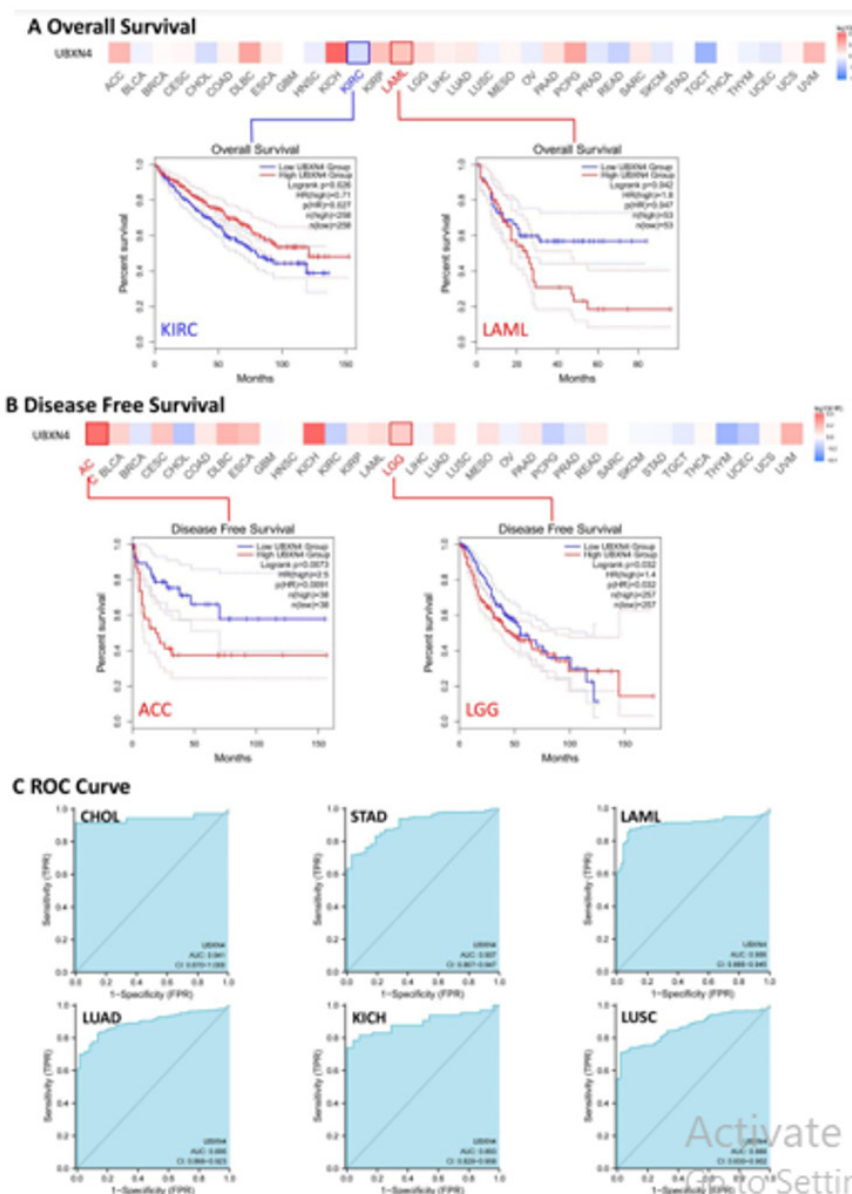


Figure 2: Research on the correlation between UBXN4 gene expression and survival prognosis of cancers in TCGA. We used the GEPIA2 tool to analyze overall survival (A) and disease-free survival (B) of different tumors in TCGA using UBXN4 gene expression. The survival map and Kaplan-Meier curves with positive results are presented. (C) The receiver operating characteristic (ROC) curve between UBXN4 and tumor prognosis was plotted based on TCGA data. The areas under the ROC curves for CHOL, STAD, LAML, LUAD, KICH, and LUSC were 0.941, 0.907, 0.906, 0.895, 0.893, and 0.866, respectively, indicating a high predictive value for tumor prognosis.

Genetic Alteration Analysis Data

We observed the genetic alteration status of UBXN4 in different tumor samples from TCGA project. As displayed in Fig. 3A, the highest alteration frequency of UBXN4 (>4%) was in patients with prostate adenocarcinoma with “Deep Deletion” as the primary CNA (Copy number alterations) type. The “Mutation” of CNA was the predominant type in the uterine corpus endometrial carcinoma, lung squamous cell carcinoma and colorectal adenocarcinoma cases, which show an alteration of frequency of 2.46%, 2.05% and 2.02%, respectively, while the main type of genetic alteration in sarcoma cases is the CNA “Amplification,” with a frequency of 1.18% (Figure 3A). It is worth noting that all skin cutaneous melanoma (2.93% frequency), cholangiocarcinoma (2.78% frequency), and

cervical squamous cell carcinoma (2.02% frequency) cases with genetic alteration had “Mutation” of UBXN4, and in contrast all mesothelioma (1.15% frequency) cases with genetic alteration were specified as CNA “Deep Deletion” (Figure 3A).

The types, sites, and case numbers of the UBXN4 genetic alterations are presented in detail in Figure 3B. We found that missense mutation of UBXN4 was the main type of genetic alteration, and the S361L alteration in the UBX domain, which was detected in two cases of UCEC and 1 case of HNSC (Figure 3B), was able to induce a frameshift mutation in the UBXN4 gene, translation from S (serine) to L (leucine) at the 361 site of the UBXN4 protein, and the subsequent UBXN4 protein truncation. We observed the S361 site in the 3D structure of UBXN4 protein (Figure 3C).

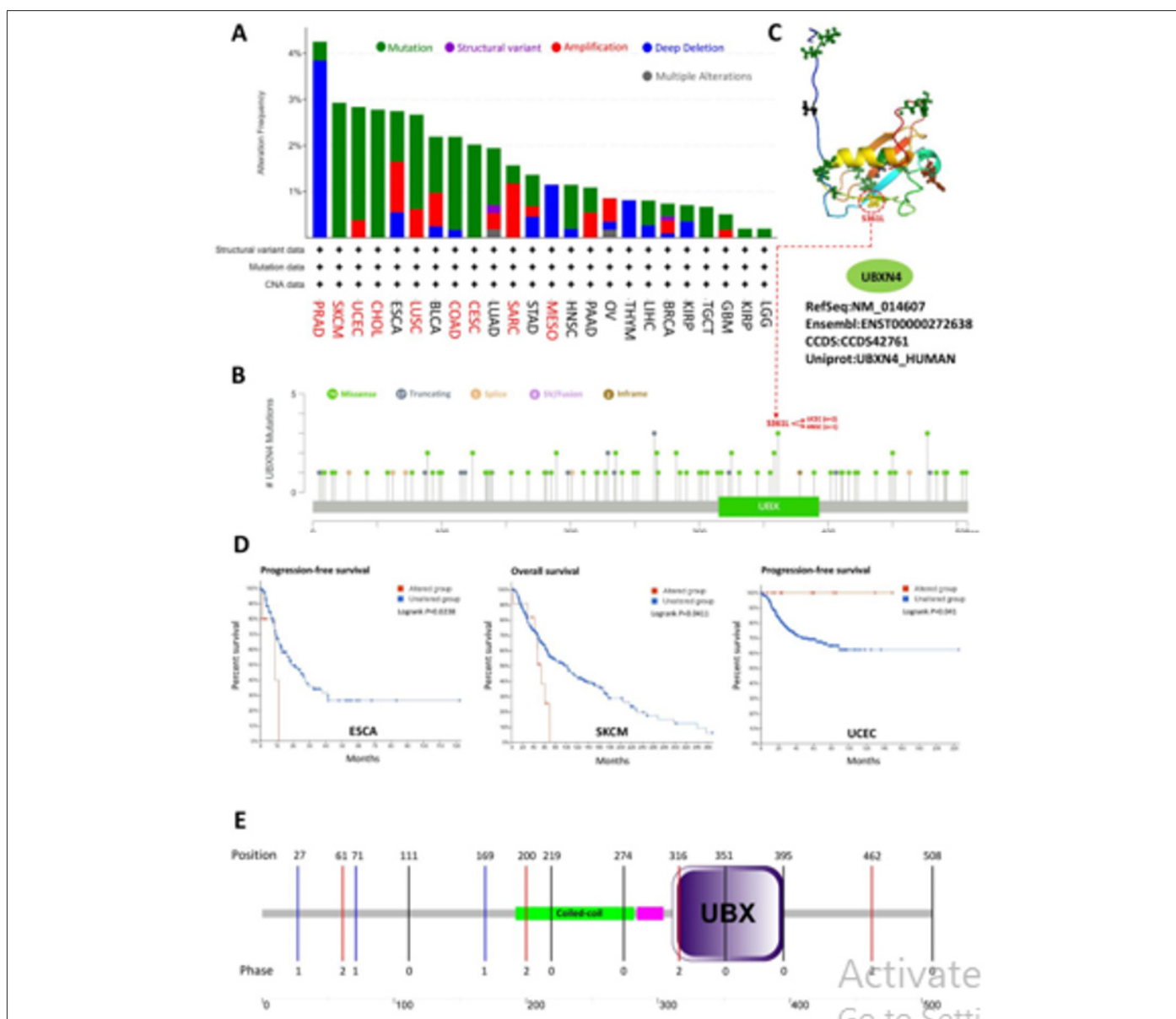


Figure 3: Mutational characteristics of UBXN4 in different tumors of TCGA.

We analyzed the mutational signature of UBXN4 in TCGA tumors using the cBioPortal tool. The frequency of alterations in the mutation type (A) and site (B) is shown.

(C) The 3D structure of UBXN4 shows the mutation site with the highest frequency of alteration (S361L).

(D) We also revealed a potential association between mutational status and ESCA, SKCM, and UCEC survival curves, using the cBioPortal tool.

(E) Schematic diagram of the structural prediction of the UBXN4 protein obtained using SMART.

In addition, we explored the potential association between the clinical survival prognosis of patients with different types of cancer and genetic alterations in UBXN4. The data in Figure 3D indicate that ESCA cases with unaltered UBXN4 showed better prognosis in progression-free survival ($P=0.0238$) and SKCM cases without UBXN4 alteration had a better prognosis in overall survival ($P=0.0411$) than those with UBXN4 alteration. However, patients with UCEC with UBXN4 alterations had a better prognosis in progression-free survival ($P=0.041$) than patients without alterations. This result warrants further investigation.

Figure 3E shows the predicted structural domains of UBXN4. We also observed that the UBXN4 protein contains two structural domains, namely the UBX domain (amino acids:316-395) and a coiled-coil domain (amino acids:192-282), where the UBX structural domain plays a major role in binding the p97/VCP

(valosin-containing protein) to participate in the endoplasmic reticulum-associated protein degradation (ERAD) process.

DNA Methylation Analysis Data

In TCGA project, we used the UALCAN web to observe differences in UBXN4 gene promoter methylation levels between different primary tumors and matched normal tissues. As shown in Figure 4, the expression levels of promoter methylation of UBXN4 were lower in BLCA, COAD, HNSC, KIRP, LIHC, LUAD, LUSC, PRAD, TGCT, UCEC ($P<0.001$), and THCA ($P<0.05$) than in the corresponding normal tissues, whereas in KIRC cases ($P<0.001$), UBXN4 promoter methylation levels were higher than those in the normal tissues. Owing to the lack of UBXN4 expression data, we did not analyze the relationship between UBXN4 promoter methylation and UBXN4 expression.

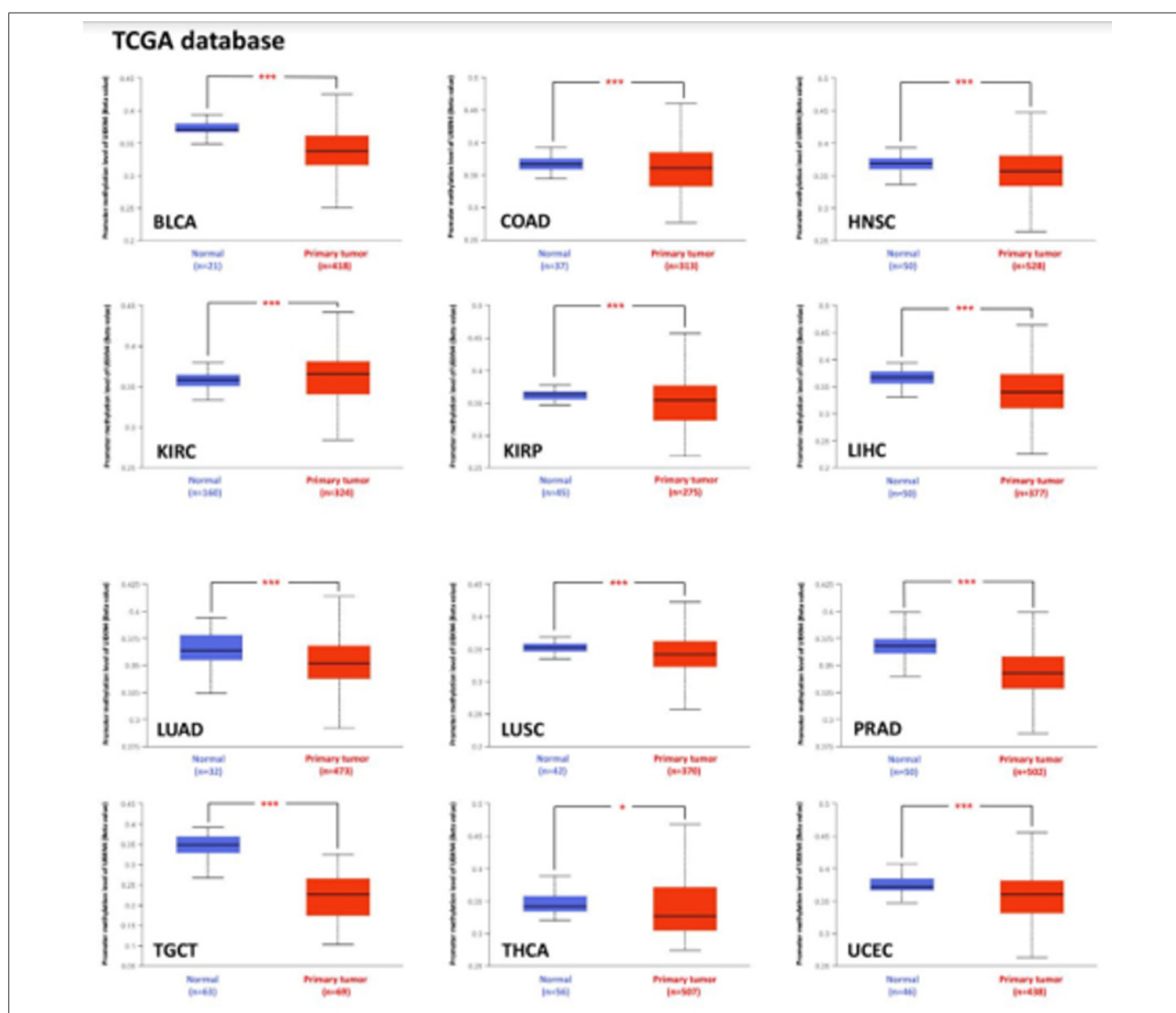


Figure 4: Analysis of UBXN4 gene promoter methylation in different tumors.

Based on TCGA database, we analyzed the expression levels of UBXN4 gene promoter methylation in different tumors and normal tissues using UALCAN. Box plots of different cancers, including BLCA, COAD, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, TGCT, THCA, and UCEC (* $P<0.05$; *** $P<0.001$).

Immune Infiltration Analysis Data

As an important component of the tumor microenvironment, tumor-infiltrating immune cells are closely associated with the initiation, progression, and metastasis of cancer [30,31]. Cancer-associated fibroblasts in the tumor microenvironment stroma have been reported to be involved in regulating the function of various tumor-infiltrating immune cells [32,33]. In this study, we used the EPIC, MCPOUNTER, XCELL, and TIDE algorithms to investigate the potential relationship between the infiltration level of cancer-associated fibroblasts and UBXN4 gene expression

in diverse cancer types in TCGA. Analysis performed using all or most of the selected algorithms revealed that cancer-associated fibroblast infiltration was significantly positively correlated with UBXN4 expression in HNSC, HNSC-[HPV (human papillomavirus)], KIRP, LGG, LIHC, PAAD, SKCM, and SKCM-metastasis but negatively correlated with STAD and TGCT (Figure 5). Scatterplot data of the tumors presented above were obtained using a single algorithm (Figure 5B). For example, the UBXN4 expression level in STAD was negatively correlated with the infiltration level of cancer-associated fibroblasts (Fig. 5B, $Rho=-0.303$, $P=1.65e-09$) based on the XCELL algorithm.

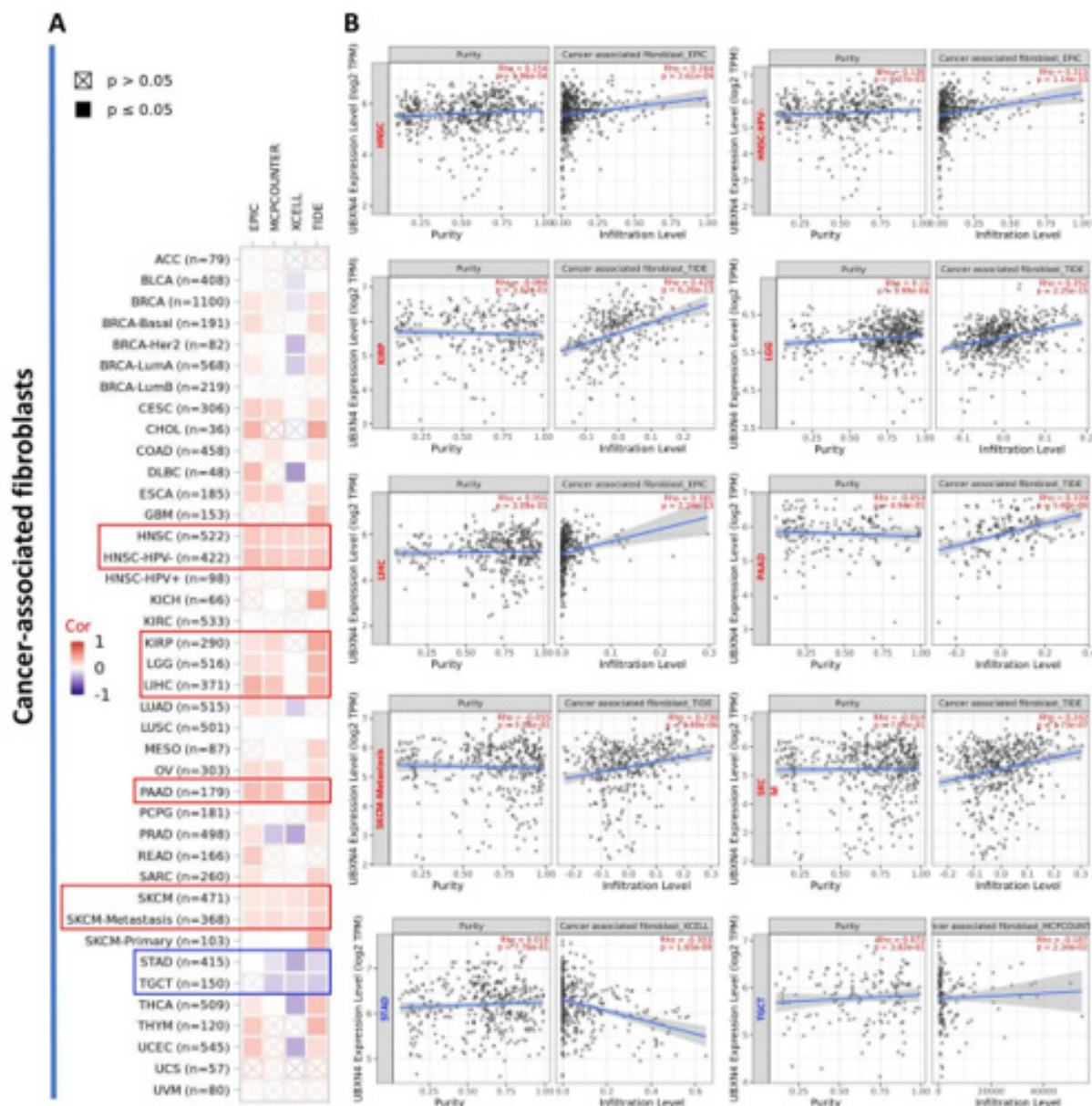


Figure 5: Correlation analysis between UBXN4 expression and immune infiltration of cancer-associated fibroblasts.

Different algorithms were used to explore the potential correlation between UBXN4 gene expression levels and the infiltration level of cancer-associated fibroblast in the TCGA project.



We also analyzed the potential relationship between other immune cells in diverse cancer types in TCGA and UBXN4 gene expression. As shown in Figure 6A & 6B, UBXN4 was closely related to multiple infiltrating immune cells in different tumor microenvironments, suggesting that UBXN4 could influence tumor progression, prognosis, and treatment. UBXN4 was positively correlated with T helper (Th) and T central memory (Tcm) cells,

which were mainly enriched sources of DLBC, KICH, KIRP, LAML, PRAD, READ, SKCM, TGCT, THCA, THYM, UCEC, and UVM in TCGA project. Cytotoxic cells and Th1 cells were mainly enriched in ACC and negatively correlated with the expression of UBXN4. Moreover, UBXN4 was negatively correlated with pDC (plasmacytoid Dendritic Cells) in PCPG, PRAD, THYM, and NK CD56bright cells in LGG and TGCT.

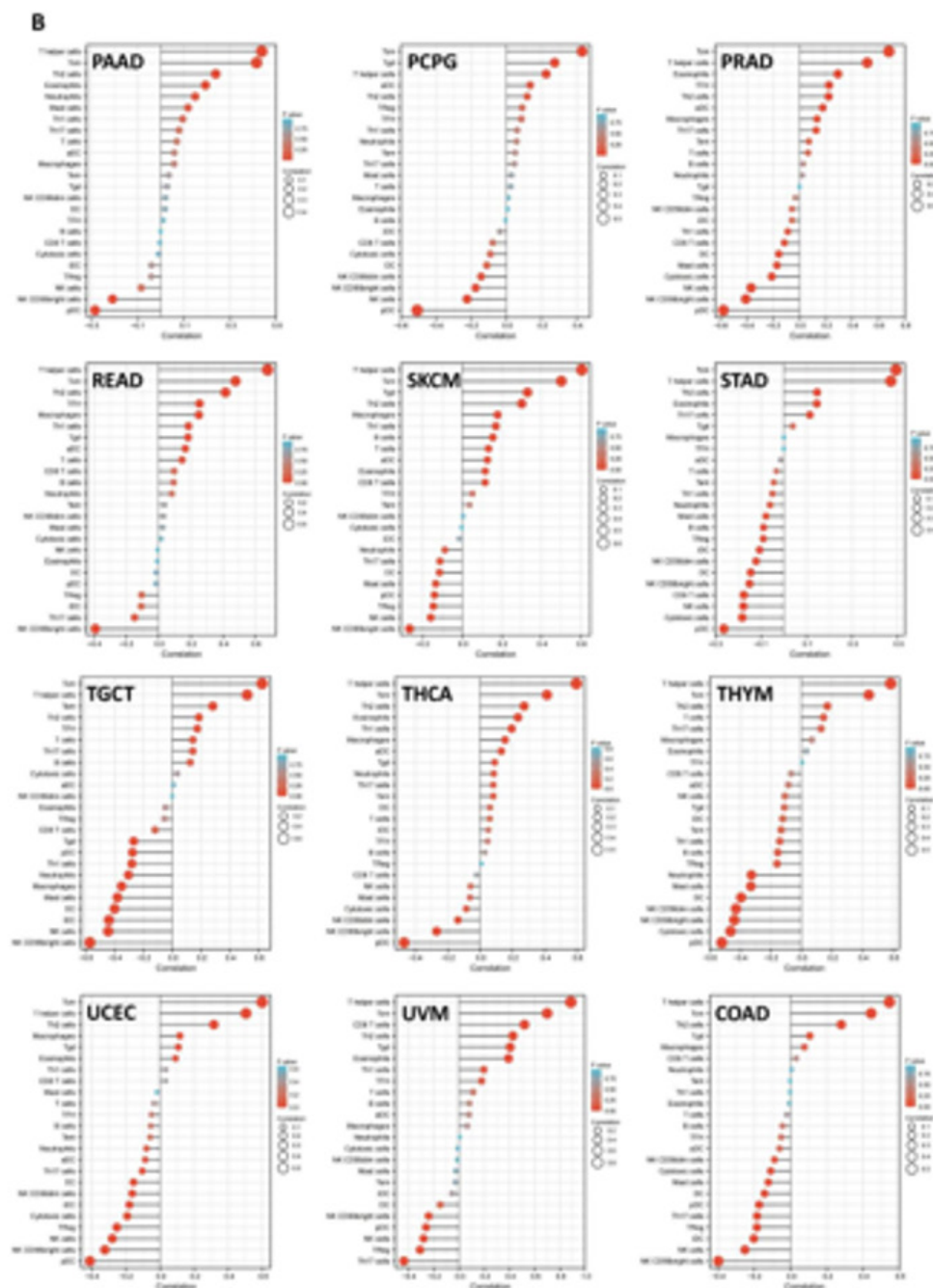


Figure 6: Gene Set Variation Analysis (GSVA) of UBXN4 enrichment and immune cells in different tumor types of TCGA. RNA-seq data and clinical information in the level 3 HTSeq-FPKM format were extracted from TCGA database and analyzed for correlation between the UBXN4 gene and various immune cells using the GSVA package from R software. P-values and partial correlation (cor) values were obtained using the Spearman's rank correlation test.

(A) Enrichment analysis of the immune cell relationship between UBXN4 and ACC, BLCA, BRCA, CHOL, DLBC, GBM, KICH, KIRC, KIRP, LAML, LGG, or LIHC.

(B) Enrichment analysis of immune cell relationships between UBXN4 and PAAD, PCPG, PRAD, READ, SKCM, STAD, TGCT, THCA, THYM, UCEC, UVM, and COAD.

UBXN4-Correlated Protein Enrichment Analysis Data

To further investigate the molecular mechanism of UBXN4 in tumorigenesis, we screened for UBXN4-binding proteins and UBXN4 expression-related genes using a series of pathway enrichment analyses. We obtained 45 UBXN4-binding proteins using the STRING tool, which was supported by experimental evidence. The interaction network between these proteins is shown in Figure 7A. We then used the GEPIA2 tool to combine all tumor expression data from TCGA database to reveal the top 100 genes associated with UBXN4 expression. Figure 7B demonstrates

a positive correlation between UBXN4 expression levels and the following genes (all $P < 0.001$): STAM2 [Signal Transducing Adaptor Molecule 2, $R = 0.78$], SP3 [Specificity Protein 3, $R = 0.76$], SMEK2 [SMEK homolog 2, $R = 0.76$], PUM2 [Pumilio RNA Binding Family Member 2, $R = 0.75$], PPIG [Peptidylprolyl Isomerase G, $R = 0.72$], RBM27 [RNA Binding Motif Protein 27, $R = 0.72$], SCYL2 [SCY1 Like Pseudokinase 2, $R = 0.71$], and ARID4B [AT-Rich Interaction Domain 4 B, $R = 0.71$]. The corresponding heatmap data also showed that UBXN4 was positively associated with these eight genes in most cancer types (Figure 7C). Intersection analysis of these two groups revealed one common member, C1orf27 (Figure 7D).

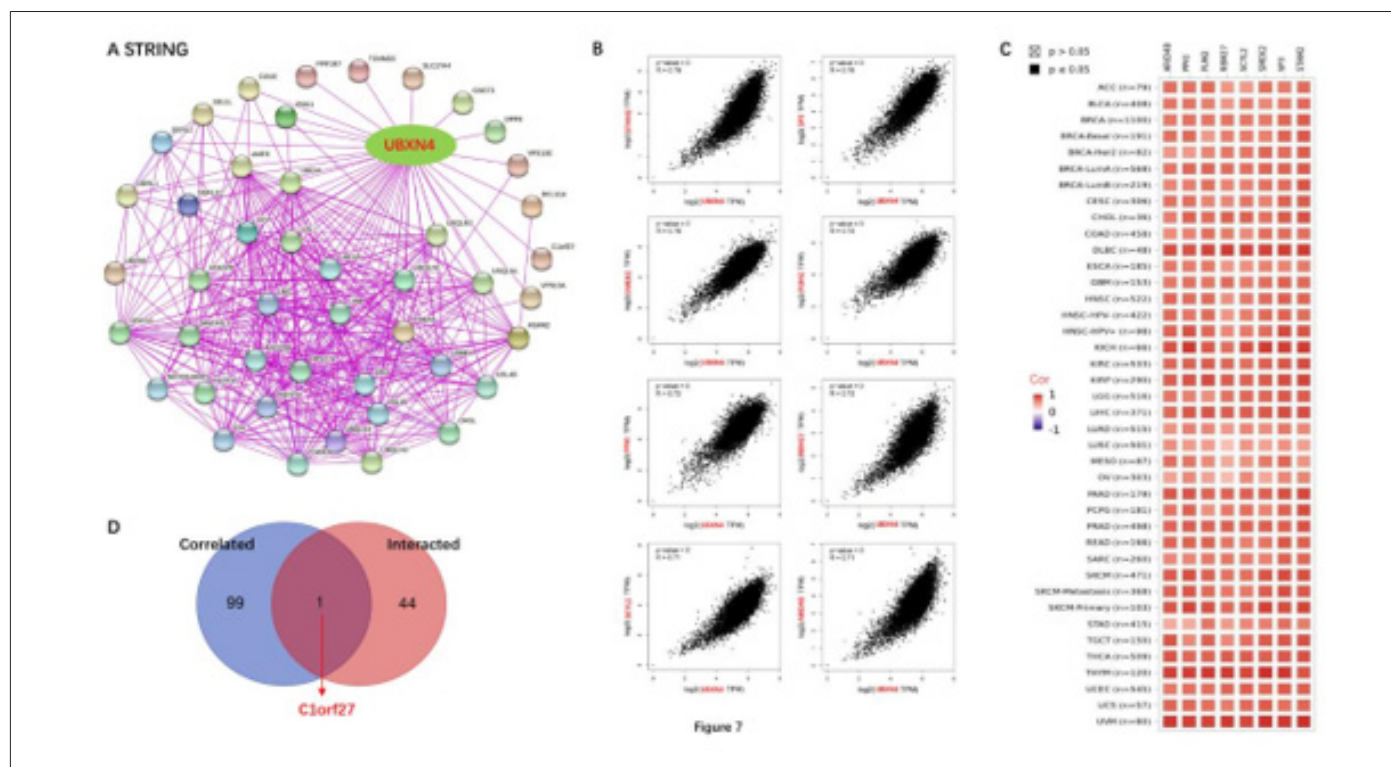


Figure 7: Enrichment analysis of UBXN4-related genes and proteins.

(A) We first obtained the available experimentally determined UBXN4-binding proteins using the STRING tool. (B) Utilizing the GEPIA2 approach, we also obtained the top 100 UBXN4-related genes in the TCGA project and analyzed the expression correlation between UBXN4 and selected targeting genes, including STAM2, SP3, SMEK2, PUM2, PPIG, RBM27, SCYL2 and ARID4B.

(C) The corresponding heatmap data in the detailed cancer types were displayed.

(D) An intersection analysis of the UBXN4-binding and correlated genes was conducted.

In addition, we integrated these two datasets into Metascape and performed Gene Ontology (GO) enrichment analysis. The data in Figure 8A suggest that UBXN4 may be associated with cellular protein catabolic processes during tumor development, such as endoplasmic reticulum protein processing, protein hydrolysis,

and macroautophagy. The analysis data also indicated that most of these genes are associated with the regulation of cell biological behaviors, such as chromosome organization, gene silencing, and mRNA processing, and might also be associated with Wnt and NF- κ B signaling pathways.

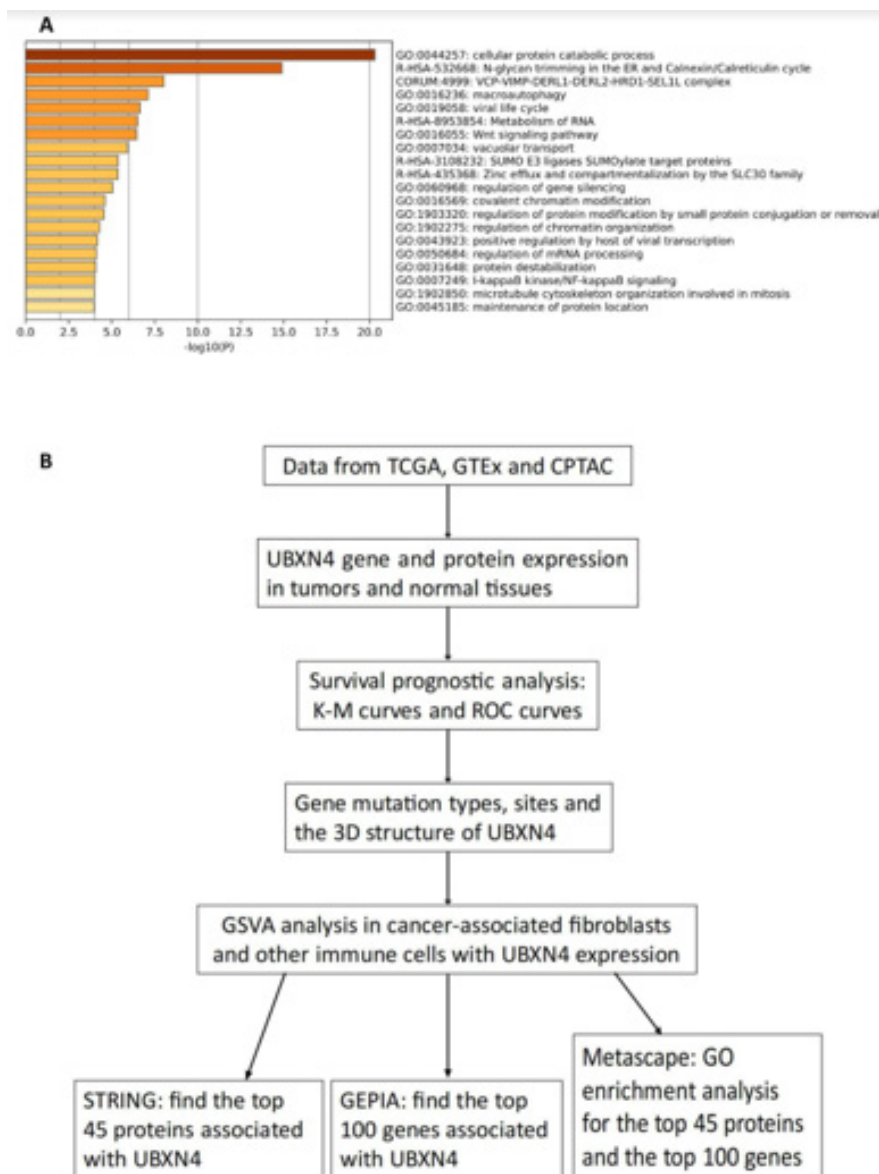


Figure 8: Gene Ontology (GO) enrichment analysis using Metascape platform.

(A) Based on the UBXN4-binding proteins and interacting genes, GO enrichment analysis was performed using the Metascape platform.

(B) A working model. Data on UBXN4 were extracted from TCGA, GEO and CPTAC for analysis of gene expression, survival status, genetic alterations, immune infiltration and related cellular pathways.

Discussion

The plasticity of cancer cells relies heavily on glycoproteins that traverse secretory pathways such as cell surface receptors and signaling molecules released in the extracellular medium. These secreted glycoproteins respond to and steer changes in the surroundings of cancer cells and contribute to tumor immunity [34], tumor growth, cancer cell division, adhesion, and metastasis. The endoplasmic reticulum is an essential organelle in eukaryotic cells that synthesizes membrane-piercing proteins that are necessary for biological survival. Protein synthesis requires proper folding and assembly in the endoplasmic reticulum, whereas proteins that are misfolded or not properly assembled are left in the endoplasmic reticulum (ER) to degrade through modifications such as ubiquitination [14]. Multiprotein complexes in the ER can identify, remove, ubiquitinate, and deliver misfolded proteins to

the 26S proteasome for degradation in the cytosol, collectively termed ER-associated degradation (ERAD) [35]. UBXN4 is a highly conserved class of proteins that is widely present in various mammalian cells and tissues. According to biochemistry, cell biology, and immunology techniques, UBXN4 is a protein localized in the endoplasmic reticulum of cells, associated with the nuclear membrane, and involved in the degradation of endoplasmic reticulum-associated proteins through its UBX structural domain binding to the p97/VCP protein [9]. Although the involvement of UBX domain-containing protein families in tumorigenesis and development has been reported in the literature [16-19], the role of UBXN4 in tumors has rarely been reported, and related pan-cancer analyses are lacking. Thus, we performed a comprehensive assay of the UBXN4 gene in 33 different tumors based on data from TCGA, CPTAC, and GEO databases, as well as molecular characterization of

gene expression, genetic alterations, DNA methylation, and immune infiltration.

UBXN4 is highly expressed in most tumors. Nevertheless, data from the survival prognostic analysis of the UBXN4 gene suggest different conclusions for different tumors. In this study, we used the GEPIA2 tool to detect the potential relationship between high UBXN4 expression and poor overall survival prognosis. Different data-processing methods or updated survival information may have contributed to this result. We then extracted TCGA data and plotted ROC curves using the Kaplan-Meier "Survival ROC" software package and concluded that UBXN4 was a reliable predictor of clinical prognosis.

Our analysis based on TCGA database revealed that UBXN4 expression was significantly higher in CHOL tumor tissues than in normal tissues, and the ROC curve showed a high area under the curve (AUC=0.941). However, high expression of UBXN4 did not seem to be associated with the clinical prognosis of CHOL patients, which might be related to the low number of CHOL cases (less than 20) with high or low UBXN4 expression. An analysis of a larger sample size may lead to more accurate conclusions. For TCGA-LAML, we found that high UBXN4 expression was associated with poor overall survival (OS) prognosis in patients, and the ROC curve from LAML indicated an AUC of 0.906, implying that UBXN4 was reliable as a prognostic indicator for LAML. In addition, we observed no difference in survival prognosis between the high and low UBXN4 expression groups in patients with UCEC tumors, but progression-free survival was significantly higher in the group with altered UBXN4 gene expression than in the group without alteration. Hence, further molecular experimental evidence is required to determine whether UBXN4 expression plays an essential role in the development of these tumors, or whether it is only the result of antitumor transformation in normal tissues.

According to DNA methylation data analysis, the expression level of the UBXN4 gene promoter methylation in BLCA, COAD, HNSC, LIHC, LUAD, and LUSC was lower than that in the corresponding normal tissues, whereas the expression level of UBXN4 in these tumors was higher than that in the normal tissues, suggesting that UBXN4 gene promoter methylation silences gene expression. In KIRP, PRAD, and UCEC tumors, even though the promoter methylation level of the UBXN4 gene was lower than that in normal tissues, the UBXN4 expression level did not differ between tumor and normal tissues; therefore, we speculated that there might be other factors influencing the expression of UBXN4.

In this study, we applied multiple immune deconvolution methods to observe a statistically positive correlation between UBXN4 expression and the immune infiltration level of cancer-associated fibroblasts in HNSC, HNSC-HPV-, KIRP, LGG, LIHC, PAAD, SKCM, and SKCM-metastasis, whereas the opposite was observed for STAD and TGCT. In most tumors, T helper and Tcm cells were positively enriched and correlated with UBXN4 expression. In contrast, cytotoxic and Th1 cells were predominantly enriched in ACC and negatively correlated with UBXN4 expression. UBXN4 is closely related to a variety of infiltrating immune cells in different

tumor microenvironments, suggesting that UBXN4 may influence tumor progression, prognosis and treatment. Additionally, we incorporated information on UBXN4 binding components and UBXN4 expression-related genes in all tumors and performed a series of enrichment analyses to identify the potential impact of "endoplasmic reticulum protein processing" and "RNA metabolism" on the etiology or pathogenesis of cancer. UBXN4 is a key cofactor in the degradation pathway of endoplasmic reticulum-related proteins [9].

In summary, our first pan-cancer analysis of UBXN4 demonstrated that UBXN4 expression was statistically correlated with clinical prognosis, DNA methylation, immune infiltration, and related cellular pathways. This study contributes to the understanding of the role of UBXN4 in tumorigenesis from the perspective of clinical tumor samples and provides informative implications for future experimental studies.

Conclusion

This is the first study to systematically assess the potential role of UBXN4 in disease progression and prognosis of several cancers. Based on TCGA, GTEx, and CPTAC databases, our analysis revealed that UBXN4 is highly expressed in most tumors and is associated with the pathological staging of certain tumors. UBXN4 may regulate the clinical survival prognosis of tumors through genetic alterations and immune cell infiltration. Gene Ontology (GO) enrichment analysis revealed that UBXN4 might be associated with endoplasmic reticulum protein processing and RNA metabolism and might also be involved in the regulation of Wnt and NF- κ B signaling pathways. Although few studies on UBXN4 in different cancers have been reported, the results of pan-cancer analysis suggest that UBXN4, a key cofactor in the endoplasmic reticulum-associated protein degradation pathway, could be a new target for tumor therapy.

Research Highlights

A first pan-cancer analysis of UBXN4

UBXN4 is associated with different prognosis in different tumor cases.

The association between UBXN4 and cancer-associated fibroblast infiltration.

UBXN4 is a critical cofactor in the endoplasmic reticulum-associated degradation pathway.

Disclosure

No potential conflicts are disclosed in this work.

Statement

This study did not include human subjects, human data, tissues, or animals.

Funding

This study was sponsored by the Key Research Projects of Henan Higher Education Institutions, China (grant number 17A310031).

References

1. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020) Pan-cancer analysis of whole genomes. *Nature*. 578(7793): 82-93.
2. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Gordon B Mills, Kenna R Mills Shaw, et al. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45(10): 1113-1120.
3. Clough E, Barrett T (2016) The Gene Expression Omnibus Database. *Methods Mol Biol* 1418: 93-110.
4. Buchberger A (2002) From UBA to UBX: new words in the ubiquitin vocabulary. *Trends Cell Biol* 12(5): 216-221.
5. Buchberger A, Howard MJ, Proctor M, Bycroft M. (2001) The UB domain: a widespread ubiquitin-like module. *J Mol Biol* 307(1): 17-24.
6. Alexandru G (2012) Exploring the role of p97 and its UB domain cofactors through identification of their interacting proteins. *Methods Mol Biol* 832: 305-312.
7. Yi L, Kaler SG. Interaction between the AAA ATPase p97/VCP and a concealed UB domain in the copper transporter ATP7A is associated with motor neuron degeneration. *J Biol Chem*. 2018;293(20):7606-7617.
8. Hurn DM, Kornfilt DJ, Wipf P. p97: An Emerging Target for Cancer, Neurodegenerative Diseases, and Viral Infections. *J Med Chem*. 2020;63(5):1892-1907.
9. Liang J, Yin C, Doong H, Shengyun Fang, Corrine Peterhoff, et al. (2006) Characterization of erasin (UBXD2): a new ER protein that promotes ER-associated protein degradation. *J Cell Sci* 119(Pt 19): 4011-4024.
10. Wolf LM, Lambert AM, Haenlin J, Boutros M (2021) EVI/WLS function is regulated by ubiquitination and linked to ER-associated degradation by ERLIN2. *J Cell Sci* 134(16): jcs.257790.
11. Meyer H, Weihl CC (2014) The VCP/p97 system at a glance: connecting cellular function to disease pathogenesis. *J Cell Sci* 127(Pt 18): 3877-3883.
12. Dicks N, Gutierrez K, Michalak M, Bordignon V, Agellon LB (2015) Endoplasmic reticulum stress, genome damage, and cancer. *Front Oncol* 5: 11.
13. Oakes SA, Papa FR (2015) The role of endoplasmic reticulum stress in human pathology. *Annu Rev Pathol* 10: 173-194.
14. Oakes SA (2020) Endoplasmic Reticulum Stress Signaling in Cancer Cells. *Am J Pathol* 190(5): 934-946.
15. Urrea H, Dufey E, Avril T, Chevet E, Hetz C (2016) Endoplasmic Reticulum Stress and the Hallmarks of Cancer. *Trends Cancer* 2(5):252-262.
16. Rezvani K (2016) UBXD Proteins: A Family of Proteins with Diverse Functions in Cancer. *Int J Mol Sci* 17(10): 1724.
17. Sane S, Abdullah A, Boudreau DA, R K Autenried, B K Gupta, et al. (2014) Ubiquitin-like (UBX)-domain-containing protein, UBXN2A, promotes cell death by interfering with the p53-Mortalin interactions in colon cancer cells. *Cell Death Dis* 5(3): e1118.
18. Huang K, Yang C, Wang QX, Yan-Sheng Li 1, Chuan Fang, et al. (2017) The CRISPR/Cas9 system targeting EGFR exon 17 abrogates NF- κ B activation via epigenetic modulation of UBXN1 in EGFRwt/vIII glioma cells. *Cancer Lett* 388: 269-280.
19. Chai RC, Chang YZ, Chang X, Bo Pang, Song Yuan An, et al. (2021) YTHDF2 facilitates UBXN1 mRNA decay by recognizing METTL3-mediated m6A modification to activate NF- κ B and promote the malignant progression of glioma. *J Hematol Oncol* 14(1): 109.
20. Li T, Fu J, Zeng Z, David Cohen, Jing Li, et al. (2020) TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res* 48(W1): W509-W514.
21. Tang Z, Kang B, Li C, Chen T, Zhang Z (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 47(W1): W556-W560.
22. Chandrashekar DS, Bashel B, Balasubramanya SAH, Chad J Creighton, Israel Ponce-Rodriguez, et al. (2017) UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 19(8): 649-658.
23. Hoo ZH, Candlish J, Teare D (2017) What is an ROC curve?. *Emerg Med J* 34(6): 357-359.
24. Gao J, Aksoy BA, Dogrusoz U, Gideon Dresdner, Benjamin Gross, et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6(269): pl1.
25. Cerami E, Gao J, Dogrusoz U, Benjamin E Gross, Selcuk Onur Sumer, et al. (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data [published correction appears in *Cancer Discov*. *Cancer Discov* 2(5): 401-404.
26. UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49(D1): D480-D489.
27. Letunic I, Khedkar S, Bork P (2021) SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res* 49(D1): D458-D460.
28. Szklarczyk D, Gable AL, Nastou KC, David Lyon, Rebecca Kirsch, et al. (2021) The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 49(D1): D605-D612.
29. Zhou Y, Zhou B, Pache L, Max Chang, Alireza Hadj Khodabakhshi, et al. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10(1): 1523.
30. DeBerardinis RJ (2020) Tumor Microenvironment, Metabolism, and Immunotherapy. *N Engl J Med* 382(9): 869-871.
31. Kaymak I, Williams KS, Cantor JR, Jones RG (2021) Immunometabolic Interplay in the Tumor Microenvironment. *Cancer Cell* 39(1): 28-37.
32. Kato T, Noma K, Ohara T, Hajime Kashima, Yuki Katsura, et al. (2018) Cancer-Associated Fibroblasts Affect Intratumoral CD8+ and FoxP3+ T Cells Via IL6 in the Tumor Microenvironment. *Clin Cancer Res* 24(19): 4820-4833.
33. De Jaeghere EA, Denys HG, De Wever O (2019) Fibroblasts Fuel Immune Escape in the Tumor Microenvironment. *Trends Cancer* 5(11): 704-723.
34. Eggleton P, Michalak M, Bremer E (2015) Editorial: Endoplasmic Reticulum and Its Role in Tumor Immunity. *Front Oncol* 5: 252.
35. Needham PG, Guerriero CJ, Brodsky JL (2019) Chaperoning Endoplasmic Reticulum-Associated Degradation (ERAD) and Protein Conformational Diseases. *Cold Spring Harb Perspect Biol* 11(8): a033928.