Comprehensive Pan-Cancer Analysis on CBX3 as a Prognostic and Immunological Biomarker

Hongjuan Niu  
Minzu University of China

Peiqiong Chen  
Zhengzhou Ninth People's Hospital

Lu Fan  
Minzu University of China

Boyu Sun (nhj19901113@163.com)  
Qingdao Third People's Hospital Affiliated to Qingdao University

Research Article

Keywords: CBX3, pan-cancer, expression, prognosis, immune

Posted Date: September 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-737601/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: Increased evidence supports the relationship between chromobox protein homolog 3 (CBX3) and tumorigenesis of some cancers. However, the role of CBX3 in pan-cancers remains poorly defined. In the research, we aimed to investigate the prognostic value and the immunological functions of CBX3.

Results: We explored the potential oncogenic roles of CBX3 in mRNA and protein levels based on the diverse databases, including the expression, the correlation with prognosis, tumor microenvironment (TME), DNA methylation, protein phosphorylation and enrichment analysis across all TCGA tumors. The results show that CBX3 is overexpressed in multiple cancers, and significant correlations exist between high expression and adverse prognosis in most tumor patients. We observed an enhanced phosphorylation level in uterine corpus endometrial carcinoma, colon cancer and lung adenocarcinoma. A distinct relationship was also found between CBX3 expression and TME, including immune infiltration of tumor-infiltrating lymphocytes (TILs) and cancer-associated fibroblasts (CAFs), immune score or matrix score, immune checkpoints. The correlative transcription factors and miRNAs of CBX3-binding hub genes were analyzed to investigate the molecular mechanism. Moreover, alcoholism and alteration of DNA cellular biology may be involved in the functional mechanisms of CBX3.

Conclusion: The first pan-cancer study offers a relatively comprehensive cognition on the oncogenic roles of CBX3 as a prognostic and immunological marker in various malignant tumors.

Introduction

The tumorigenesis is a complex process involving immune system, where TME plays a crucial role. Due to the elusive regulation mechanism, the interaction between tumor and immune remains a focus. It is urgent to conduct a pan-cancer analysis of genes to assess its value for clinical prognosis and as potential therapeutic targets. Currently, the expression pattern, survival analysis and tumor immune infiltration of genes in different tumors across numerous databases, including available TCGA project, GEO database, cBioPortal, GEPIA and TIMER database [1-3], can allow us to conduct pan-cancer analysis.

CBX3 is a member of the heterochromatin protein 1 (HP1) family. The protein encoded by this gene binds DNA and composes heterochromatin. CBX3 has several synonyms including HECH, HP1γ, heterochromatin protein 1 homolog gamma, heterochromatin-like protein 1 and modifier 2 protein. Moreover, together with CBX1 and CBX5, the three belong to HP1 orthologs. It is believed that CBX3 plays an important role for transcriptional repression or activation, epigenetic modification, as well as cell differentiation [4, 5]. Additionally, it has been reported to be abnormally expressed and dysregulated in many cancers. Currently, the expression of CBX3 was found increased in the breast cancer [6], prostate cancer [7], colorectal cancer [8] and non-small cell lung cancer [9], which may predict the poor prognosis of patients. Some literatures reported that CBX3 was related to immunity in certain cancers. On this account, CBX3 may potentially be of great importance in epigenetic regulation of cancer development, and is expected to become a therapeutic target for cancers. Nowadays, CBX3 has obtained widespread
attention as a potential biomarker in several cancers [10]. However, the expression, function and mechanism of CBX3 in various tumors still remain unknown.

In this study, we conducted a pan-cancer analysis of CBX3 based on multiple databases. A group of factors, such as gene expression and alteration, clinical prognosis, protein phosphorylation, TME and relevant cellular pathway, were involved. Based on the results, we aimed to investigate the value as biomarker and to evaluate the potential molecular mechanism of CBX3 in different cancers.

**Methods**

**Gene expression analysis**

In the “Gene_DE” module of TIMER2.0 (tumor immune estimation resource, version 2) ([http://timer.cistrome.org/](http://timer.cistrome.org/)) [11], CBX3 was searched and observed the expression discrepancy between tumor and adjacent normal tissues for the various tumors in the TCGA project. The ONCOMINE database ([https://www.ONCOMINE.org/](https://www.ONCOMINE.org/)) [12] integrates RNA and DNA-sequence data from GEO, TCGA and published literature. We entered CBX3 in the search box, then set the P value, fold change, gene rank and data type to get the differential expression data of CBX3 in various tumors. The red indicates the overexpression, and the blue is the opposite. Besides, the dark color and the large value were representative of the high expression amount. The CCLE (Cancer Cell Line Encyclopedia) ([https://portals.broadinstitute.org/ccle](https://portals.broadinstitute.org/ccle)) database can provide the information of a gene in tumor cell lines, such as gene expression, mutation, copy number and DNA methylation. We input the CBX3, then the box diagrams were classified and colored by the average distribution of gene expression in different cell lines. GEPIA2 (Gene Expression Profiling Interactive Analysis, version 2) web server ([http://gepia2.cancer-pku.cn/#analysis](http://gepia2.cancer-pku.cn/#analysis)) [13] combines TCGA and GTEx database. In the module of “Single Gene Analysis,” we input CBX3 then can obtain the bodymap, dot plot and bar plot of the gene expression profile. Furthermore, to compare the expression of CBX3 in cancer tissues and corresponding normal tissues, we used the “Expression analysis -Box Plots” module of the GEPIA2 to obtain box plots of the expression difference of the GTEx (Genotype-Tissue Expression) database. The settings included P-value cutoff = 0.01, log2FC (fold change) cutoff =1, and “Match TCGA normal and GTEx data”. In addition, on the base of the “Pathological Stage Plot” module of HEPIA2, the violin plots of the CBX3 expression in different pathological stages (stage I, stage II, stage III, and stage IV) of all TCGA tumors were obtained. The log2 [TPM (Transcripts per million) +1] transformed expression data were applied for the box or violin plots.

The UALCAN portal ([http://ualcan.path.uab.edu/analysis-prot.html](http://ualcan.path.uab.edu/analysis-prot.html)), can be used to conduct protein expression analysis of the CPTAC (Clinical proteomic tumor analysis consortium) dataset [14]. With the help of the website, we explored the expression of the total protein or phosphoprotein of CBX3 between primary tumor and normal tissues, respectively. The Human Protein Atlas ([https://www.proteinatlas.org/](https://www.proteinatlas.org/)) is a database about gene immunohistochemistry in cancer. The distribution and expression of each protein in 48 kinds of normal human tissues and 20 kinds of tumor tissues were examined by immunohistochemical technique, attached with at least 576 staining pictures.
Survival prognosis analysis

In the part, considering that there may be other factors result in death except tumor during the follow-up, we also visualized the relationship between CBX3 expression and prognosis (OS: overall survival; DSS: disease-specific survival; DFI: disease-free interval; PFI: progression-free interval) in the form of forest plots and Kaplan-Meier curves via the Sangerbox online platform (http://www.sangerbox.com/tool). The univariate survival analysis was applied to calculate the hazard ratio (HR) and 95% confidence intervals.

DNA methylate analysis

DNMIVD (http://119.3.41.228/dnmivd/index/) [15] is a database of methylation pan-cancer analysis based on the methylation chip database of TCGA and GEO. This database can not only view the methylation results of a certain gene in pan-cancer, but also understand the influence of certain methylation sites on prognosis models. After we searched CBX3, the basic information of methylation and cg probe related to the gene was obtained. The methylation difference of this gene in the promoter region of target cancer species was shown in a box diagram. Spearman correlation was used to analyze the relationship between methylation and expression. The relationship between methylation and prognosis, including OS, DFI and PFI was grouped according to the median value in Kaplan-Meier plots.

Genetic alteration analysis

In the cBioPortal web (https://www.cbioportal.org/) [16], we entered “CBX3” in the “TCGA PanCancer Atlas Studies” section for queries of the genetic alteration characteristics across all TCGA tumors. In the “Mutations” module, the mutated site information of CBX3 was displayed in the schematic diagram of the 3D (three-dimensional) structure. Moreover, we also applied the “Comparison” module to obtain the data on the overall, disease-free, progression-free, and disease-free survival differences for the cancer cases. The corresponding Kaplan-Meier plots with the log-rank P-value were generated for the significantly different cancers with or without CBX3 genetic alteration.

Protein phosphorylation analysis

To obtain a schematic diagram of the CBX3 protein domain, first of all, we used the UniProt tool (https://www.uniprot.org/) [17] to find the ID of corresponding protein. Subsequently, with the help of the SMART (http://smart.embl-heidelberg.de/) tool, we acquired the protein domains. The UALCAN portal was applied to display the phosphorylation site and protein expression level of CBX3 across the TCGA cancers.

TME analysis

The TME analysis included the evaluation on the immune infiltration of TILs and CAFs, the immune score or matrix score as well as expression of immune checkpoint markers. The correlation between CBX3 expression and the immune infiltration of the TILs (B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages and dendritic cells) across the TCGA cancers were evaluated with the help of Sangerbox,
and the top 3 tumors with significant differences were displayed. In addition, we analyzed the relationship between CBX3 expression and the immune score or matrix score, in the form of StromalScore, Est_ImmuneScore and ESTIMATEScore. Similarly, the top 3 tumors with significant differences were presented in scatter plots, respectively. Subsequently, we estimated the correlation between the expression levels of 40 common immune checkpoint markers and CBX3 expression by Spearman and Pearson correlation analyses. Gene expression levels are shown as log2 RSEM values.

With the help of the TIMER2.0 web server, we explored the association between CBX3 expression and immune infiltrates in the “Immune-Gene” module. The CAFs was selected in the immune infiltrates part. We selected the TIMER, EPIC, MCP-COUNTER and TIDE algorithms for estimations. Similarly, we applied the “Immune-Outcome” module to explore the clinical relevance of tumor immune subsets. The immune cells of the CAFs were selected. The age and stage were searched in the clinical part, respectively. Ultimately, we got the correlation heatmap and scatter plot of CBX3 and CAFs in gene expression and clinical outcome.

**Correlation analysis of CBX3-binding hub genes**

We firstly searched the STRING website (https://string-db.org/) get help from the keywords “CBX3” and “Homo sapiens”. Afterwards, the main parameters were set as follows: “Low confidence (0.150)“, “evidence”, “no more than 50 interactors” and “experiments”. Subsequently, we screened 50 CBX3-binding genes that were experimentally verified. Ultimately, the data was exported to Cytoscape v3.8.0 software for modification of the picture. simultaneously, we obtained 10 hub ones from CBX3-binding genes via Cytohubba module.

In the NetworkAnalyst website (https://www.networkanalyst.ca/), the protein-chemical interaction was identified by the Comparative Toxicogenomics Database (CTD) (downloaded on Nov. 2016) [18] to identify potential chemical material that could influence CBX3. In a sense, it can work in the prevention of diseases caused by CBX3 in advance. To identify transcription factors (TFs) that regulate the hubs at a transcriptional level, TF-hub interactions were obtained via the JASPAR database [19] and topological parameters (degree and betweenness centrality) were identified via NetworkAnalyst [20]. Similarly, to identify regulatory miRNAs that influence hubs at the post-transcriptional level, miRNAs-target gene interactions were gained from TarBase [21]. The database included experimentally supported miRNA-gene interactions and topological parameters were analyzed using NetworkAnalyst [20].

**CBX3-related genes enrichment analysis**

In the “Similar Gene Detection” module of GEPIA2, we obtained the top 100 CBX3-correlated target genes. In addition, the “correlation analysis” module was applied to perform Pearson correlation analysis for five pairwise genes of CBX3, respectively. Then, we acquired five corresponding dot plots marked with the P-value and the correlation coefficient (R). Moreover, based on “Gene_Corr” module in the “Exploration” of TIMER2.0, we obtained the heatmap data of the five genes in TCGA cancers.
To further screen the genes, we applied venny2.1.0 web servicer (https://bioinfogp.cnb.csic.es/tools/venny/index.html), an interactive venn diagram viewer, to conduct an intersection analysis on the CBX3-binding and interacted genes. Moreover, we combined the two sets of data to perform Gene ontology (GO) enrichment analysis and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis with R package. The data for biological process (BP), cellular component (CC), and molecular function (MF) in the GO enrichment analysis were visualized as bubble chart. The R language software [R-4.0.2, 64-bit] (https://www.r-project.org/) was used in the process. Two-tailed P <0.05 was considered statistically significant.

Results

Gene expression analysis data

The relationship between human CBX3 (NM_007276.4 for mRNA or NP_009207.2 for protein) and pan-cancer was explored from many aspects in the study. We analyzed the distribution and expression of CBX3 in different tumor cells and tissues based on the combination of the TCGA, GTEx, CCLE and HPA datasets.

The TIMER2.0 tool was applied to analyze the expression of CBX3 in cancers of TCGA. As shown in Figure 1a, CBX3 in most tumor tissues, including BLCA, BRCA, CHOL, COAD, ESCA, GBM, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, STAD, THCA (P<0.001), CESC, UCEC (P<0.01) has a higher expression level, compared with that in normal tissue. While, the KICH (P<0.01), and PCPG (P<0.05) have lower expression level of CBX3 than the corresponding control tissues.

With the help of GEPIA2 web server, we combined TCGA and GTEx database to compare the expression of gene in tumor tissues and normal ones. The expression difference of CBX3 in SKCM, GBM, LAML, DLBC, LGG, PAAD, TGCT, UCS and THYM (Figure 1b, P<0.05) were further evaluated. However, for other tumors, such as ACC, SARC, OV and CESC (Figure S1), it showed no significant difference.

In order to analyze the mRNA expression of CBX3 in cancer patients, the ONCOMINE database was used. As were shown in Figure 1c, mRNA expressions of CBX3 in various types of cancers were measured and compared to normal tissues. Significantly higher mRNA expression of CBX3 was found in multiple tumor tissues.

We visualized the distribution and expression of CBX3 in normal and tumor human tissues via the GEPIA2 website. As shown in Figure 1d, the body map, dot plot and bar plot indicate that CBX3 is distributed in multiple tissues of the human body. In addition, in general, tumor tissues have a higher expression level of CBX3 compared with that in normal ones, apart from KICH, PCPG and LAML. The result is consistent with that from Figure 1a, and further confirms that CBX3 is lowly expressed in KICH, PCPG compared with normal controls.
Considering that tissue-based RNA expression detection might be complicated by the non-tumor tissues that are adjacent to tumor cells, we further analyzed the expression and distribution of CBX3 in 1457 cell lines from 26 tumor types based on the CCLE database. The box plot in Figure 1e demonstrates that the cell lines from Burkitt lymphoma, neuroblastoma and lung_small-cell were the top ones expressing the highest levels of CBX3 mRNA. However, the cell lines from Hodgkin lymphoma showed relatively lower CBX3 mRNA expression.

In order to clarify the expression level of CBX3 in pathological stages of different tumors, we applied the “Pathological Stage Plot” module of GEPIA2 tool. Among them, the cancers including LIHC (P<0.001), ACC (P<0.01), THCA, TGCT, KICH, OV (P<0.05) showed difference between gene expression and pathological stages (Figure 1f).

The results of the CPTAC dataset showed compared with the normal tissues, the primary tumors of breast cancer, colon cancer, ovarian cancer, clear cell RCC, LUAD and UCEC had higher expression of CBX3 total protein (Figure 1g, P<0.001). Moreover, the additional analysis results in the ONCOMINE database (Figure 1c) further confirmed that CBX3 is highly expressed in breast cancer, cervical cancer, lung cancer, kidney cancer, colorectal cancer and ovarian (P<0.01) compared with normal controls.

We tried to explore the protein expression patterns of CBX3 in cancers by the Human Protein Atlas. As was shown in Figure 1h, CBX3 proteins were not detected in normal liver and testis tissues, whereas low expressions of them were observed in normal skin and pancreas tissues. In addition, medium protein expressions of CBX3 were expressed in normal tissues of breast, lung, cervix, ovary and colon. While high protein expressions in their corresponding cancers were observed. Taken together, the results showed that genic and proteinic expressions of CBX3 were over-expressed in most cancers.

Data from the multiple databases showed that CBX3 mRNA and protein expression in tumor tissues was significantly higher, compared to that in normal tissues, indicating that it might function as an oncogenic molecule in the development of diverse tumors.

**Survival analysis data**

Next, we evaluated the effect of CBX3 expression on prognostic value for the patients in pan-cancer. According to the expression levels of CBX3, the cancer cases were divided into high-expression and low-expression groups. Notably, CBX3 expression was significantly correlated with OS in 11 types of cancer (ACC, CESC, HNSC, KICH, LGG, LIHC, LUAD, MESO, PAAD, THYM and UVM) (Figure 2a). In addition, except as a protective factor in THYM (P=1.7e-02, HR=0.98), CBX3 appeared to be a risk factor in other cancer types. Specifically, the results were shown in Figure 2b, ACC (P=2.7e-07, HR=1.01), CESC (P=1.4e-02, HR=1), HNSC (P=2.0e-02, HR=1), KICH (P=1.7e-04, HR=1.04), LGG (P=4.6e-07, HR=1.01), LIHC (P=3.6e-07, HR=1.01), LUAD (P=3.4e-03, HR=1), MESO (P=2.8e-02, HR=1), PAAD (P=2.8e-04, HR=1.01) and UVM (P=6.3e-03, HR=1). Subsequently, we investigated the relationship between CBX3 expression and DFI and found that increased gene expression was correlated with poor prognosis in ACC (P= 0.0021, HR= 1.01), CESC (P=0.0400, HR=1.01), KIRP (P=0.0140, HR=1.01), LIHC (P=0.0130, HR=1.01) and
PAAD (P=0.0280, HR =1.01), but with favorable prognosis in OV (P= 0.0031, HR =0.99) (Figure 2c, d). Next, the relationship between CBX3 expression and PFI in 33 cancers from TCGA was analyzed. Results showed CBX3 expression impacted patients’ PFI in 12 cancer types (Figure 2e). Specifically, the Kaplan-Meier curves of all the cancers (Figure 2f) showed that high expression of CBX3 was significantly correlated with poor prognosis of patients in ACC (P =9.9e-10, HR =1.01), CESC (P =5.0e-03, HR =1.01), HNSC (P =1.8e-03, HR =1), KICH (P =3.1e-04, HR =1.04), KIRP (P =1.1e-02, HR =1.01), LGG (P =3.0e-07, HR =1), LIHC (P =1.5e-04, HR =1.01), LUAD (P =1.7e-03, HR =1), MESO (P =1.8e-02, HR =1.01), PAAD (P =4.2e-04, HR =1.01), PRAD (P =3.5e-03, HR =1.01) and UVM (P =1.1e-04, HR =1.02). Ultimately, we assessed the relationship between CBX3 expression and DSS to exclude non-tumor death factors. The results showed that high expression of CBX3 affected DSS unfavorably in ACC (P =3.1e-07, HR =1.01), CESC (P =3.9e-03, HR =1.01), HNSC (P =1.1e-02, HR =1), KICH (P =6.8e-05, HR =1.05), LGG (P =6.1e-07, HR =1.01), LIHC (P =5.8e-04, HR =1.01), LUAD (P =1.9e-02, HR =1), MESO (P =1.4e-02, HR =1.01), PAAD (P =1.4e-03, HR =1.01), THCA (P =2.0e-02, HR =1.03), UVM (P =6.0e-03, HR =1.01) (Figure 2g, h). In conclusion, these results indicated that CBX3 expression level is a key influencing factor for the prognosis of patients, especially those with ACC, CESC, LIHC and PAAD. Moreover, the above data mainly from the datasets of TCGA revealed that elevated expression of CBX3 is associated with the poor prognosis of cases in most tumors.

DNA methylation analysis data

The DNMIVD tool was used to investigate the potential association between CBX3 DNA methylation and the prognosis of pan-cancers based on TCGA and GEO databases. As shown in the TABLE 1, 21 methylation sites of CBX3 and related information were summarized. The correlation between promoter methylation in target cancers of KIRC, LUSC and UCEC and CBX3 gene expression were shown in Figure 3a, c, e, respectively. Except in UCEC, we observed a significant positive correlation of CBX3 DNA methylation and gene expression in KIRC and LUSC. Furthermore, the Figure 3b, 3d and 3f displayed the relationship between gene methylation and prognosis in KIRC, LUSC and UCEC, respectively. In contrast to that in LUSC, the high methylation expression of CBX3 in KIRC showed the worse prognosis, compared with the low expression.

Genetic alteration analysis data

The genetic variation status of CBX3 in various cancer samples were observed via the cBioportal tool. As shown in Figure 4a, the patients with UCEC occupied the highest alteration frequency of CBX3, nearly 5%. Furthermore, the primary type of alteration performed “mutation” (~3% frequency). The “amplification” type of CNA was the primary type in the esophageal cancer cases, which show an alteration frequency of ~4% (Figure 4a). In addition, all sarcoma cancer cases with genetic alteration (~3% frequency) presented “amplification” alteration (Figure 4a). As displayed in Figure 4b, the types, sites and case number of the CBX3 genetic alteration are presented clearly. We found that missense mutation of CBX3 was the main type of genetic alteration, and K14Nfs*23 alteration in the Chromo, which was detected in 2 cases of UCEC, 2 cases of STAD and 1 case of COAD (Figure 4b). We presented the 3D structure of CBX3 protein
Additionally, we explored the potential association between genetic alteration of CBX3 and the clinical survival prognosis of cases with different types of cancer. The data of Figure 4d indicated that STAD cases with altered CBX3 showed poorer prognosis in disease-free survival (P=4.046e-03), compared with cases without CBX3 alteration. However, no significant differences were observed in overall (P=0.835), disease-specific (P=0.490) and progression-free (P=0.945) survival.

Protein phosphorylation analysis data

The CBX3 phosphorylation sites from SMART web were summarized in the pattern diagram of Figure 5a. The differences in protein phosphorylation levels of CBX3 between normal tissues and primary tumor ones were compared from the CPTAC dataset via UALCAN tool. Among them, we analyzed six types of tumors, including breast cancer, LUAD, clear cell RCC, ovarian cancer, colon cancer and UCEC. The S93 locus of CBX3 exhibits a significant difference on phosphorylation level in five primary tumor tissues except clear cell RCC, compared with normal tissues (Figure 5c-g, all P <0.05). In contrast with the ovarian cancer, the other four cancers expressed higher phosphorylation level than corresponding controls. The second change was an increased phosphorylation level of the S95 locus for LUAD (Figure 5c, P=3.1e-02), colon cancer (Figure 5e, P=2.1e-08) and UCEC (Figure 5g, P=5.8e-10). While, clear cell RCC (Figure 5f, P=5.2e-07) in the S95 locus displayed decreased level. We also analyzed and summarized the expression difference in other phosphorylation sites. As shown in Figure 5, most of the primary tumors displayed a higher phosphorylation level than the corresponding normal tissues.

TME analysis data

The TME contains various cells, yet infiltrating immune cells account for a large proportion [22, 23]. As prominent components of TME, these cells were believed to be closely related with the initiation, progression or metastasis of cancer [24]. It was reported that the TILs and CAFs participated in modulating the function of various tumor-infiltrating immune cells [25]. Therefore, it is necessary to evaluate the correlation between CBX3 and the characters of immune infiltration in pan-cancer. In the part, we investigated the potential relationship between the infiltration level of different immune cells and CBX3 gene expression or clinical outcome in diverse cancer types of TCGA.

We obtained the scores of 6 infiltrating immune cells from TCGA database and analyzed the correlation with CBX3. As shown in Figure 6a, results revealed that CBX3 expression was appreciably positively correlated with the infiltration levels of 6 immune cells in KIRC and LIHC, while negatively in LUSC. For the StromalScore, the top 3 cancers with significant difference all showed negatively correlated with CBX3 expression in TGCT, BRCA and LUSC. The increased gene expression was correlated with decreased Est_ImmuneScore in BRCA, UCEC and SKCM. In addition, CBX3 expression was negatively correlated with the ESTIMATEScore in BRCA, UCEC and LUSC (Figure 6b). Subsequently, we analyzed the correlation between CBX3 expression and that of 40 common immune checkpoint genes. The heat plot showed CBX3 expression was correlated with more than 20 immune checkpoint markers in UVM and KIRC. While, in LIHC, it had even close to 40 checkpoint markers (Figure 6c). To be conclude, these results strongly suggest that CBX3 plays a vital role in tumor immunity.
Moreover, we observed a statistical positive correlation of CBX3 expression and the estimated infiltration value of CAFs for the CESC, ESCA, HNSC-HPV-, KICH, LGG, LIHC, PAAD, and THCA, but noted a negative correlation for TGCT and THYM (Figure 6d) based on the TIDE, MCP-COUNTER and EPIC algorithms. The scatterplot data of the above tumors produced using one algorithm are presented in Figure 6e. For instance, the CBX3 expression level in KICH is positively correlated with the infiltration level of CAFs (cor=0.486, P=4.06e-05) based on the TIDE algorithm. Similarly, as shown in the Figure 6f and 6h, we revealed the potential relationship between the level of cancer associated fibroblast and the clinical outcomes of age or stage in various cancer types in the form of heat map, respectively. In addition, the cumulative survival graphs on the cancer with statistical difference were displayed in the Figure 6g and 6i. For example, as the age of LGG patients with high infiltration levels increases, the survival outcome is worse. To sum up, the results indicated that CBX3 expression was closely correlated with the TME in cancers.

**Correlation analysis of CBX3-binding hub genes**

To further investigate the molecular mechanism of the CBX3 gene in the occurrence and development of tumors, we attempted to screen out and analyze the CBX3-binding hub genes. Based on the STRING tool, we obtained 50 CBX3-binding genes, which were supported by experimental evidence (Figure 7a). Afterwards, 10 hub genes (CBX3, TRIM24, HIST1H1E, HIST1H1C, H2AFY, CHD1L, POLA1, EHMT2, L3MBTL1, HIST1H1B) were further screened with the help of Cytohubba module (Figure 7b).

In the protein-chemical interactions network, hubs were associated with the chemical substances, Aflatoxin B1, Valproic Acid, Formaldehyde, Copper Sulfate, Cyclosporine (Figure 7c). These toxic chemicals, as a predisposing factor, may induce cancer by affecting CBX3 or its binding genes.

We presented hubs-TFs interactions (Figure 7d) and hubs-miRNAs interactions (Figure 7e) in the form of network diagram, respectively. We further detected central regulatory biomolecules (TFs and miRNAs) according to the set topological parameters. Seven TFs (GATA3, E2F1, TFAP2A, GATA2, FOXC1, JUN, CREB1) and four miRNAs (Hsa-mir-182-5p, Hsa-mir-320a, Hsa-mir-24-3p, Hsa-mir-1-3p) were detected from the hubs-TFs and hubs-miRNAs interaction networks, respectively. The TFs and miRNAs participated the tumorigenesis by regulating CBX3-binding hub genes in transcription level and post-transcriptional level.

**Enrichment analysis of CBX3-related partners**

Similarly, the CBX3 expression-correlated genes were also screened out for a series of pathway enrichment analyses to further investigate the molecular mechanism from another point of view. The GEPIA2 tool was used to screen the top 100 genes that correlated with CBX3 expression from all tumor expression data of TCGA. The scatter plots (Figure 8a) displayed the positive correlation between CBX3 expression level and that of HNRNPA2B1(R=0.67), PLEKHA8(R=0.64), LSM5(R=0.64), FTSJ2(R=0.63) and DHX9(R=0.62) genes (all P <0.001). Moreover, the corresponding heatmap further confirmed the positive association between CBX3 and the five genes in the majority of tumors (Figure 8b). We did an
intersection analysis of the correlated gene group and interacted gene group, then obtained one common member, centromere protein A CENPA (Figure 8c). Finally, the KEGG and GO enrichment analyses were performed with the combined two datasets via R language, respectively.

The GO enrichment analysis data were presented in three parts, including BP (Biological Process), CC (Cellular Component) and MF (Molecular Function). The results indicated that most of these genes are linked to the cellular biology of DNA. For example, in the BP part, DNA conformation change and protein-DNA subunit organization played a major role. In the CC part, nuclear chromatin and chromosomal region were main enrichment position. The change in molecular function of nucleosome binding and chromatin DNA binding is linked to the pathways (Figure 8d).

The bubble plot of KEGG data suggested that the two pathways of “alcoholism” and “neutrophil extracellular trap formation” might be involved in the tumor pathogenesis of CBX3 to a large extent (Figure 8e).

Discussion

The pan-cancer analysis of gene is of importance for prevention of cancer, development of biomarkers and search for a potential therapeutic target [26]. It has been reported that CBX3 participates in a series of cellular biological processes, such as chromosome segregation, transcriptional regulation, DNA repair and RNA splicing [27]. Increased literatures have reported high expression of CBX3 can induce clinical diseases, especially the development of multiple cancers, and associated with poor prognosis [28-30]. The CBX3 gene played tumor-promoting role in PAAD and induced the increased proliferation, migration and invasion of the PAAD cells. Similarly, the silence of CBX3 can inhibit glioma cell proliferation by blocking the cell cycle in the G2/M phase [31]. The CBX3 may function as a tumor promoter. In the previous study, CBX5, the HP1 orthologs, is the most studied HP1 protein, while CBX3 is barely analyzed. Through a literature search, we failed to retrieve any publication about a pan-cancer analysis of CBX3. Whether CBX3 can play a key role in the pathogenesis of multiple tumors has not been identified. Thus, we comprehensively analyzed the CBX3 gene in a wide variety of tumors based on the data from TCGA, CPTAC, cBioPortal and GEO databases. In addition, the molecular features in diverse expression patterns, genetic alteration, protein phosphorylation, prognostic values, as well as immune infiltration were examined by comprehensive applications of bioinformatics.

In this study, we found that abnormal expression of CBX3, mostly overexpression in genic and proteinic level exists in human pan-cancer including BRCA, COAD, PAAD, UCEC and LUAD. Nevertheless, the survival prognosis analysis data from different forms of prognosis for the CBX3 gene suggested distinct conclusions for different tumors. In multiple cancers, survival analysis showed CBX3 overexpression was associated with poor prognosis, especially in CESC, LIHC, LUAD and PAAD. The result was consistent with the one from a recent article on CBXs and LUAD from Kaplan-Meier analysis [32]. The previous research screened proteins with alterations in expression, DNA sequences, and copy numbers in LUAD from 124 histone methylation reader proteins based on several databases. The result indicated CBX3/HP1γ was
the most frequently overexpressed and amplified one, and high mRNA levels were associated with poor prognosis in LUAD patients [33]. CBX3, as a binding modules-containing protein, contribute to tumorigenesis by mediating the effect of histone methylation. In addition, further regression model analysis showed the role of independent prediction of CBX3 for OS and PFI in LUAD. For lung cancer, the data from LUSC were also analyzed. And a high mRNA expression was indicated but not a obvious correlation with the survival prognosis. Moreover, the related literature reports were not shown in the search process. Similarly, CBX3 was found to be overexpressed in PAAD based on the data from GEO datasets. Furthermore, for the proliferation and invasiveness of PAAD cells in vitro, CBX3 played a role of acceleration. The high level of expression was also associated with poor prognosis of overall and disease-free survival [34]. The different data source and survival information processing both demonstrated similar results. To sum up, our data strongly indicate CBX3 may be a potential prognostic biomarker and molecular therapeutic target in LUAD and PAAD.

For cancers in digestive system, besides PAAD, the relationship between gastric cancer, hepatocellular carcinoma, colorectal cancer and CBX3 were analyzed in the literatures. An article reported the overexpression of CBX3 was significantly correlated with a short OS prognosis in STAD [35]. However, other study has shown the OS prognosis was better in STAD patients with high CBX3 expression [36]. In our results, we failed to observe the significant correlation. The diversity of sample numbers and grouping method in the prognostic analysis may be the reason. Thus, larger sample sizes are required to confirm the effect of CBX3 in the survival prognosis of STAD. CBX3 was reported to show a mutation rate of more than 5% in STAD [36]. Our results also observed a mutation in the site of K14Nfs*23. Further, the mutation was significantly negative correlated with disease free survival in patients with STAD. For LIHC, other research based on Kaplan-Meier and Cox regression analyses [37] verified our survival analysis result about correlation between CBX3 high expression and poor clinical prognosis. However, the previous one just examined the overall survival, our study further analyzed the remaining prognostic model, including PFI, DFI and DSS.

HP1 family proteins are extensively modified in a manner analogous to histones, including phosphorylation, acetylation, methylation and others, which can modulate their sub-nuclear location and bioactivity [38]. HP1γ can specifically bind methylated histone H3, such as H3K36, through their conserved chromodomain [39]. Furthermore, data analysis indicated multiple phosphorylation sites were existed on HP1γ. However, there is almost no research about the influence of protein modification on tumors. We first used the DNMIVD and CPTAC datasets to explore the expression level of the CBX3 methylation and phosphorylation across all TCGA cancers. The findings indicated the correlation between methylation and expression or prognosis, as well as the phosphorylation level of CBX3 at a certain domain in some tumors. Further, additional evidence is required to evaluate the potential role of CBX3 methylation and phosphorylation in the tumorigenesis.

It has been reported that immune cells in TME played a role in either tumor-promoting or tumor-suppressing and associated with proliferation and progression of cancer cell. Our results first suggested that the expression of CBX3 was significantly associated with the TILs and cancer associated fibroblast
in certain tumors, implying that CBXs may potentially reflect the immune status besides the prognosis. At present, there are few studies and literature on CBX3 and immune infiltration. There is a research report suggesting CBX3 affects factors related to immunotherapy in STAD [40]. These results were in line with our data in TILs and tumor mutational burden (TMB, figure S2, S5). Similarly, our observation (S3) was consistent with the previous finding that CBX3 was positively associated with the infiltration of CD8+T cells and macrophages in COAD [41]. In addition, we also observed a statistical correlation between CBX3 expression and the immune infiltration level of TILs in the tumors of KIRC, LIHC and LUSC. Besides, other immune associated indexes including immunization score, matrix score, check point, neoantigen (Figure S4) and TMB in multiple cancers were examined. These findings might present detailed immunization evidences across all TCGA tumors to assist for new immunotherapies.

Furthermore, we integrated the information on CBX3-binding genes and CBX3-related partners for a series of analyses and obtained one common component, CENPA. The identified TFs and miRNAs participated the tumorigenesis by regulating CBX3-binding hub genes in transcription level and post-transcriptional level. The related chemical substances, change in cellular biology of DNA and other enriched pathway have a potential impact in the etiology and molecular mechanism. The CENPA is an epigenetic mark that determines centromere identity. Existing research reported overexpression of the gene is crucial to certain cancer growth [42]. The expression of CENPA was estimated across all tumors from TCGA and GTEx database (Figure S6). Interestingly, all illustrated statistical difference compared the corresponding normal group, implying the analytical value for biomarkers or therapeutic target in pan-cancers.

However, even though we systematically investigated information of pan-cancer from multiple databases, there were some deficiencies in the current study, inevitably. Above all, although the comprehensive bioinformatic analysis provided some meaningful insights of CBX3 in pan-cancer, further biological studies will be beneficial for clarifying the influence of CBX3 in the molecular level. Next, despite the finding showed that CBX3 expression was correlated with tumor immunity and clinical survival, future deep studies focusing on CBX3 expression and tumor immune microenvironment should be researched, thus providing a therapeutic strategy based on the immune.

Conclusion

Taken together, the first pan-cancer analyses of CBX3 indicated statistical correlations of CBX3 expression with survival prognosis, DNA methylation, protein phosphorylation, immune cell infiltration across multiple tumors. Because CBX3 showed overexpression in multiple cancers and correlation with poor prognosis, CBX3 might be considered as a novel biomarker and target for cancer therapy. In addition, our analysis provided a potential mechanism that CBX3 might regulate tumor immune microenvironment. The results contribute to explain the role of CBX3 in tumorigenesis from the perspective of clinical tumor samples.

Abbreviations
ACC: Adrenocortical carcinoma; BLCA: Bladder urothelial carcinoma; BRCA: Breast invasive carcinoma; CAFs: cancer-associated fibroblasts; CBX3: chromobox protein homolog 3; CENPA: centromere protein A; CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL: Cholangiocarcinoma; COAD: Colon adenocarcinoma; DFI: disease-free interval; DLBC: Lymphoid neoplasm diffuse large B-cell lymphoma; DSS: disease-specific survival; ESCA: Esophageal carcinoma; GBM: Glioblastoma multiforme; HNSC: Head and neck squamous cell carcinoma; KICH: Kidney chromophobe; KIRC: Kidney renal clear cell carcinoma; KIRP: Kidney renal papillary cell carcinoma; LAML: Acute myeloid leukemia; LGG: Brain lower grade glioma; LIHC: Liver hepatocellular carcinoma; LUAD: Lung Adenocarcinoma; LUSC: Lung squamous cell carcinoma; OS: overall survival; OV: Ovarian serous cystadenocarcinoma; PAAD: Pancreatic adenocarcinoma; PCPG: Pheochromocytoma and paraganglioma; PFI: progression-free interval; PRAD: Prostate adenocarcinoma; SARC: Sarcoma; SKCM: Skin cutaneous melanoma; STAD: Stomach adenocarcinoma; TGCT: Testicular germ cell tumors; THCA: Thyroid carcinoma; THYM: Thymoma; TIL: tumor-infiltrating lymphocyte; TMB: tumor mutational burden; UCEC: Uterine corpus endometrial carcinoma; UCS: Uterine carcinosarcoma; HNSC-HPV: head and neck squamous carcinoma-HPV; HR: hazard ratio.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The authors ensure that the data analyzed in the research is publicly available. The data can be found as follows: TIMER2.0 (http://timer.cistrome.org/), UALCAN (http://ualcan.path.uab.edu/), cBioPortal (https://www.cbioportal.org/), CCLE (https://portals.broadinstitute.org/ccle), GEPIA2 (http://gepia2.cancer-pku.cn/#analysis), Human Protein Atlas (https://www.proteinatlas.org/), DNMIVD (http://119.3.41.228/dnmivd/index/), ONCOMINE (https://www.ONCOMINE.org/) (accession numbers: 19400242@muc.edu.cn), et al.

Competing interest

The authors declare that they have no competing interests.

Funding

There is no available project funding.

Author's contributions
HN wrote the paper. PC edited and adjusted the figures. LF edited the manuscript. BS provided funding and reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Blum A, Wang P, Zenklusen J: SnapShot: TCGA-Analyzed Tumors. Cell 2018, 173(2):530.
2. Clough E, Barrett T: The Gene Expression Omnibus Database. Methods in molecular biology (Clifton, NJ) 2016, 1418:93–110.
3. Tomczak K, Czerwińska P, Wiznerowicz M: The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemporary oncology (Poznan, Poland) 2015, 19:A68-77.
4. Fan Y, Li H, Liang X, Xiang Z: CBX3 promotes colon cancer cell proliferation by CDK6 kinase-independent function during cell cycle. Oncotarget 2017, 8(12):19934–19946.
5. Xu Y, Pan S, Song Y, Pan C, Chen C, Zhu X: The Prognostic Value of the Chromobox Family in Human Ovarian Cancer. Journal of Cancer 2020, 11(17):5198–5209.
6. Chang S, Lai Y, Chen Y, Wang N, Wang W, Lai J: CBX3/heterochromatin protein 1 gamma is significantly upregulated in patients with non-small cell lung cancer. Asia-Pacific journal of clinical oncology 2018, 14(5):e283-e288.
7. Slezak J, Truong M, Huang W, Jarrard D: HP1γ expression is elevated in prostate cancer and is superior to Gleason score as a predictor of biochemical recurrence after radical prostatectomy. BMC cancer 2013, 13:148.
8. Liu M, Huang F, Zhang D, Ju J, Wu X, Wang Y, Wang Y, Wu Y, Nie M, Li Z et al: Heterochromatin protein HP1γ promotes colorectal cancer progression and is regulated by miR-30a. Cancer research 2015, 75(21):4593–4604.
9. Liang Y, Lin H, Chen C, Zeng: Prognostic values of distinct CBX family members in breast cancer. Oncotarget 2017, 8(54):92375–92387.
10. Chen L, Cheng C, Qu C, Wang P, Chen H, Meng Z, Chen Z: CBX3 promotes proliferation and regulates glycolysis via suppressing FBP1 in pancreatic cancer. Biochemical and biophysical research communications 2018, 500(3):691–697.
11. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Li B, Liu X: TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic acids research 2020, 48:W509-W514.
12. Rhodes D, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs B, Barrette T, Anstet M, Kincead-Beal C, Kulkarni P et al: Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia (New York, NY) 2007, 9(2):166–180.
13. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic acids research 2017, 45:W98-W102.
14. Chandrashekar D, Bashel B, Balasubramanya S, Creighton C, Ponce-Rodriguez I, Chakravarthi B, Varambally S: **UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses.** *Neoplasia (New York, NY)* 2017, 19(8):649–658.

15. Ding W, Chen J, Feng G, Chen G, Wu J, Guo Y, Ni X, Shi T: **DNMIVD: DNA methylation interactive visualization database.** *Nucleic acids research* 2020, 48:D856-D862.

16. Gao J, Aksoy B, Dogrusoz U, Dresdner G, Sumer S, Sun Y, Jacobsen A, Sinha R, Larsson E et al: **Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.** *Science signaling* 2013, 6(269):pl1.

17. Consortium U: **UniProt: the universal protein knowledgebase in 2021.** *Nucleic acids research* 2021, 49:D480-D489.

18. Davis A, Wiegers T, Wiegers J, Grondin C, Johnson R, Mattingly C: **CTD Anatomy: analyzing chemical-induced phenotypes and exposures from an anatomical perspective, with implications for environmental health studies.** *Current research in toxicology* 2021, 2:128–139.

19. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon J, van der Lee R, Bessy A, Chêneby J, Kulkarni S, Tan G et al: **JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework.** *Nucleic acids research* 2018, 46:D1284.

20. Xia J, Gill E, Hancock R: **NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data.** *Nature protocols* 2015, 10(6):823–844.

21. Sethupathy P, Corda B, Hatzigeorgiou A: **TarBase: A comprehensive database of experimentally supported animal microRNA targets.** *RNA (New York, NY)* 2006, 12(2):192–197.

22. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf A, Angell H, Fredriksen T, Lafontaine L, Berger A et al: **Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer.** *Immunity* 2013, 39(4):782–795.

23. Bonanno L, Pavan A, Dieci M, Di Liso E, Schiavon M, Comacchio G, Attili I, Pasello G, Calabrese F, Rea F et al: **The role of immune microenvironment in small-cell lung cancer. Distribution of PD-L1 expression and prognostic role of FOXP3-positive tumour infiltrating lymphocytes.** *European journal of cancer (Oxford, England: 1990)* 2018, 101:191–200.

24. Steven A, Seliger B: **The Role of Immune Escape and Immune Cell Infiltration in Breast Cancer.** *Breast care (Basel, Switzerland)* 2018, 13(1):16–21.

25. Chen X, Song E: **Turning foes to friends: targeting cancer-associated fibroblasts.** *Nature reviews Drug discovery* 2019, 18(2):99–115.

26. Consortium ITP-CAoWG: **Pan-cancer analysis of whole genomes.** *Nature* 2020, 578(7793):82–93.

27. Canzio D, Larson A, Narlikar G: **Mechanisms of functional promiscuity by HP1 proteins.** *Trends in cell biology* 2014, 24(6):377–386.

28. Zhang H, Chen W, Fu X, Su X, Yang A: **CBX3 promotes tumor proliferation by regulating G1/S phase via p21 downregulation and associates with poor prognosis in tongue squamous cell carcinoma.** *Gene* 2018, 654:49–56.
29. Zhao S, Wang F, Yang M, Wang X, Jin C, Ji Q, Li S, Zhao X: CBX3 promotes glioma U87 cell proliferation and predicts an unfavorable prognosis. Journal of neuro-oncology 2019, 145(1):35–48.

30. Li X, Gou J, Li H, Yang X: Bioinformatic analysis of the expression and prognostic value of chromobox family proteins in human breast cancer. Scientific reports 2020, 10(1):17739.

31. Wang S, Liu F, Wang Y, Fan W, Zhao H, Liu L, Cen C, Jiang X, Sun M, Han P: Integrated analysis of 34 microarray datasets reveals CBX3 as a diagnostic and prognostic biomarker in glioblastoma. Journal of translational medicine 2019, 17(1):179.

32. Zhang Q, Liu W, Zhang H, Xie G, Miao Y, Xia M, Guo A: hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. Genomics, proteomics & bioinformatics 2020, 18(2):120–128.

33. Alam H, Li N, Dhar S, Wu S, Lv J, Chen K, Flores E, Baseler L, Lee M: HP1γ Promotes Lung Adenocarcinoma by Downregulating the Transcription-Repressive Regulators NCOR2 and ZBTB7A. Cancer research 2018, 78(14):3834–3848.

34. Chen L, Cheng C, Qu C, Wang P, Chen H, Meng Z, Chen Z: Overexpression of CBX3 in Pancreatic Adenocarcinoma Promotes Cell Cycle Transition-Associated Tumor Progression. International journal of molecular sciences 2018, 19(6).

35. Ma T, Ma N, Chen J, Tang F, Zong Z, Yu Z, Chen S, Zhou T: Expression and prognostic value of Chromobox family members in gastric cancer. Journal of gastrointestinal oncology 2020, 11(5):983–998.

36. Lin K, Zhu J, Hu C, Bu F, Luo C, Zhu X, Zhu Z: Comprehensive analysis of the prognosis for chromobox family in gastric cancer. Journal of gastrointestinal oncology 2020, 11(5):932–951.

37. Zhong X, Kan A, Zhang W, Zhou J, Zhang H, Chen J, Tang S: CBX3/HP1γ promotes tumor proliferation and predicts poor survival in hepatocellular carcinoma. Aging 2019, 11(15):5483–5497.

38. Harouz H, Rachez C, Meijer B, Marteyn B, Donnadieu F, Cammas F, Muchardt C, Sansonetti P, Arbibe L: Shigella flexneri targets the HP1γ subcode through the phosphothreonine lyase OspF. The EMBO journal 2014, 33(22):2606–2622.

39. Zaidan N, Sridharan R: HP1γ regulates H3K36 methylation and pluripotency in embryonic stem cells. Nucleic acids research 2020, 48(22):12660–12674.

40. Lin H, Lian J, Xia L, Guan G, You J: CBX3 Promotes Gastric Cancer Progression and Affects Factors Related to Immunotherapeutic Responses. Cancer management and research 2020, 12:10113–10125.

41. Li Q, Pan Y, Cao Z, Zhao S: Comprehensive Analysis of Prognostic Value and Immune Infiltration of Chromobox Family Members in Colorectal Cancer. Frontiers in oncology 2020, 10:582667.

42. Saha A, Contreras-Galindo R, Niknafs Y, Iyer M, Qin T, Padmanabhan K, Siddiqui J, Palande M, Wang C, Qian B et al: The role of the histone H3 variant CENPA in prostate cancer. The Journal of biological chemistry 2020, 295(25):8537–8549.
Figure 1

Expression level of CBX3 gene in different tumors and pathological stages in multiple databases. (a) The gene expression status of CBX3 in different cancers or specific cancer subtypes in TCGA database was analyzed through TIMER2.0. * P<0.05; ** P<0.01; *** P<0.001. (b) For the absent types of cancers in the TCGA project, the corresponding normal tissues of the GTEx database were included as controls. The box plot data were supplied. * P <0.05. (c) Transcriptional expression of CBX3 in different types of cancer diseases (ONCOMINE database). Difference of transcriptional expression was compared by students’ t-test. Cut-off of p value and fold change were as following: p value: 0.01, fold change: 1.5, gene rank: 10%, data type: mRNA. (d)The expression and distribution of CBX3 in tumor and normal samples of human tissues were illustrated in various types. It is highly expressed in most cancers in GEPIA2 database. (e)The mRNA expression of CBX3 in tumor cell lines was revealed from CCLE database. (f) Based on the TCGA dataset, the expression levels of the CBX3 gene were analyzed by the main pathological stages (stage I, stage II, stage III, and stage IV) of ACC, KICH, OV, LIHC, THCA and TGCT. Log2 (TPM+1) was
applied for log-scale. (g) Based on the CPTAC dataset, we analyzed the expression level of CBX3 total protein between normal tissue and primary tissue of breast cancer, colon cancer, ovarian cancer, clear cell RCC, LUAD and UCEC. *** P<0.001. (h) We showed representative and contrastive immunohistochemistry images of CBX3 in tumor tissues and normal tissues (Human Protein Atlas).

Figure 2
Correlation between CBX3 gene expression and survival prognosis of cancers in TCGA. We used the Sangbox to perform OS (a,b), DFI (c,d), PFI(e,f) and DSS(g,h) analyses of different tumors in TCGA by CBX3 gene expression. The forest plots and Kaplan-Meier curves with positive results are given.

Figure 3

DNA methylation analysis of CBX3 gene in unequal tumors. The boxplot of expression and DNA methylation for CBX3 in KIRC(a), LUSC(c), UCEC(e) together with spearman correlation between gene expression and promoter methylation were displayed. The relationship between gene methylation of CBX3 and prognosis including OS, DFI and PFI in KIRC(b), LUSC(d), UCEC(f) were analyzed.
Figure 4

Mutation feature of CBX3 in different tumors of TCGA. The cBioPortal tool was used to analyze the mutation features of CBX3 for the TCGA tumors. The alteration frequency with mutation type (a) and mutation site (b) are displayed. We display the mutation site with the highest alteration frequency (K14Nfs*23) in the 3D structure of CBX3 (c). We also analyzed the potential correlation between mutation status and overall, disease-specific, disease-free and progression-free survival of STAD (d) with the cBioPortal tool.
Figure 5

Phosphorylation analysis of CBX3 protein in multiple tumors. Based on the UniProt and SMART tools, we analyzed the protein domain of CBX3 phosphoprotein (S93, S95, S97, S99, S102 and S176 sites). The phosphoprotein sites with positive results are also displayed in the schematic diagram of CBX3 protein between normal tissue and primary tissue of selected tumors via the UALCAN (a). We also supply the box plots for different cancers, including ovarian cancer (b), LUAD (c), breast cancer (d), colon cancer (e), clear cell RCC (f) and UCEC (g).
Figure 6

Correlation analysis between CBX3 and TME with the help of TIMER2.0 tool and Sangerbox. (a) The correlation between CBX3 expression and TILs was showed in top 3 cancers. (b) The scatter plots presented the relationship between CBX3 expression and the StromalScore, Est_ImmuneScore, ESTIMATEScore in top 3 cancers. (c) The correlation between gene expression and the common immune checkpoint marker was displayed. The heat map (d) and the scatter diagram (e) of the correlation
between CBX3 expression and infiltration of CAFs were illustrated, respectively. The heat map (f) and the survival curve (g) indicated the correlation between the clinical outcome of age and infiltration of CAFs. The fig 6h and fig 6i showed the relation on clinical stage and infiltration of CAFs.

**Figure 7**

Correlation analysis of CBX3-binding hub genes. (a) We first obtained 50 available experimentally determined CBX3-binding genes using the STRING tool. (b) 10 CBX3-binding hub genes were analyzed
with Cytohubba plug-in component in Cytoscape 3.8.0 software. We did interaction analysis of hub genes with chemical (c), transcription factor (d) and miRNA (e) based on the NetworkAnalyst approach.

**Figure 8**

CBX3-related gene enrichment analysis. (a) Based on the GEPIA2 approach, we obtained the top 100 CBX3-correlated genes in TCGA projects and analyzed the expression correlation between CBX3 and selected targeting genes, including HNRNPA2B1, PLEKHA8, LSM5, FTSJ2, and DHX9. (b) The
corresponding heatmap data of selected targeting genes in the detailed cancer types were displayed. (c) An intersection analysis of the CBX3-binding and correlated genes was conducted in the form of a Venn diagram. Based on the CBX3-binding and interacted genes, the bubble chart for the molecular function data in GO analysis (d) and the KEGG pathway analysis (e) were performed with R 4.0.2 software.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- table1.pdf
- Supplementarymaterial.docx
- supplement.pdf