Prognostic and immunological value of Chromobox family genes in Pan-Cancer

Hongmiao XU  
Graduate School, Hunan University of Chinese Medicine

Yingchun HE  (✉ 003452@hnucm.edu.cn )  
Medical School, Hunan University of Chinese Medicine, Changsha

Fang ZOU  
Medical School, Hunan University of Chinese Medicine, Changsha

Pan ZOU  
Graduate School, Hunan University of Chinese Medicine

Fangliang ZHOU  
Medical School, Hunan University of Chinese Medicine, Changsha

Xianwen WANG  
The First Affiliated Hospital of Hunan University of Chinese Medicine Changsha, Hunan

Hongjian SHI  
Medical School, Hunan University of Chinese Medicine, Changsha

Lan HE  
The First Affiliated Hospital of Hunan University of Chinese Medicine Changsha, Hunan

Research Article

Keywords: chromobox family, Pan cancer, survival analysis, immune subtypes, tumor microenvironment

Posted Date: July 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1734872/v1

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

Objectives: This study aims to analyze the expression and prognostic significance of chromobox (CBX) family genes in pan cancer and to associate their roles with immune subtypes, tumor microenvironment and drug sensitivity.

Methods: TCGA +GTEX method was used to analyze five members of the CBX2/4/6/7/8, Using 11,093 samples from 33 tumor types in TCGA database to multidimensional analysis of five members of the CBX family, CBX2, CBX4, CBX6, CBX7, CBX8. including the differential expression of the CBX family in pan cancer, as well as the correlation between immune subtypes, tumor purity and stemness. Finally, the correlation data of CellMiner database was used to analyze the relationship between CBXs and drug sensitivity. Then, GO plot was performed, and GSEA analysis and weighted gene co-expression network analysis were generated. CellMiner database was used to analyze the correlation and mechanism of CBX family genes with drug sensitivity.

Results: CBX 2/4/8 were highly expressed in pan-cancers. The expression level of CBX6 was heterogeneous among different tumor types, while that of CBX7 was low. The results show that CBX2, CBX4, and CBX8 are highly expressed in pan-cancer, and CBX6 is heterogeneous among different tumor types, while CBX7 is lowly expressed in almost tumor types. The expression of CBXs was correlated with the overall survival of patients, but the correlation direction depended on the cancer type. CBXs were significantly correlated with immune infiltration subtypes (P< 0.001). The expression of CBX2/CBX7s had differences in stromal cell score and RNAss stemness score. High expression of CBX82 was correlated with the high purity of most types of cancers, while CBX7 was contrary to this pattern. Rich analysis indicated that CBXs may affect the occurrence and progression of tumors by regulating RNA splicing and transcription coactivator activity. The core genes of hsa-mir-181a and EGR1 were identified in the established mRNA-TF-miRNA WGCNA network. Finally, our study revealed that CBX2 gene may be associated with sensitivity to Acrichine, Curcumin, Nelarabine, Ixabepilone, Ifosfamide, tfdu and Tamoxifen. The high expression of CBX7 was positively correlated with the drug resistance of cells to Aminoflavone, AFP464 and Lificguat.

Conclusion: The CBX family showed different expression in different tumor types. CBX2 acted as an oncogene and might be overexpressed in a variety of human cancers, and CBX7 acted as a tumor suppressor gene and might become a promising therapeutic target. This study revealed multiple expression patterns of CBXs in pan cancer, providing new insights for cancer treatment, but further laboratory validation is needed.

Introduction

The Cancer Genome Atlas (TCGA) launched "Pan-Cancer" in 2012, aiming to integrate different cancer types in TCGA dataset and analyze the commonness and heterogeneity among different cancers[1]. Chromobox (CBX) represents an important member of Polycomb group proteins (PcG)[2]. Five CBX family
members are identified in the human genome, including CBX2, CBX4, CBX6, CBX7 and CBX8\textsuperscript{[2]}. CBX protein family plays a vital role in tumor initiation and progression by mediating the differentiation and self-renewal of tumor stem cells\textsuperscript{[3]}. CBX family is considered as a prognostic biomarker of various types of cancer, including rectal cancer\textsuperscript{[4]} bladder cancer\textsuperscript{[5]}, kidney cancer\textsuperscript{[6]} and ovarian cancer\textsuperscript{[7]} Still, the research on CBX gene is limited within a few cancer types with a single gene, and there is no systematic research on CBX family genes in different human cancers. With the rapid development of high-throughput genomic technology, emerging bioinformatics tools have brought new opportunities for cancer research. Pan cancer analysis aims to examine the similarities and differences of genomic and cellular changes found in different cancer types. It prompted us to systematically analyze the role of CBXs from a pan-cancer perspective.

1 Methods

1.1 Data source

The gene expression data of 33 types of cancer (HTSeq-FPKM) in TCGA, clinical characteristic data (including survival data and expression data) and the stemness score data based on immune subtypes were downloaded from the websites https://xena.ucsc.edu/. The relevant data of CellMiner database was downloaded from its official website (https://discover.nci.nih.gov/cellminer/loadDownload.do). We also utilize the Sangerbox online platform (http://sangerbox.com/) to assist in multi-dimensional analysis.

1.2 Expression of pan oncogene

In this study, CBX family was selected to analyze its expression in pan cancer. Only 22 of 33 tumor types had three or more normal tissue samples, so 22 tumors were first differentially analyzed for gene expression, the rma function in the R software (version 3.4.1, http://www.r-project.org) was used to filter the entire data collection, delete missing and duplicate results, and convert via log2 (TPM +1); ggpubr package of R software was used to analyze the differential expression of CBX between tumor and normal tissues, and unpaired Wilcoxon Rank Sum and Signed Rank Tests were used to perform differential significance analysis on these tumor types, P < 0.05 was considered to indicate differential expression between tumors and normal tissues. This study also examined the genetic alterations in the CBXs gene in cancer patients by using the cBioPortal online database (https://www.cbioportal.org).

It includes 33 cancer types: ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; DLBC, lymphoid neoplasm diffuse large B-cell lymphoma; 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;
MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, colon adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous Melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal Melanoma.

1.3 Survival analysis

According to the median gene expression value, the patients were divided into high and low expression groups to determine whether gene expression was correlated with survival time and survival status. The clinical correlation analysis of different tissue types of tumors in the three stages of clinical staging was carried out. The forest plot of hazard ratio (HR) and 95% confidence interval (CI) of overall survival showed the survival advantages (HR < 1) and disadvantages (HR > 1) of CBX genes in different cancer types. Significantly different cancer types used the Kaplan–Meier plotter method and log-rank test for survival analysis for each cancer type (P< 0.05). One-way Cox proportional hazard regression model was used for correlation test, and the P< 0.05 meant the statistical significance.

1.4 Analysis of immune subtypes

TCGA dataset identifies 6 immune subtypes of tumor types, corresponding to tumor promotion and tumor inhibition, including C1 (wound healing), C2 (IFN-blocking dominance), C3 (inflammation), C4 (lymphocyte depletion), C5 (immune quiet), C6 (TGF-multinuclear dominance) subtypes. The distribution of immune subtypes in tumors is different, and each immune subtype shows different biological and clinical characteristics, which determines the effect of anti-tumor treatment to a certain extent. TISIDB[8] (http://cis.hku.hk/TISIDB/index.php) was used to analyze the expression of CBX2, CBX4, CBX6, CBX7 and CBX8 in different immune subtypes, so as to understand how each member of CBX family in TGCA tumor type was correlated with immune components. Kruskal test was used to analyze the correlation between CBX expression and immune subtypes.

1.5 Correlation analysis of tumor microenvironment (TME) and tumor stem cells

TME is mainly composed of tumor cells, immune cells, inflammatory cells and fibroblasts[10]. Stromal-score, immune-score and estimate-score[11] indicate the tumor purity, and the sum of the former two is equal to the latter. It is estimated that the higher the estimate-score, the higher the content of stromal cells and immune cells in TME and the lower the content of tumor cells. Estimate algorithm was used to study the correlation between the expression of CBX and the infiltration of stromal cells and immune cells in tumor[12]. To further understand the regulatory role of CBXs in tumor immunity. According to the gene expression, the deconvo mcpcounter[13] method of the R software package “IOBR”[14] was used to evaluate the infiltration scores of T cells, CD8 T cells, Cytotoxic lymphocytes, B lineage, NK cells,
Monocytic lineage, Myeloid dendritic cells, Neutrophils, Endothelial cells, and Fibroblasts in tumor patients.

1.6 Drug sensitivity analysis

To explore the correlation between gene expression and drug response in tumor cells, we used CellMiner (https://discover.nci.nih.gov/cellminer/) to access the NCI-60 database containing data of 60 different cancer cell lines from 9 different cancers. The data (GI50) of gene expression and cell sensitivity of 263 drugs approved by FDA or clinical trial drugs were retrieved, and Z-score was performed to explore the correlation between gene expression and drug sensitivity. The high Z-score, that is, the increase of gene expression is corresponded to good drug response, indicates that the cells were sensitive to drugs, and vice versa.

1.7 Statistical analysis

All data on gene expression were converted to expression data by log2(TPM+1). The comparison between normal tissue and cancer tissue was performed by two-group t test; P<0.05 indicated statistical significance. Kaplan-Meier curves, log-rank test, and Cox proportional hazards regression models were used for all survival analyses in this study. Correlation analysis between two variables was performed using Spearman's or Pearson's test; P< 0.05 was considered significant.

2 Results

2.1 Differential expression and co-expression analysis of CBX between tumor and normal samples

First, the differences in CBXs mRNA expression between 22 tumor samples and normal tissues were compared. It was then determined whether they were upregulated or downregulated in tumor tissues. Specifically, CBX2/CBX4/CBX8 were highly expressed in most types of tumors. CBX7 was the only gene commonly downregulated in CBX family. However, CBX6 was inconsistently upregulated or downregulated in different cancers. The expression of CBXs caused a proportional predominance of up-regulation or down-regulation in pan-cancer, suggesting that they may play a role in tumor development. cBioPortal online database is available to understand CBXs genetic alterations, The spectrum of genetic alterations in CBX2/CBX4/CBX8 suggests that its amplification was one of the most important single factors for alteration in SKCM,BRCA,OV,LIHC,UCS,MESO,etc.In addition,CBX6/CBX7 have less mutation and amplification. Results are shown in Figure 1.

2.2 Correlation analysis of CBXs expression and survival

The distinct differential expression pattern of CBXs in pan-cancer prompted us to explore their prognostic value. We further tested whether the expression of CBX could predict the survival of patients. For survival analysis, all cancer types were analyzed by Kaplan-Meier. According to the median value of gene expression, patients were divided into two groups: high and low expression groups. The differential
expression was expressed by P value, and the P < 0.05 meant significant difference. The results of P value were shown in Table 1. According to Figure 2, The high expression of CBX2 in ACC, KIRP, MESO and CBX4 in ACC, KIRC and KIRP predicted poor prognosis. In contrast, CBX6 and CBX8 show inconsistent prognostic outcomes in different cancers. CBX7 was highly expressed in pan cancer and showed better prognosis. In the analysis of clinical stage correlation (Figure 3), it was found that compared to stage I, CBX2 and CBX4 were expressed more in stage II and III patients (P < 0.05). It indicates that the increase in its possible expression may lead to poor survival results. The expression of CBX7 decreased with the increase of clinical stage. Cox proportional hazard regression model was used to measure the prognostic value of CBXs in 33 types of cancers. CBX with HR > 1 was considered as a risk factor for this type of cancer, which was unfavorable to the prognosis. The forest map (Figure 4) showed that CBX4 had pan cancer significance in ACC, DLBC, ESCA, KICH, KIRC and KIRP, HR > 1. CBX2 in DLBC and THCA, CBX6 in PCPG, CBX7 in PCPG and PRAD, and CBX8 in TGCT and THYM were considered to be favorable to the survival, HR < 1. In this part, CBX6 does not have uniform consistency in pan-cancer, considering that it may not have the potential of tumor markers. In some cancer types especially exhibited recurrent negative associations between high CBX2 mRNA levels and lower survival, and CBX7 predicted a better prognosis. Therefore, our study indicated that the majority of its high expression and poor prognosis suggest CBX2, possibly as an adverse factor, and CBX7 may be a protective factor in patients.

Table 1

<table>
<thead>
<tr>
<th>survival</th>
<th>P value result</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>CancerType</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBX2</td>
<td>ACC</td>
<td>2.69E-07</td>
</tr>
<tr>
<td>CBX2</td>
<td>BRCA</td>
<td>0.04319</td>
</tr>
<tr>
<td>CBX2</td>
<td>KIRP</td>
<td>0.001561</td>
</tr>
<tr>
<td>CBX2</td>
<td>LAML</td>
<td>0.024321</td>
</tr>
<tr>
<td>CBX2</td>
<td>LGG</td>
<td>0.03992</td>
</tr>
<tr>
<td>CBX2</td>
<td>LIHC</td>
<td>0.001288</td>
</tr>
<tr>
<td>CBX2</td>
<td>MESO</td>
<td>0.000215</td>
</tr>
<tr>
<td>CBX2</td>
<td>OV</td>
<td>0.042056</td>
</tr>
<tr>
<td>CBX2</td>
<td>PCPG</td>
<td>0.035438</td>
</tr>
<tr>
<td>CBX2</td>
<td>THYM</td>
<td>0.031716</td>
</tr>
<tr>
<td>CBX4</td>
<td>ACC</td>
<td>0.012857</td>
</tr>
<tr>
<td>CBX4</td>
<td>KIRC</td>
<td>4.99E-08</td>
</tr>
<tr>
<td>CBX4</td>
<td>KIRP</td>
<td>0.001216</td>
</tr>
<tr>
<td>CBX6</td>
<td>KIRC</td>
<td>0.046466</td>
</tr>
<tr>
<td>CBX6</td>
<td>KIRP</td>
<td>0.029178</td>
</tr>
<tr>
<td>CBX6</td>
<td>LGG</td>
<td>2.68E-06</td>
</tr>
<tr>
<td>CBX6</td>
<td>PAAD</td>
<td>5.99E-05</td>
</tr>
<tr>
<td>CBX6</td>
<td>PCPG</td>
<td>0.005805</td>
</tr>
<tr>
<td>CBX6</td>
<td>READ</td>
<td>0.019745</td>
</tr>
<tr>
<td>CBX6</td>
<td>THYM</td>
<td>0.018514</td>
</tr>
<tr>
<td>CBX6</td>
<td>UVM</td>
<td>0.04051</td>
</tr>
<tr>
<td>CBX7</td>
<td>ACC</td>
<td>0.021448</td>
</tr>
<tr>
<td>CBX7</td>
<td>CESC</td>
<td>0.002526</td>
</tr>
<tr>
<td>CBX7</td>
<td>KIRC</td>
<td>0.001729</td>
</tr>
<tr>
<td>CBX7</td>
<td>KIRP</td>
<td>0.008273</td>
</tr>
<tr>
<td>CBX7</td>
<td>LGG</td>
<td>0.002094</td>
</tr>
<tr>
<td>CBX7</td>
<td>LUAD</td>
<td>0.024615</td>
</tr>
<tr>
<td>CBX7</td>
<td>MESO</td>
<td>0.000256</td>
</tr>
</tbody>
</table>
### 2.3 Correlation analysis of CBX gene and immune response

Figure 5 showed the expression of CBX in different subtypes of immune infiltration. Among the 6 immune subtypes, C3 and C5 subtypes were associated with better survival rate, while C4 and C6 subtypes were on the contrary (15). Different CBX genes had different correlation patterns with different immune subtypes. The expression of CBX6 and CBX7 were increased in C5 and decreased in C1, C2 and C6, suggesting that higher expression of CBX6 and CBX7 might have better prognosis. The highly expressed genes were correlated with the favorable immune components, which indicated that these genes might play an anti-tumor role to some extent. On the contrary, CBX2 was highly expressed in C1 and C2, suggesting that CBX2 played a pro-tumor role. There was no significant difference in the expression of CBX4 among all 6 subtypes. The different expression patterns of CBX in different immune subtypes indicated that the function of each gene might be immune subtype dependent. Although the relationship of CBXs in them varies by tumor types, at least we observed that CBX2 and CBX7 have different roles in tumorigenesis, CBX2 may act as a covariate of patient survival and promote the aggressive progression of the disease, while CBX7 may be a tumor suppressor gene. CBX4, CBX6, and CBX8 cannot determine their homogeneous properties due to the incomprehensible laws in tumor types. Combined with the previous performance, in the following studies, we will focus on CBX2 and CBX7 to discuss the impact of their different expressions in tumors.

### 2.4 Stemness Indices and Microtumor Environment in Pan-Cancer

An increasing number of reports indicate that the TME plays a crucial role in tumor initiation and progression. Therefore, it is important to further explore the pan-cancer relationship between TME and CBXs expression. According to the estimation algorithm, a correlation matrix was drawn to show the relationship between the expression level of CBX family genes and the stromal scores of 33 different types of cancer. Spearman correlation analysis was generated. As shown in Figure 6A-C, the positive
correlation between the CBX gene and stromal cells and immune cells in tumors is shown in red. It can be seen from the graph that the CBX family genes are negatively correlated with the three scores in most tumors. This means that, the higher the gene expression, the lower the content of stromal cells and immune cells. Furthermore, relatively speaking, tumor cells have increased. Our results demonstrate that CBX2 expression is significantly negatively correlated with immune scores in GBM, UCEC, CESC, LUAD, ESCA, SARC, STAD, HNSC, LUSC, SKCM, BLCA, THCA, TGCT, LAML and ACC, but except in PRAD, BRCA, THYM and KICH. According to the R and P values, The correlation between CBX2 or CBX7 expression in the top 4 tumors and the three scores is shown in Figure 7. In this study, we found that the expression level of CBX2 and CBX7 gene was significantly correlated with immune cell infiltration of T cells, NK cells, Monocytic lineage, The latter was also significantly associated with Neutrophils, Endothelial cells infiltrating cells. As shown in Figure 8, CBX2 could at least reflect immune status of GBM, SARC, LUAD, LUSC, etc. CBX2 reduces the number of immune cells, promotes the increase in the number of tumor cells, and may also shape an inflamed TME for tumor progression, while CBX7 can increase the number of immune cells, contribute to immune surveillance, and reduce tumorigenesis.

### 2.5 Correlation analysis of CBX gene and drug sensitivity

Figure 9 showed the scatter diagram of the correlation between CBX gene expression and drug sensitivity in P value order. The abscissa was the gene expression, and the ordinate was the Z-score. P < 0.05. As shown in Figure 9, CBX2 was positively correlated with the sensitivity of Acrichine (cor = 0.428, P < 0.001), Curcumin (cor = 0.418, P < 0.001), Nelarabine (cor = 0.412, P < 0.001), Ixabepilone (cor = 0.408, P < 0.001), Ifosfamide (cor = 0.386, P < 0.001), tfdu (cor = 0.366, P = 0.004) and Tamoxifen (cor = 0.365, P = 0.004), but was positively correlated with the resistance of Dasatinib (cor = -0.437, P < 0.001) and Staurosporine (cor = -0.673, P < 0.001). The high expression of CBX7 was positively correlated with the drug resistance of cells to Aminoflavone (cor = -0.482, P < 0.001), Afp464 (cor = -0.399, P = 0.002) and lifigquat (cor = -0.389, P = 0.002). We also noted that CBX2 and CBX4 were correlated with the drug resistance of cells to Dasatinib and Staurosporine.

### Discussion

CBX protein constitutes one of the important members of polycomb family. Polycomb protein family is an epigenetic regulatory complex and exists in the form of Polycomb repressive complexes (PRCs)[16, 17], which plays a vital role in regulating cell differentiation, aging (inhibition of INK4/ARF sites), death, tumorigenesis and metastasis by modifying chromatin to inhibit the transcription of target genes[3, 17–19]. Previously, several studies have shown that aberrant CBXs expression were observed in certain single cancers tissues. CBX2[20] expression was upregulated in breast cancer whereas CBX7[21] was decreased. CBX4/CBX8 was elevated in Hepatocellular carcinoma[22, 23]. However, litter is understood in the expression of individual CBXs in pan-cancer.

This study was the first to provide a pan cancer analysis of CBX. We focused on the expression and prognostic value of CBX family in pan cancer. The results showed that CBX2, CBX4 and CBX8 were highly
expressed in most types of tumors, high expression brings poor prognosis. The expression of CBX6 in different tumors was heterogeneous, while the expression of CBX7 in tumor tissues was lower than that in normal tissues.

and those with high expression have a more advantageous clinical prognosis. The effect of CBX7 in different types of malignant tumors is opposite to that of CBX2. These findings are consistent with past observations\[5, 24-26\]. The above studies suggest that CBX7 may act as a tumor suppressor gene and can be considered as a new tumor suppressor gene and CBX2 may be a possible oncogene. Our research strengthens this evidence.

Immune cells in the tumor microenvironment help cancer cells evade immune surveillance and participate in cancer through a series of mechanisms. Anti-tumor immune cells are mainly composed of CD4/8+ T cells, natural killer cells, macrophages and neutrophils. Patients with high expression levels of CBX2 had a significantly reduced number of T cells, CD8 T cells, Cytotoxic lymphocytes, B lineage, NK cells, Monocytic lineage. Conversely, patients with higher levels of CBX7 expression had significantly increased infiltration of all six types. Our analysis suggests that CBX2/7 may have the ability to modulate immune infiltration. So, CBX2 may serve as a biomarker for cancer immune infiltration and poor prognosis. Anti-CBX2 therapy may be a valuable clinical strategy for the treatment of certain human cancers. While CBX7 contributes to immune surveillance, Jian Li's research supports this theory, it shows that CD4+ T cells express Cbx7, the latter prevents FasL (CD95L) expression and the activation-induced CD4+ T cell apoptosis\[27\].

Since regulating PcG function can inhibit tumor growth, people have been working to develop drugs for different PcG family members\[28\]. In this regard, chemical inhibitors of CBX are being developed, mainly focusing on three types of molecules: calixarene, FALKme3 and UNC\[29\]. The purpose of these molecules is to inhibit the reader function of the CBX protein or to dissociate the PRC1 complex. UNC3866\[30\] is a recently reported peptidyl chemical probe for PRC1. It can effectively inhibit the proliferation of PC3 cells by selectively binding to CBX7\[31\]. It is recommended that more specific PRC inhibitors can be developed in the future. The combined use of chemotherapy drugs such as platinum compounds and topoisomerase inhibitors increases the anti-tumor effect. Therefore, in the end, we used the Z-score method to predict the sensitive drugs of the CBX family. We found that CBX2 is more sensitive to Acrichine, Curcumin, Nelarabine, Ixabepilone, Ifosfamide, tfdu, and Tamoxifen, and may have clinical significance for anti-tumor therapy.

In conclusion, this study elucidated the clinical significance and prognostic value of CBX family in pan cancer. CBX2 acts as an oncogene in a variety of human cancers. The role of CBX7 in different types of malignant tumors is contrary to that of other CBX families, suggesting that CBX7 may act as a tumor suppressor. The role of other members of CBX in tumor varies with different tumor types. Although this study is the first multi-level analysis of CBX in pan cancer, it merely focuses on the potential mechanism of CBX for bioinformatics research and is only independent of TCGA dataset, which has not been confirmed by other additional datasets and experiments. CBX family plays a complex role in tumor
biology. Biostatistical correlation alone cannot clarify the interaction relationship, and more experimental and clinical evidence is warranted in the future.

**Abbreviations**

CBX, chromobox; PcG, Polycomb group proteins; r, Correlation coefficient; TCGA, The cancer genome atlas.

**Declarations**

**Conflict of interest**: The authors declare that they have no conflict of interest.

**Ethical approval**: This article does not contain any studies with human participants performed by any of the authors.

**Funding**

This study was supported by Science and technology project of Changsha (kq1907027), Scientific research project of Hunan Education Department (21C0260), National Natural Science Foundation of China (81973914 and 82104941).

**Data availability** These data are all from The Cancer Genome Atlas (TCGA) database (https://xena.ucsc.edu/). All relevant data are within the manuscript and its Supporting Information files.

**References**


Figures
CBXs expression levels and mutation frequency in different types of human cancers. The expression level of CBXs between tumor (the red portion) and normal tissues (The blue portion) were compared in twenty two cancer types based on the TCGA dataset (A). Consistent high expression level of CBX2/CBX4/CBX8 was observed in tumor samples of BLCA, BRCA, COAD, ESCA, HNSC, LUAD, LUSC, PRAD, STAD, THCA, UCEC than normal tissue based on both comparisons, CBX6 does not have uniform consistency in pan-cancer, considering that it may not have the potential of tumor markers, however, the expression of CBX7 was opposite in the above cancer. (ns:no significance, * P< 0.05, ** P < 0.01, *** P < 0.001.) The above data are available from the SangerBox online platform. The alteration frequency of CBXs were plotted in pan cancers (B). The alteration frequency included Mutation(green), Amplification(red), Deep Deletion(blue) or Multiple Iterations(grey). The above data are available from the cBioPortal online platform.
Figure 2

Kaplan-Meier survival curves comparing high and low expression of CBXs in different cancer types. Significantly different cancer types used the Kaplan–Meier plotter method and log-rank test for survival analysis for each cancer type (P< 0.05). CBX2 (OS in ACC(A), LIHC(B), BRCA(C)) and CBX4 (OS in KIRC(D), KIRP(E), ACC(F)) expression were negatively correlated with advanced stages of cancers. Instead, CBX7 (OS in LGG(J), CESC(K), SKCM(L)) and CBX8 (OS in PAAD(M), THYM(N), STAD(O)) expression were...
positively correlated with advanced stages of cancers. CBX6 (G-I) did not have the above consistency. OS, overall survival.

**Figure 3**

CBXs Correlation between gene expression and clinical stage. CBX2, CBX4 and CBX8 were expressed more in stage  and  patients (BLCA, COAD, READ, BRCA, LIHC, ESCA, LUAD, STAD, (A-H)) compared to patients that in stage  (P < 0.05). The expression of CBX7 decreased with the increase of clinical stage in BLCA(A), COAD(B), READ(C), BRCA(D), LUAD(G). CBX6 (A-H) did not have the above consistency. (Pairwise significance analysis using unpaired t-Test. ns: no significance, * P < 0.05, ** P < 0.01, *** P < 0.001.)

**Figure 4**

Cox proportional hazard analysis shows the hazard ratio (HR) of CBX in 33 cases of TCGA tumors. The forest map showed that CBX4 had pan cancer significance in ACC, DLBC, ESCA, KICH, KIRC and KIRP, HR
> 1. CBX2 in DLBC and THCA, CBX6 in PCPG, CBX7 in PCPG and PRAD, and CBX8 in PAAD, TGCT and THYM were considered to be favorable to the survival, HR < 1.

**Figure 5**

Analysis of immune subtypes. The abscissa represents the immune subtype, and the ordinate represents the expression of the CBX gene. The expression of CBX6 and CBX7 were decreased in C1, C2 and C6, suggesting that higher expression of CBX6 and CBX7 might have better prognosis. CBX2 was highly expressed in C1 and C2, suggesting that CBX2 played a protumor role. There was no significant difference in the expression of CBX4 among all 6 subtypes. *** P < 0.001;
Figure 6

Correlation analysis results between CBXs members and tumor microenvironment score. (A-C) The heat map shows the correlation between Stromal score, Immune score, Estimate score and CBXs mRNA expression. The size of the point represents the absolute value of the correlation coefficient. The larger the size, the higher the correlation. Note: The red dots represent positive correlation, and the blue dots represent negative correlation. To establish the association of Stromal, Immune and Estimate proportions with CBXs expression, Spearman correlation analysis of CBXs gene expression levels and matrix score was performed by the Estimation and Limma piece packages in the R package.

Figure 7

Correlation analysis between CBX2 or CBX7 expression and the tumor microenvironment in top 4 tumors. In GBM, SARC, OV and LUSC, CBX2 expression was significantly negatively correlated with immune cell infiltration, while in COAD, STAD, PRAD, READ, CBX7 expression was significantly positively correlated.

Figure 8
Relationship between CBX2/CBX7 and tumor infiltration of different immune cells with MCPCounter. The upper left triangle represents the correlation coefficient, RNA modification-related genes positively correlated with CBX2/CBX7 expression are marked in green. The lower right triangle represents the P-value. Correlation analysis between two variables was performed using Pearson's test; * P<0.05, ** P<0.01, and ***/*** P<0.001.

Figure 9

The correlation between CBXs expression and predicted drug response. CBX2 was positively correlated with the sensitivity of Acrichine, Curcumin, Nelarabine, Ixabepilone, Ifosfamide, tfdu, and Tamoxifen, but was positively correlated with the resistance of Dasatinib and Staurosporine. The high expression of CBX7 was positively correlated with the drug resistance of cells to Aminoflavone, AFP464 and Lificguat.