

Expression and prognostic significance of CBX2 in colorectal cancer: Database mining for CBX family members in malignancies

He Zhou

Affiliated Hospital of North Sichuan Medical College

Yongfu Xiong

Affiliated Hospital of North Sichuan Medical College

Zuoliang Liu

Affiliated Hospital of North Sichuan Medical College

Songlin Hou

Affiliated Hospital of North Sichuan Medical College

Tong Zhou (✉ zhoutong@nsmc.edu.cn)

Affiliated Hospital of North Sichuan Medical College <https://orcid.org/0000-0001-9294-8234>

Primary research

Keywords: Chromobox, Colorectal cancer, CBX2, Prognosis, Cell proliferation

Posted Date: April 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-352837/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Cancer Cell International on July 28th, 2021. See the published version at <https://doi.org/10.1186/s12935-021-02106-4>.

Abstract

Background: The Chromobox (CBX) domain protein family, a core component of polycomb repressive complexes 1, is involved in transcriptional repression, cell differentiation, and program development by binding to methylated histone tails. Each CBX family member plays a distinct role in various biological processes through their own specific chromatin domains, due to differences in conserved sequences of the CBX proteins. It has been demonstrated that colorectal cancer (CRC) is a multiple-step biological evolutionary process, whereas the roles of the CBX family in CRC remain largely unclear.

Methods: In the present study, the expression and prognostic significance of the CBX family in CRC were systematically analyzed through a series of online databases, including Cancer Cell Line Encyclopedia (CCLE), Oncomine, Human Protein Atlas (HPA), and Gene Expression Profiling Interactive Analysis (GEPIA).

Results: Most CBX proteins were found to be highly expressed in CRC, but only the elevated expression of CBX2 could be associated with poor prognosis in patients with CRC. Further examination of the role of CBX2 in CRC was performed through several *in vitro* experiments. CBX2 was overexpressed in CRC cell lines via the CCLE database and the results were verified by RT-qPCR. Moreover, the knockdown of CBX2 significantly suppressed CRC cell proliferation and invasion. Furthermore, the downregulation of CBX2 was found to promote CRC cell apoptosis.

Conclusion: Based on these findings, CBX2 may function as an oncogene and potential prognostic biomarker. Thus, the association between the abnormal expression of CBX2 and the initiation of CRC deserves further exploration.

Background

Colorectal cancer (CRC) remains one of the top three causes of tumor-related deaths worldwide [1]. In 2019, an estimated 145,600 new cases of CRC and 51,020 deaths from the disease were reported in the USA [2]. Although the incidence of CRC has decreased in developed countries, due to improved treatments and early screening, the 5-year overall survival (OS) rate of CRC patients in developing countries is still not ideal [3, 4]. The TNM staging system developed by the American Joint Committee on Cancer and Union for International Cancer Control is the most universal tumor staging standard in the world and an important reference index for predicting the prognosis of tumor patients [5]. Increasing evidence has indicated that CRC is a highly heterogeneous disease with multiple molecular pathways involved in its progression [6]. Nevertheless, current TNM staging systems cannot reflect the intrinsic biological heterogeneity of CRC, especially in patients with atypical early symptoms. This results in <50% of CRC being diagnosed early, with certain CRC patients diagnosed through this system already presenting with distant metastasis [7]. Therefore, to improve the diagnosis, prognosis and targeted therapy of CRC, it is necessary to find biomarkers that can accurately predict its progression and therapeutic effect, and explore the mechanism of CRC development at the molecular level.

Chromatin has been divided into different domains according to the function of associated genomes, including euchromatin and heterochromatin [8]. The structural distribution and assembly of chromatin are influenced by numerous factors. The proteins that control chromatin dynamics play a pivotal role in the epigenetic regulation of gene expression [9]. The Chromobox (CBX) family proteins are crucial components of chromatin-related complexes heterochromatin protein 1 (HP1) and polycomb (Pc), which are involved in transcriptional regulation, chromatin structural modification, and the cell development process [10]. To date, eight members of the CBX family proteins have been identified in eukaryotic organisms, each containing a single N-terminal chromosomal domain[11]. The CBX family can be subdivided into two groups: One consisting of CBX1, CBX3, and CBX5, with a similar structural feature of HP1 homologs (HP1 α , HP1 β , and HP1 γ), and another made of Pc paralog proteins, known as CBX2, CBX4, CBX6, CBX7, and CBX8, which can recruit Pc repressive complexes 1 to maintain expression patterns of different genes during cell proliferation [12].

CBX family proteins are widely involved in a variety of biological process in all metazoans, including cell cycle control, induction of cell differentiation, and maintenance of pluripotency of embryonic stem cells [13]. Existing evidence has revealed that the dysregulation of CBX proteins results in numerous cell divisions that initiate cancer [14]. For instance, three isoforms of HP1 (CBX1, CBX3, and CBX5) act as organizers of pericentric heterochromatin in conjunction with H3K27me₃, which hinders cell cycle progression, leading to transcriptional activation, cell proliferation and cancer [15]. Recent studies have suggested that CBX2 and CBX6 act as oncogenes in hepatocellular carcinoma (HCC). The overexpression of both CBX2 and CBX6 is associated with poor prognosis in HCC patients [16, 17]. Xia *et al* [18] demonstrated that mutation of the CBX4 gene causes the transcriptional repression of proto-oncogenes, and can interact with CBX2 and Bmi-1 to alter pre-splicing mRNA. This ultimately causes an abnormal transformation of cells. Unlike other Pc family members, the role of CBX7 as a proto-oncogene or suppressor depends on its specificity in cells and tissues, as well as various epigenetic factors [19].

CBX2 is a key regulator of developmental genes. It shows a stronger effect on cancer progression than other CBX members by repressing the transcription of the Ink4a/Arf locus [20]. Further research has shown that CBX2 is the main protein expressed in ESC and can repress the expression of pluripotency genes that promote stem cell differentiation [20, 21]. Although CBX2 has been reported to be abnormally expressed in a number of cancer types, the role of CBX2 in CRC remains largely unclear. In the present study, integrated analysis of the CBX protein family was performed through several online databases, with the purpose of searching for potential therapeutic biomarkers of CRC patient survival.

Methods

Cancer Cell Line Encyclopedia (CCLE) database analysis

CCLE (<https://portals.broadinstitute.org/ccle/home>) is an open-access database covering large-scale deep sequencing information of 947 human cancer cell lines from >30 varieties of tissue sources. The mRNA expression of the CBX family in cell lines derived from different tumor types was analyzed by

CCLE, to deepen the understanding of DNA mutations, gene expression and chromosome copy number information for specific genes. Gene expression data of the CBX family was downloaded directly from the CCLE website. According to the website, raw microarray data of CRC cell lines were converted to a single value for each probeset using the Robust Multi-array Average algorithm and quantile normalization.

Oncomine database analysis

Oncomine (<http://www.oncomine.org>) is a classic oncogene chip database, which integrates data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus databases. It provides a variety of analytical tools and visually shows the differences between cancer and normal tissue expression, co-expression analysis, mutation analysis, etc. In the current study, the mRNA expression of distinct CBX family proteins was analyzed between tumor and normal tissues in different types of cancers. The results were filtered using the following threshold: Fold change, 2; $P=1\times 10^{-4}$; gene rank, top 10%.

The Human Protein Atlas database (HPA) analysis

The HPA database (<https://www.proteinatlas.org/>) provides information on the tissue and cell distribution of all 26,000 human proteins and is publicly available free of charge [22]. This database uses special antibodies and immunohistochemical (IHC) techniques to examine the distribution and expression of each protein in 48 normal human tissues, 20 tumor tissues, 47 cell lines and 12 blood cells. These samples were collected from different clinical individuals, ensuring that the staining results were sufficiently representative. In the present study, IHC images of the CBX protein expression in clinical samples of patients with CRC and normal tissues were obtained from the HPA database.

Gene Expression Profiling Interactive Analysis (GEPIA) database analysis

GEPIA (<http://www.gepia.cancer-pku.cn/>) is an open-access database for the interactive exploration of multiple cancer genomics datasets. The utilization of GEPIA significantly reduces the barriers in accessing complex genomic data and facilitates rapid, intuitive, high-quality access to molecular profiling and clinical prognostic correlations for large-scale cancer genomics projects. The website query interface is combined with multiple databases that store DNA copy numbers and mRNA expression, gene coexpression, Reverse Phase Protein Array, DNA methylation and clinical survival data. This allows researchers to investigate the interaction between gene alterations and clinical case samples by directly entering gene names. Kaplan–Meier plots were used to compare OS and disease-free survival (DFS) in CRC cases with the mRNA expression of each CBX member.

Cell culture

The CRC HCT116 and HT29 cell lines, and the normal human colon mucosal epithelial cell line NCM460, were obtained from the Cell bank of the Chinese Academy of Science. Cells were cultured in McCoy's 5A medium (Gibco, Carlsbad, CA) containing 10% fetal bovine serum (PAN-Biotech, Adenbach, Bavaria) in an incubator with 5% CO₂ at 37°C.

Lentivirus transduction and stable cell line selection

The lentivirus-mediated GV248 vector (Genechem, China) was used to express short hairpin RNA (shRNA) targeting CBX2. The shRNAs sequences targeting CBX2 were as follows: 5'-GAG GTC AAC CCA GGA GAG AGA-3' (sense) and 5'-ACC TAA ATA TCC ACT CAG TAT-3' (antisense). Lentiviral particles were transfected into 293T cells with GV248-shCBX2 constructs. The stable cell lines were selected using puromycin (4µg/ml) 72 h after transfection. The medium containing puromycin was replaced every three days for two weeks.

RNA isolation and RT-qPCR

Total RNA was extracted from cells using RNAiso Plus reagent (Takara, Dalian, China). The Prime Script RT Reagent Kit (Takara) was used to perform reverse transcription following the manufacturer's protocol. RT-qPCR was performed on a Bio-Rad CFX96 system (Bio-Rad, Hercules, CA). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a normalizing control. The primer sequences used in the present study were as follows: CBX2, 5'-GGA ACA TGA GAA GGA GGT GCA G-3' (forward) and 5'-GAA GAG GAG GAA CTG CTG GAC T-3' (reverse); GAPDH, 5'-CAG GAG GCA TTG CTG ATG AT-3' (forward) and 5'-GAA GGC TGG GGC TCA TTT-3' (reverse).

Western blotting

Cells were harvested at a density of >90% and subsequently lysed in RIPA buffer (Beijing Solarbio Science & Technology, China) on ice. Equal amounts of protein lysate (15µg) were separated using 10% sodium dodecyl sulfate polyacrylamide gels and then transferred onto polyvinylidene fluoride membranes (Millipore, County Cork, Ireland) via electroblotting. Following blocking in rapid block buffer (Sangon Biotech, China) for 15 mins, the membranes were incubated at 4°C overnight with relevant primary antibodies. The Horse Radish Peroxidase-conjugated secondary antibodies were then used to incubate the membranes for 1 h at room temperature. The primary antibodies used were rabbit anti-CBX2 (dilution, 1:2,000, cat. no. ab80044; Abcam), and rabbit anti-GAPDH (dilution, 1:1,000, cat. no. 2118; Cell Signaling Technology).

Cell colony formation and invasion assays

Colony formation assay was used to test the cell proliferation capacity following CBX2 knockdown. Transfected cells were seeded in six-well plates in triplicate at a density of 500 cells/well and incubated at 37°C for two weeks. The cell colonies were then washed with PBS, fixed with 4% paraformaldehyde for 20 min and stained with 0.2% crystal violet for 30 min. Cell colonies with >50 cells were identified as positive colonies and the colony numbers were counted under a microscope.

Transwell assay was used to evaluate the invasion ability of colon cancer cells after knocking down CBX2. In brief, 1×10^5 cells were plated in the upper chamber coated of 24-well plate with Matrigel and supplemented with serum-free medium. The lower chamber was filled with culture medium containing

15% FBS. Incubation was carried out for 48 hours at 37°C. The noninvasive cells were scraped off with cotton swabs. The cells that had successfully translocated were fixed with 4% paraformaldehyde, stained with 0.5% crystal violet. The number of invaded cells was observed by using an inverted microscope and calculated by counting six random views.

Cell apoptosis assay

For the cell apoptosis assay, transfected cells were disaggregated using 0.25% trypsin-EDTA solution and washed in PBS twice. Next, 1×10^6 cells were resuspended in PBS and stained with annexin V and 4',6-diamidino-2-phenylindole, according to the manufacturer's recommendations.

Statistical analysis

The GraphPad Prism (Version 8.0 GraphPad Software, CA) was used for statistical analysis. The significance of differences between groups was evaluated using the Student's t-test. Statistical significance of CBX family expression between CRC and normal tissues from the Oncomine database was provided by the program. Survival data of CBX family mRNA expression were obtained from the GEPIA database. Survival curves were plotted using the Kaplan-Meier method and compared using the log-rank test. $P < 0.05$ were considered to indicate a statistically significant difference.

Results

Dysregulated mRNA expression of CBX family members in CRC

The expression of CBX2 in CRC ranked 21th highest among all cancer types, as determined by CCLE analysis (Fig. 1a). Oncomine database analysis showed that the mRNA levels of CBX2, CBX3, CBX4, CBX5, and CBX8 were significantly higher in CRC than in normal tissues, based on a wide variety of datasets. Conversely, CBX6 and CBX7 were confirmed to have a lower expression in CRC, as compared with normal tissue (Fig. 1b).

Association between CBX mRNA expression and pathological stages of CRC

Next, a large sample from the TCGA dataset was analyzed in the GEPIA database. Consistent with the Oncomine database, CBX1-5 and CBX8 transcripts in colon and rectal adenocarcinoma tissues were higher than in normal tissues (Fig. 2a-h). The association between CBX expression and CRC pathological stage was then investigated. Of note, significantly statistical differences between tumor stages I-IV were only identified in the CBX2 group ($P=0.021$; Fig. 3b). There was no association between the other CBX members and pathological stage ($P > 0.05$; Fig. 3a and c-h).

Protein expression of CBXs in patients with CRC

To further verify the trend of CBX expression in CRC tissues, the results of IHC analysis of CBXs were obtained from the HPA database (Fig. 4). According to the degree of staining, the protein expression of

CBX1-5 and CBX8 in CRC tissues was also higher than that in normal tissues. Conversely, CBX6 and CBX7 were lower in CRC tissues, as compared to normal tissues. These results were consistent with the mRNA expression.

Association between CBX family expression and prognosis in patients with CRC Further assessment by GEPIA database analysis of the prognostic effect of the CBX family mRNA expression in CRC revealed that a high CBX2 expression in patients was significantly associated with a worse DFS, as compared with a low CBX2 expression ($P=0.049$; Fig. 6b). However, no significant difference ($P > 0.05$) was observed in the OS and DFS associated with or without mRNA expression alteration in the remaining CBX family members (Fig. 5 and Fig. 6a and c-h). A comparison of the above databases indicated that CBX2 might be a potential prognostic target for CRC.

Upregulation of CBX2 in CRC cell lines

In numerous CRC cell lines, including HCT116 and HT29, the relative expression of CBX2 is higher (Fig. 7a). Therefore, HCT116 and HT29 cell lines were selected for subsequent analysis. To further validate the results of the CCLE, the expression levels of CBX2 between CRC and normal cell lines were detected by RT-qPCR. As shown in Fig. 7b, the mRNA expression of CBX2 was higher in the HCT116 and HT29 cell lines, as compared with the normal cell line.

CBX2 knockdown inhibited CRC cells proliferation and invasion

Accumulating evidence has demonstrated that cancer is closely associated with abnormal proliferation [23]. A large number of cases in the GEPIA database have suggested that an elevated CBX2 expression can lead to poor prognosis. We therefore wondered whether CBX2 was involved in maintaining the malignant phenotype of CRC cells. To explore the biological function of CBX2 in the tumorigenesis of CRC, we investigated whether CBX2-knockdown could inhibit cell proliferation. First, as shown in Fig. 7c, the protein expression of CBX2 was significantly inhibited following transfection by western blotting. Next, a colony formation assay indicated that CBX2-silencing led to a marked reduction of colony numbers in HCT116 and HT29 cell lines (Fig. 8a). Furthermore, flow cytometry, performed to determine whether apoptosis was involved in the CBX2-knockdown-induced inhibition of proliferation, showed apoptotic cells were significantly increased in the shCBX2-expressing group, as compared with the shCtrl group (Fig. 8b). In addition, as compared with the shCtrl group, transwell assays revealed a significant reduction in cell invasion in the shCBX2-expressing group (Fig. 8c).

Discussion

It has been well-established that the process of CRC initiation can be attributed to cumulative genomic mutations [24, 25]. Mutations of numerous oncogenes and tumor suppressor genes during the multiple-step evolution of CRC could lead to the transformation of tissues from normal epithelial to carcinoma [26]. The presence and diversity of these mutations are suspected to dictate a tumor's clinical course. However, the clinical value of mutation-based prognostic biomarkers in several studies is

inconsistent. For instance, a specific oncogene may be considered a bad feature in certain studies and a good feature in others [27-30]. With the rapid development of high-throughput sequencing technology, a large number of tumor transcript data containing clinical information have been accumulated, but most of them have not been analyzed. In the present study, a large sample size integrated analysis of CRC was performed using several large tumor databases to identify potential biomarkers that play important roles in CRC.

In addition to gene mutations that lead to the activation of proto-oncogenes or inactivation of tumor suppressor genes, abnormal alterations in epigenome modification also play a key role in the initiation and progression of a variety of cancers, including lung, breast and liver cancer, and CRC [31-34]. Several studies have indicated that endogenous and exogenous stimulation can reorganize the chromatin structure of cells, resulting in the expression or suppression of abnormal genes, allowing them to obtain the hallmarks of cancer [35, 36]. Therefore, the reversibility of epigenetic therapies for these changes has profound implications for the prevention and clinical prognosis of cancer patients [37].

CBX2, also known as CDCA6 or M33, is a crucial component of PcG histone complexes involved in epigenetic controls [38]. The PcG complexes are highly conserved in evolution and contain a variety of enzymes that catalyze histone modification to act as gene suppressors or activators [39]. Increasing evidence has suggested that phenotypic changes caused by histone posttranslational modification dysregulation are one of the pathogenic mechanisms of human carcinogenesis. These events are often described as the biological transformation of cellular molecular hallmarks into a malignant molecular phenotype process [40]. Epigenetic control has been considered a prior response to gene activity caused by changes in chromatin structure, which stems from self-maintenance, post-translational modification of mRNA, and binding between different histones [41, 42]. Hence, further research on the epigenome will greatly improve our understanding of the mechanisms of complex diseases, including cancer. For instance, Chen *et al* found that CBX2 was abnormally highly expressed in breast cancer, and an elevation that led to poor prognosis [43]. Clermont *et al* reported that CBX2 suppressed cell viability by inducing caspase-3 enzyme, which caused apoptosis in metastatic prostate cancer cells [44]. Further studies found the knockdown of CBX2 to facilitate the sumoylation activation of SUMO2/3, leading to the occurrence of leukemia [45]. Mechanistically, it was revealed by Mao *et al* [16] that CBX2 could activate the Hippo pathway via the downregulation of the YAP expression, thus regulating the proliferation, apoptosis and DNA repair of hepatocellular carcinoma cells. Moreover, CBX2 was also reported by Han *et al* [46] to act as a tumor promoter by binding miRNA let-7a to downregulate the expression of RAS, resulting in the progression of osteosarcoma.

In the present study, a total of eight CBX family members were evaluated in 20 common human samples and normal control tissues through the Oncomine database. The Oncomine database results showed that most members of CBX were highly expressed in CRC, implying their unique roles in the disease. The present analysis further confirmed this conclusion based on a large-sample TCGA cohort study. Intriguingly, in the survival analysis of CBX family members in patients with CRC, it was found that only a high CBX2 mRNA expression was associated with poor outcome, as compared with patients with a

low mRNA expression. These findings suggested that CBX2 may be a potential prognostic target. Further analysis of the CBX2 function found it to be highly expressed in CRC cell lines HCT116 and HT29, as compared to normal colon mucosal epithelial cell line. These results were consistent with the CCLE database. As mentioned above, CBX2 has been proven by numerous researchers to affect cell proliferation by binding to the Ink4a/Arf locus [47]. Consistent with previous observations, the present *in vitro* experiments indicated that CBX2-knockdown significantly inhibited cell proliferation and invasion in HCT116 and HT29 cell lines. In addition, Daub *et al*/performed a proteomics study to search for cellular targets in cancer cells, and found that CBX2 was involved in the cell cycle, whose dysregulation could induce apoptosis [48]. Through apoptosis assays, it was confirmed that the downregulation of CBX2 could markedly promote apoptosis in HCT116 and HT29 cell lines.

Even though the underlying molecular mechanism between the abnormal expression of CBX2 and CRC requires more in-depth research, this study clearly confirmed that CBX2 is upregulated in CRC and CBX2-overexpression is significantly associated with poor survival outcomes. In conclusion, the present study suggested that CBX2 could function as an oncogene and serve as a potential prognostic biomarker in CRC.

Abbreviations

CBX: Chromobox; RT-qPCR: CCLE: Cancer cell line encyclopedia; HPA: Human protein atlas database; GEPIA: Gene expression profiling interactive analysis; TCGA: The Cancer Genome Atlas.

Declarations

Authors' contributions

HZ and TZ designed study; ZLL researched literature; HZ and YFX performed experiments; SLH analyzed experimental results; HZ wrote the manuscript.

Funding

This work was supported by Youth Science Foundation Project, National Natural Science Foundation of China (82003147); Scientific Research Project of Affiliated Hospital of North Sichuan Medical College (2020ZD001, 2021ZD001); The Doctoral Research Startup Fund Project of North Sichuan Medical College (CBY20-QD03); Pre-research of State-level Project of North Sichuan Medical College (CBY19-YZ17); Free Exploring Basic Research Project of Science and Technology of Sichuan Province (2020YJ0186, 2021YJ0456); Scientific and Technological Cooperation Project of Nanchong City (19SXHZ0290, 18SXHZ0511).

Availability of data and materials

The datasets generated and analyzed during the current study are publicly available from the following online databases: CCLE (<https://portals.broadinstitute.org/ccle/home>); Oncomine database (<http://www.oncomine.org>); HPA database (<https://www.proteinatlas.org/>); GEPIA database (<http://gepia.cancer-pku.cn/>).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests in this work.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65(2):87-108.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019; 69(1):7-34.
3. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut.* 2017; 66(4):683-91.
4. Sigurdsson JA, Getz L, Sjönell G, Vainiomäki P, Brodersen J. Marginal public health gain of screening for colorectal cancer: modelling study, based on WHO and national databases in the Nordic countries. *J Eval Clin Pract.* 2013; 19(2):400-07.
5. Compton CC. Optimal pathologic staging: defining stage II disease. *Clin Cancer Res.* 2007; 13(2):6862s-70s.
6. Cuyle PJ, Prenen H. Current and future biomarkers in the treatment of colorectal cancer. *Acta Clin Belg.* 2017; 72(2):103-15.
7. Gonzalez-Pons M, Cruz-Correa M. Colorectal Cancer Biomarkers: Where Are We Now? *BioMed Res Int.* 2015; 2015:149014.
8. Li G, Reinberg D. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev.* 2011; 21(2):175-86.
9. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet.* 2011; 12(1):7-18.
10. Ma RG, Zhang Y, Sun TT, Cheng B. Epigenetic regulation by polycomb group complexes: focus on roles of CBX proteins. *J Zhejiang Univ SC B.* 2014; 15(5):412-28.
11. Wotton D, Merrill JC. Pc2 and SUMOylation. *Biochem SOC T.* 2007; 35(6):1401-04.

12. Vincenz C, Kerppola TK. Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *PNAS*. 2008; 105(43):16572-77.
13. Aloia L, Di Stefano B, Di Croce L. Polycomb complexes in stem cells and embryonic development. *Development*. 2013; 140(12):2525-34.
14. Gieni RS, Hendzel MJ. Polycomb group protein gene silencing, non-coding RNA, stem cells, and cancer. *Biochem Cell Biol*. 2009; 87(5):711-46.
15. Carone DM, Lawrence JB. Heterochromatin instability in cancer: from the Barr body to satellites and the nuclear periphery. *Semin Cancer Biol*. 2013; 23(2):99-108.
16. Mao J, Tian Y, Wang C, Jiang K, Li R, Yao Y, et al. CBX2 Regulates Proliferation and Apoptosis via the Phosphorylation of YAP in Hepatocellular Carcinoma. *J Cancer*. 2019; 10(12):2706-19.
17. Zheng H, Jiang WH, Tian T, Tan HS, Chen Y, Qiao GL, et al. CBX6 overexpression contributes to tumor progression and is predictive of a poor prognosis in hepatocellular carcinoma. *Oncotarget*. 2017; 8(12):18872-84.
18. Xia ZB, Anderson M, Diaz MO, Zeleznik-Le NJ. MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. *PNAS*. 2003; 100(14):8342-7.
19. Simhadri C, Gignac MC, Anderson CJ, Milosevich N, Dheri A, Prashar N, et al. Structure-Activity Relationships of Cbx7 Inhibitors, Including Selectivity Studies against Other Cbx Proteins. *ACS Omega*. 2016; 1(4):541-51.
20. van den Boom V, Rozenveld-Geugien M, Bonardi F, Malanga D, van Gosliga D, Heijink AM, et al. Nonredundant and locus-specific gene repression functions of PRC1 paralog family members in human hematopoietic stem/progenitor cells. *Blood*. 2013; 121(13):2452-61.
21. Morey L, Pascual G, Cozzuto L, Roma G, Wutz A, Benitah SA, et al. Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell*. 2012; 10(1):47-62.
22. Pontén F, Jirström K, Uhlen M. The Human Protein Atlas—a tool for pathology. *J Pathol*. 2008; 216(4):387-93.
23. Markert EK, Levine AJ, Vazquez A. Proliferation and tissue remodeling in cancer: the hallmarks revisited. *Cell Death Dis*. 2012; 3(10):e397.
24. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000; 100(1):57-70.
25. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009; 458(7239):719-724.
26. Ni Z, Wang X, Zhang T, Li L, Li J. Comprehensive analysis of differential expression profiles reveals potential biomarkers associated with the cell cycle and regulated by p53 in human small cell lung cancer. *Exp Ther Med*. 2018; 15(4):3273-82.
27. Guan JL, Zhong WZ, An SJ, Yang JJ, Su J, Chen ZH, et al. KRAS mutation in patients with lung cancer: a predictor for poor prognosis but not for EGFR-TKIs or chemotherapy. *Ann Surg Oncol*. 2013; 20(4):1381-88.

28. Marabese M, Ganzinelli M, Garassino MC, Shepherd FA, Piva S, Caiola E, et al. KRAS mutations affect prognosis of non-small-cell lung cancer patients treated with first-line platinum containing chemotherapy. *Oncotarget*. 2015; 6(32):34014-22.
29. Scoccianti C, Vesin A, Martel G, Olivier M, Brambilla E, Timsit JF, et al. Prognostic value of TP53, KRAS and EGFR mutations in nonsmall cell lung cancer: the EUELC cohort. *Eur Respir J*. 2012; 40(1):177-84.
30. Sun JM, Hwang DW, Ahn JS, Ahn MJ, Park K. Prognostic and predictive value of KRAS mutations in advanced non-small cell lung cancer. *Plos One*. 2013; 8(5):e64816.
31. Li J, Li WX, Bai C, Song Y. Particulate matter-induced epigenetic changes and lung cancer. *Clin Respir J*. 2017; 11(5):539-46.
32. Byler S, Goldgar S, Heerboth S, Leary M, Housman G, Moulton K, et al. Genetic and epigenetic aspects of breast cancer progression and therapy. *Anticancer Res*. 2014; 34(3):1071-7.
33. Babaei K, Khaksar R, Zeinali T, Hemmati H, Bandegi A, Samidoust P, et al. Epigenetic profiling of MUTYH, KLF6, WNT1 and KLF4 genes in carcinogenesis and tumorigenesis of colorectal cancer. *Biomedicine*. 2019; 9(4):22.
34. Yokomichi N, Nishida N, Umeda Y, Taniguchi F, Yasui K, Toshima T, et al. Heterogeneity of Epigenetic and Epithelial Mesenchymal Transition Marks in Hepatocellular Carcinoma with Keratin 19 Proficiency. *Liver Cancer*. 2019; 8(4):239-54.
35. Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science*. 2016; 352(6290):aad9780.
36. Lao VV, Grady WM. Epigenetics and colorectal cancer. *Nat Rev Gastro Hepat*. 2011; 8(12):686-700.
37. Barchitta M, Maugeri A, Li Destri G, Basile G, Agodi A. Epigenetic Biomarkers in Colorectal Cancer Patients Receiving Adjuvant or Neoadjuvant Therapy: A Systematic Review of Epidemiological Studies. *Int J Mol Sci*. 2019; 20(15):3842.
38. Morey L, Helin K. Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci*. 2010; 35(6):323-32.
39. Obuse C, Yang H, Nozaki N, Goto S, Okazaki T, Yoda K. Proteomics analysis of the centromere complex from HeLa interphase cells: UV-damaged DNA binding protein 1 (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase. *Genes Cells*. 2004; 9(2):105-20.
40. Parfett CL, Desaulniers D. A Tox21 Approach to Altered Epigenetic Landscapes: Assessing Epigenetic Toxicity Pathways Leading to Altered Gene Expression and Oncogenic Transformation In Vitro. *Int J Mol Sci*. 2017; 18(6):1179.
41. Herceg Z, Lambert MP, van Veldhoven K, Demetriou C, Vineis P, Smith MT, et al. Towards incorporating epigenetic mechanisms into carcinogen identification and evaluation. *Carcinogenesis*. 2013; 34(9):1955-67.
42. Holmberg J, Perlmann T. Maintaining differentiated cellular identity. *Nat Rev Genet*. 2012; 13(6):429-39.

43. Chen WY, Zhang XY, Liu T, Liu Y, Zhao YS, Pang D. Chromobox homolog 2 protein: A novel biomarker for predicting prognosis and Taxol sensitivity in patients with breast cancer. *Oncol Lett.* 2017; 13(3):1149-56.
44. Clermont PL, Crea F, Chiang YT, Lin D, Zhang A, Wang JZ, et al. Identification of the epigenetic reader CBX2 as a potential drug target in advanced prostate cancer. *Clin Epigenetics.* 2016; 8:16.
45. Di Costanzo A, Del Gaudio N, Conte L, Dell'Aversana C, Vermeulen M, de Thé H, et al. The HDAC inhibitor SAHA regulates CBX2 stability via a SUMO-triggered ubiquitin-mediated pathway in leukemia. *Oncogene.* 2018; 37(19):2559-72.
46. Han Q, Li C, Cao Y, Bao J, Li K, Song R, et al. CBX2 is a functional target of miRNA let-7a and acts as a tumor promoter in osteosarcoma. *Cancer Med.* 2019; 8(8):3981-91.
47. Jangal M, Lebeau B, Witcher M. Beyond EZH2: is the polycomb protein CBX2 an emerging target for anti-cancer therapy? *Expert Opin Ther Targets.* 2019; 23(7):565-78.
48. Daub H, Olsen JV, Bairlein M, Gnad F, Oppermann FS, Körner R, et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol Cell.* 2008; 31(3):438-48.

Figures

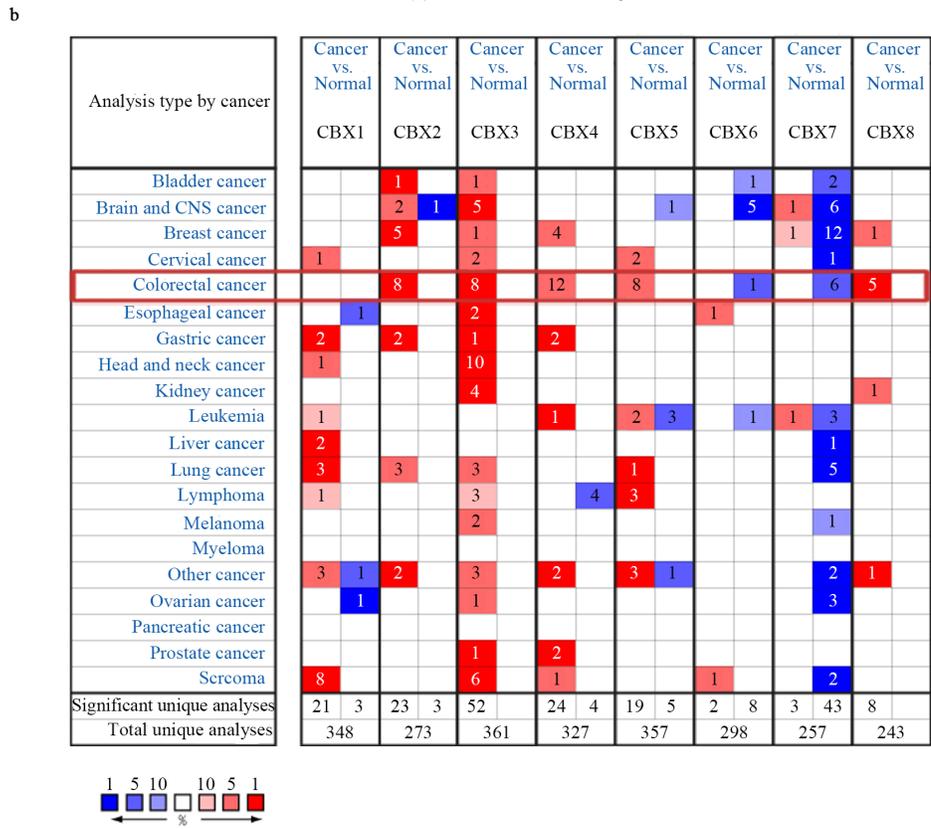
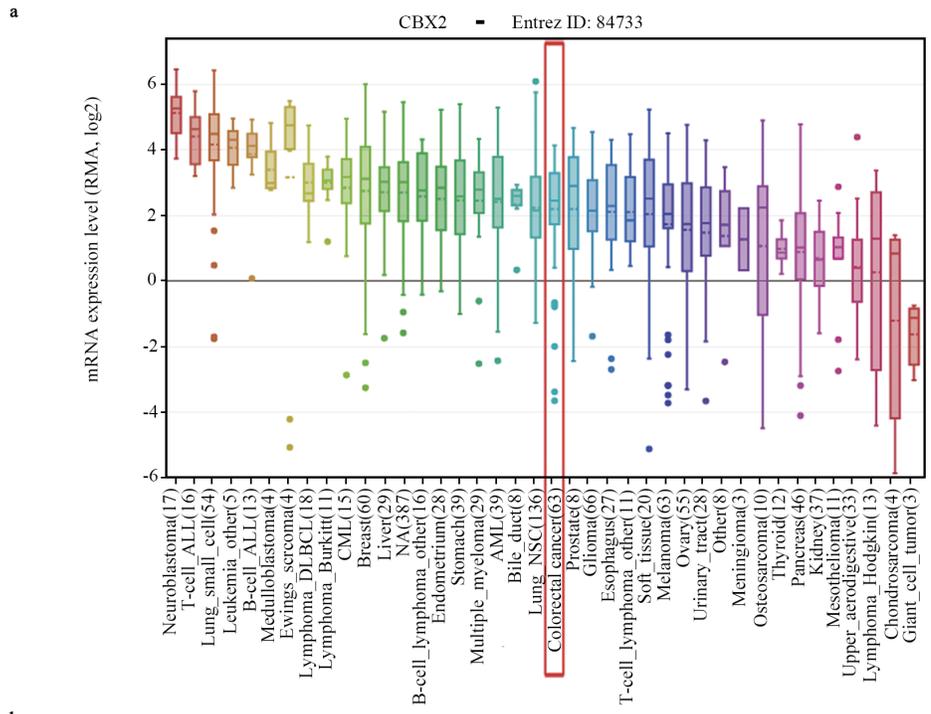


Figure 1

mRNA expression of CBX family members in different types of cancer. a mRNA expression level of CBX2 from the CCLE database. The CBX2 mRNA expression level ranked 21th among different human types of cancer (shown in red frame). b mRNA expression levels of CBX family members in various types of cancer vs. normal tissues in the Oncomine database. The blue box in the graph indicates that the target gene was lowly expressed in the corresponding tumor, while red indicates highly expressed genes, with

statistically significant differences ($P=1 \times 10^{-4}$). The number in the cell represents the number of studies that meet the set threshold. The color of the cells is determined by the rank of gene expression differences. CBX, Chromobox; CCLE, Cancer Cell Line Encyclopedia.

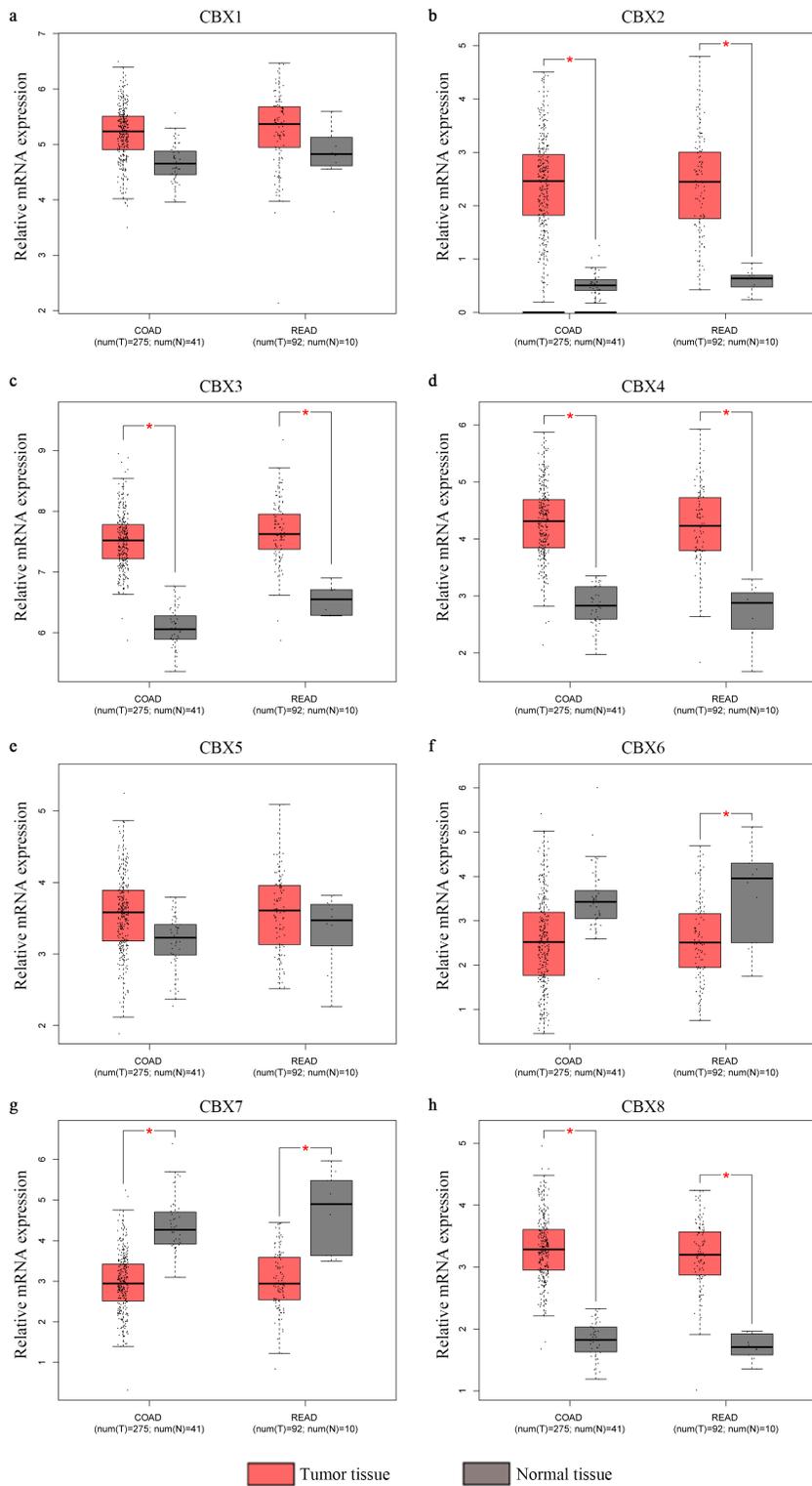


Figure 2

Expression analysis of CBX family members in CRC and normal tissues (GEPIA database). Box plots derived from gene expression data comparing expression levels of a specific CBX family member in CRC

and the corresponding normal tissue. Comparison of a CBX1, b CBX2, c CBX3, d CBX4, e CBX5, f CBX6, g CBX7, and h CBX8 mRNA expression. *P < 0.05. COAD, colon adenocarcinoma; READ, rectum adenocarcinoma; T, tumor; N, normal; CBX, Chromobox; CRC, colorectal cancer; GEPIA, Gene Expression Profiling Interactive Analysis.

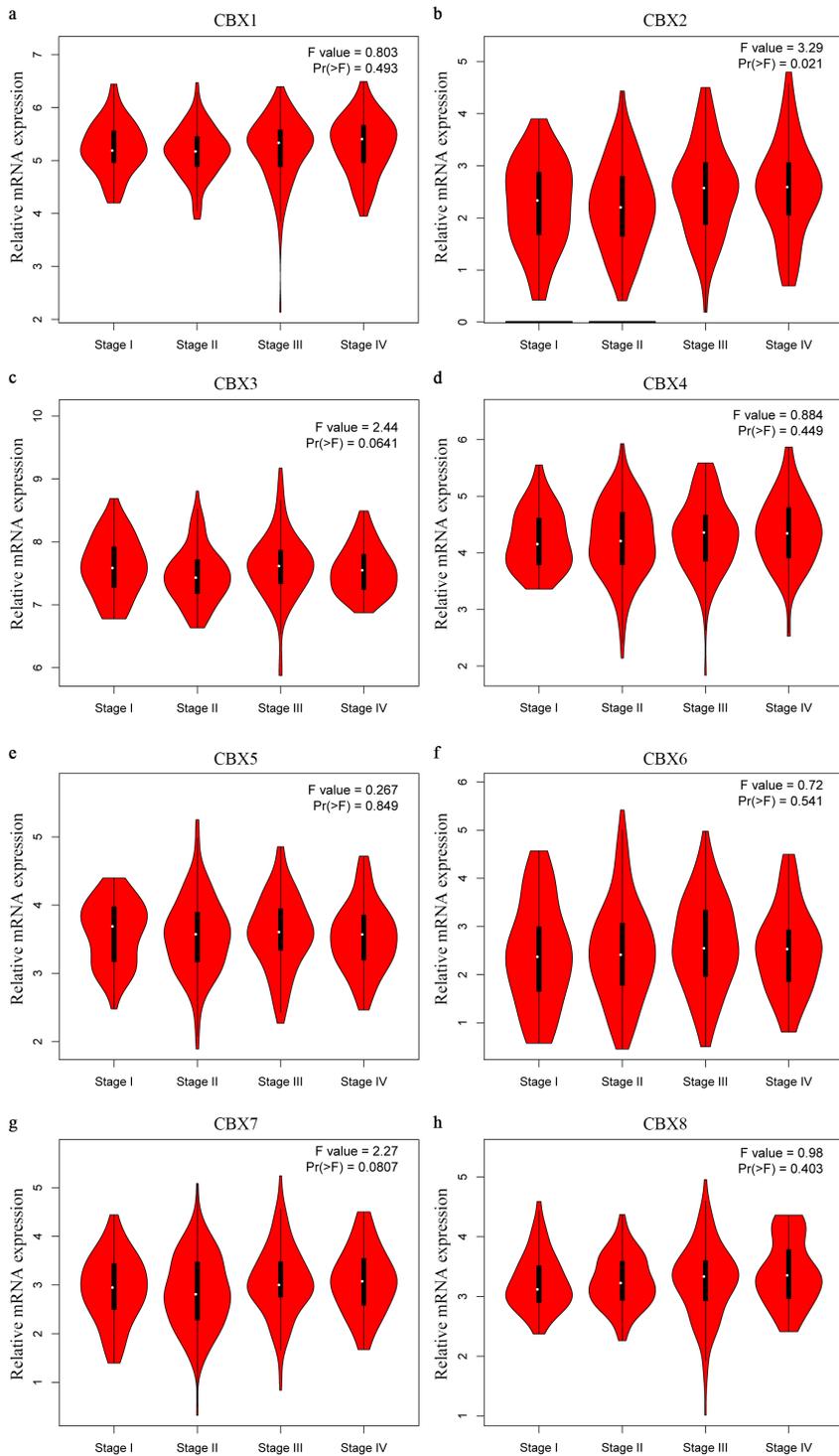


Figure 3

Association between mRNA expression of CBXs and tumor stages in patients with CRC (GEPIA database). a CBX1, b CBX2, c CBX3, d CBX4, e CBX5, f CBX6, g CBX7, and h CBX8. In the violin plots, the white dots indicate the median; the black box indicates the quartile range; the thin black line indicates 95% confidence interval; the size of the red area indicates the density. F-value, statistical value of the F test; Pr (>F), P-value. CBX, Chromobox; CRC, colorectal cancer; GEPIA, Gene Expression Profiling Interactive Analysis.

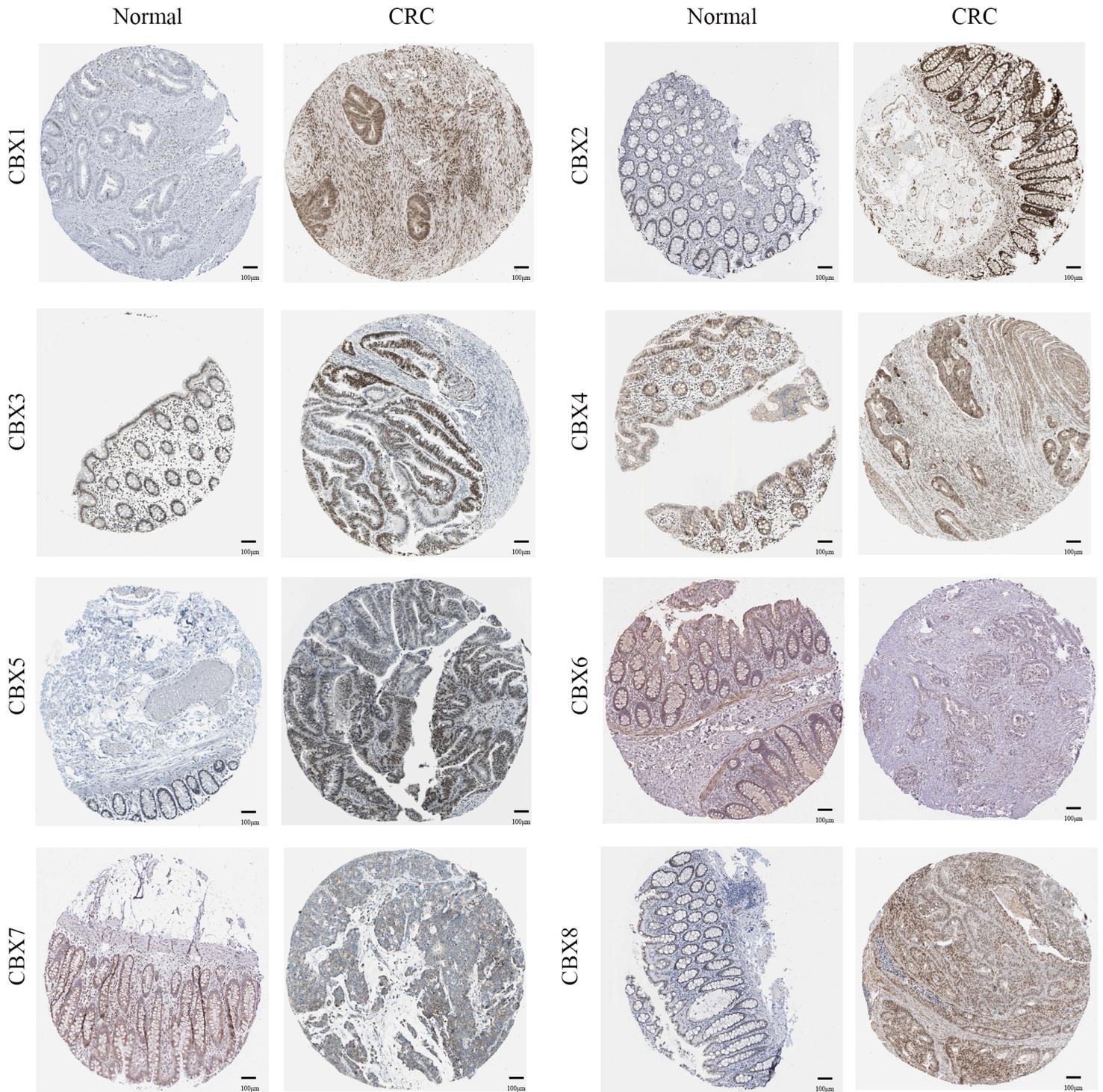


Figure 4

Immunohistochemical analysis of protein expression in CRC and normal tissues (The Human Protein Atlas Database). The brown areas represent positive expression and the blue negative. Scale bar, 100µm. CRC, colorectal cancer.

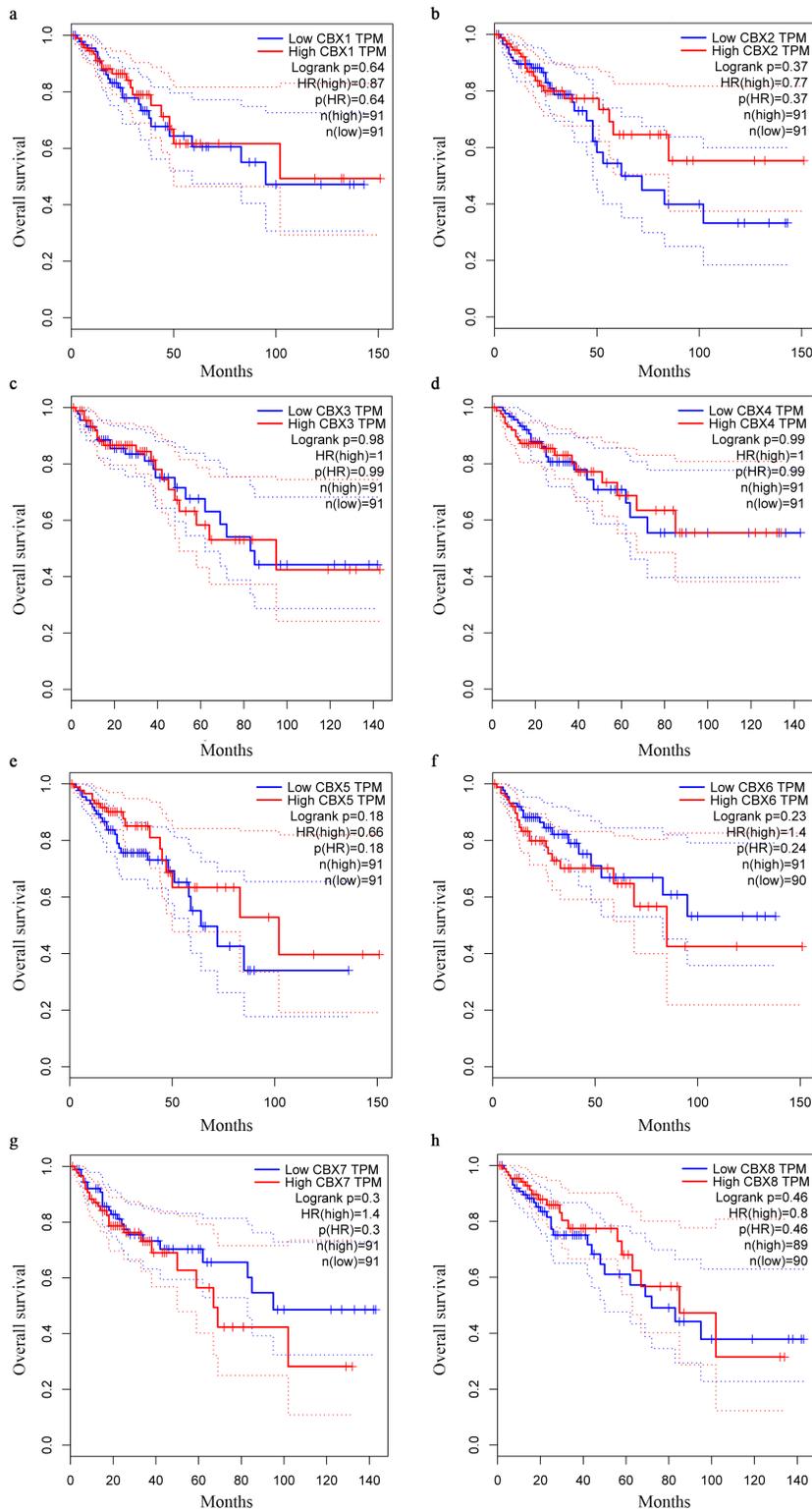


Figure 5

Survival analysis and prognostic values of CBX family in patients with CRC (GEPIA database). a-h OS curves of CBX1, CBX2, CBX3, CBX4, CBX5, CBX6, CBX7, and CBX8 in patients with CRC. OS, overall

survival; HR, hazard ratio; CRC, colorectal cancer.

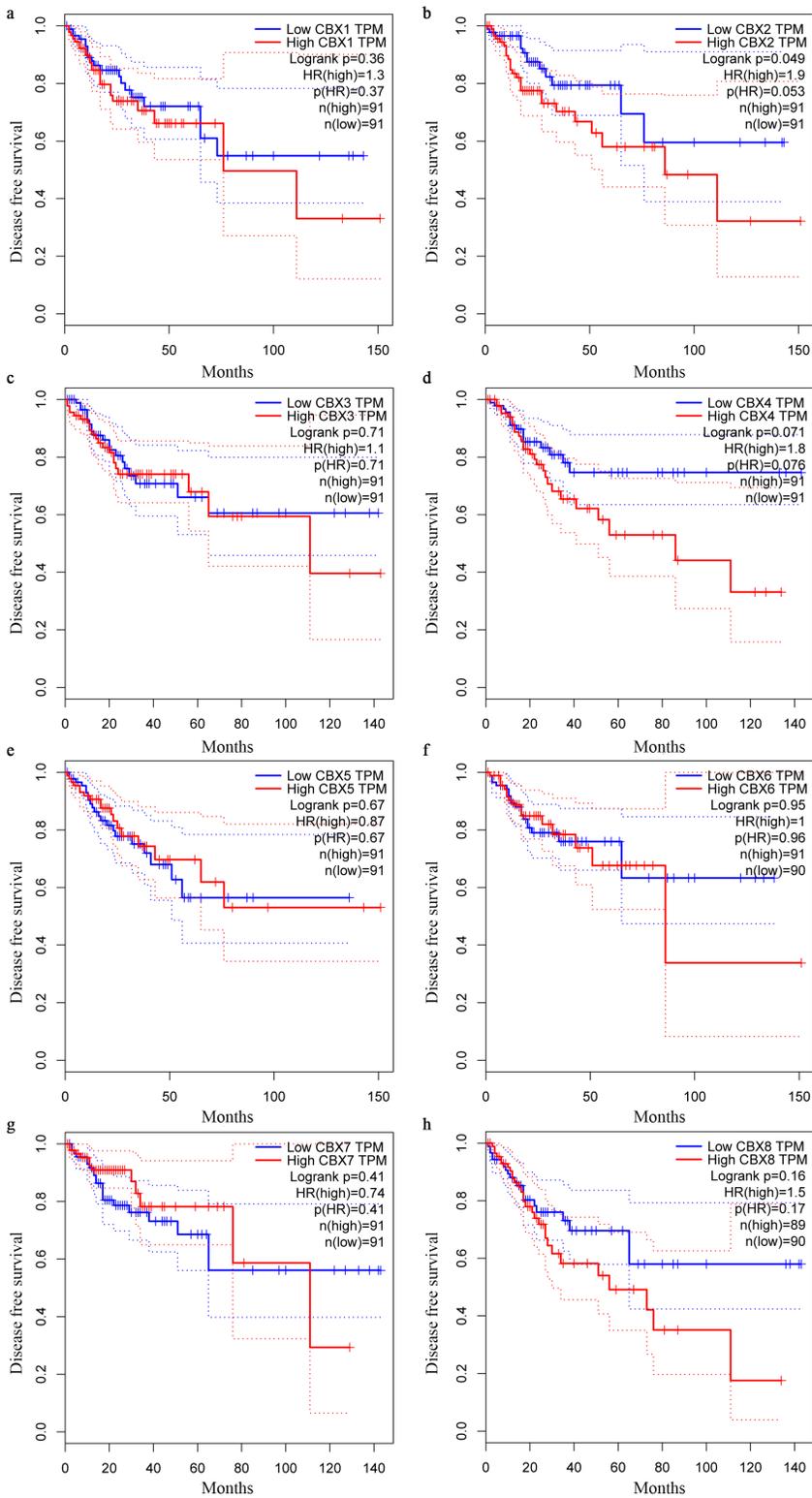
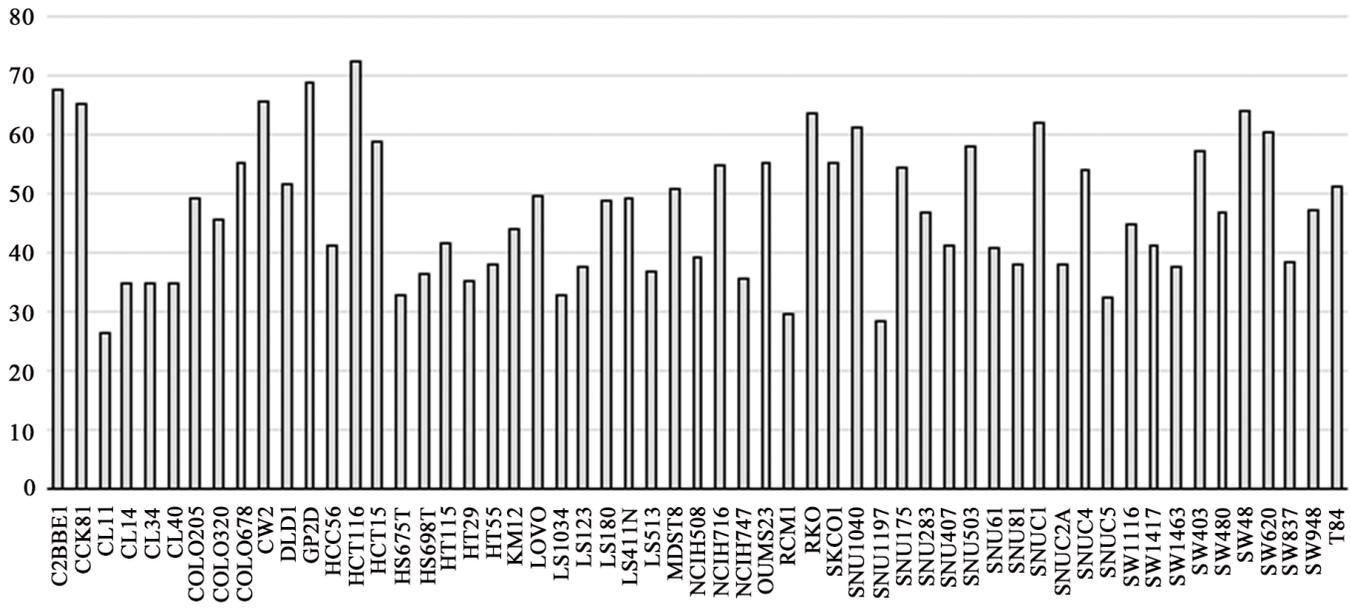
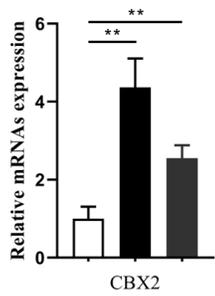
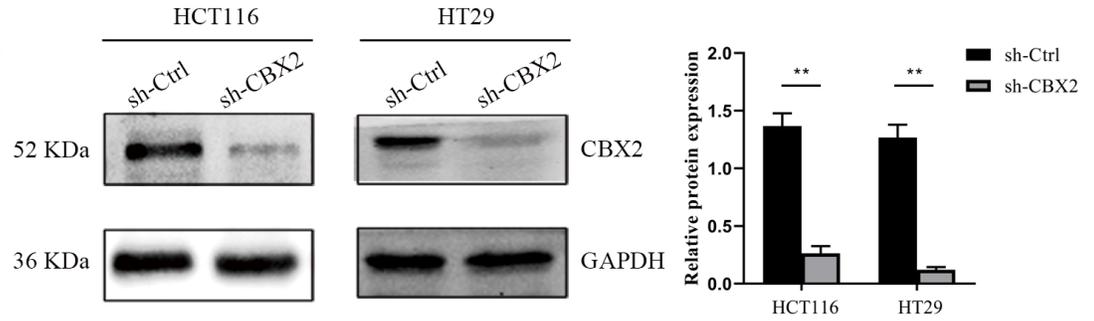


Figure 6

Survival analysis and prognostic values of CBX family in patients with CRC (GEPIA database). a-h DFS curves of CBX1, CBX2, CBX3, CBX4, CBX5, CBX6, CBX7, and CBX8 in patients with CRC. DFS, disease-free survival; HR, hazard ratio; CRC, colorectal cancer.

a

CBX2

**b****c****Figure 7**

CBX2 is upregulated in CRC cell lines. a mRNA expression level of CBX2 in different CRC cell lines, as determined by CCLE analysis. b The mRNA expression level of CBX2 in HCT116, HT29, and normal colon mucosal epithelial cell lines was detected by RT-qPCR (n=3 independent experiments. **P < 0.01). c The protein expression level of CBX2 in stably-transduced shCtrl and shCBX2 HCT116 and HT29 cell lines was analyzed by western blotting (n=3 independent experiments. **P < 0.01). CRC, colorectal cancer.

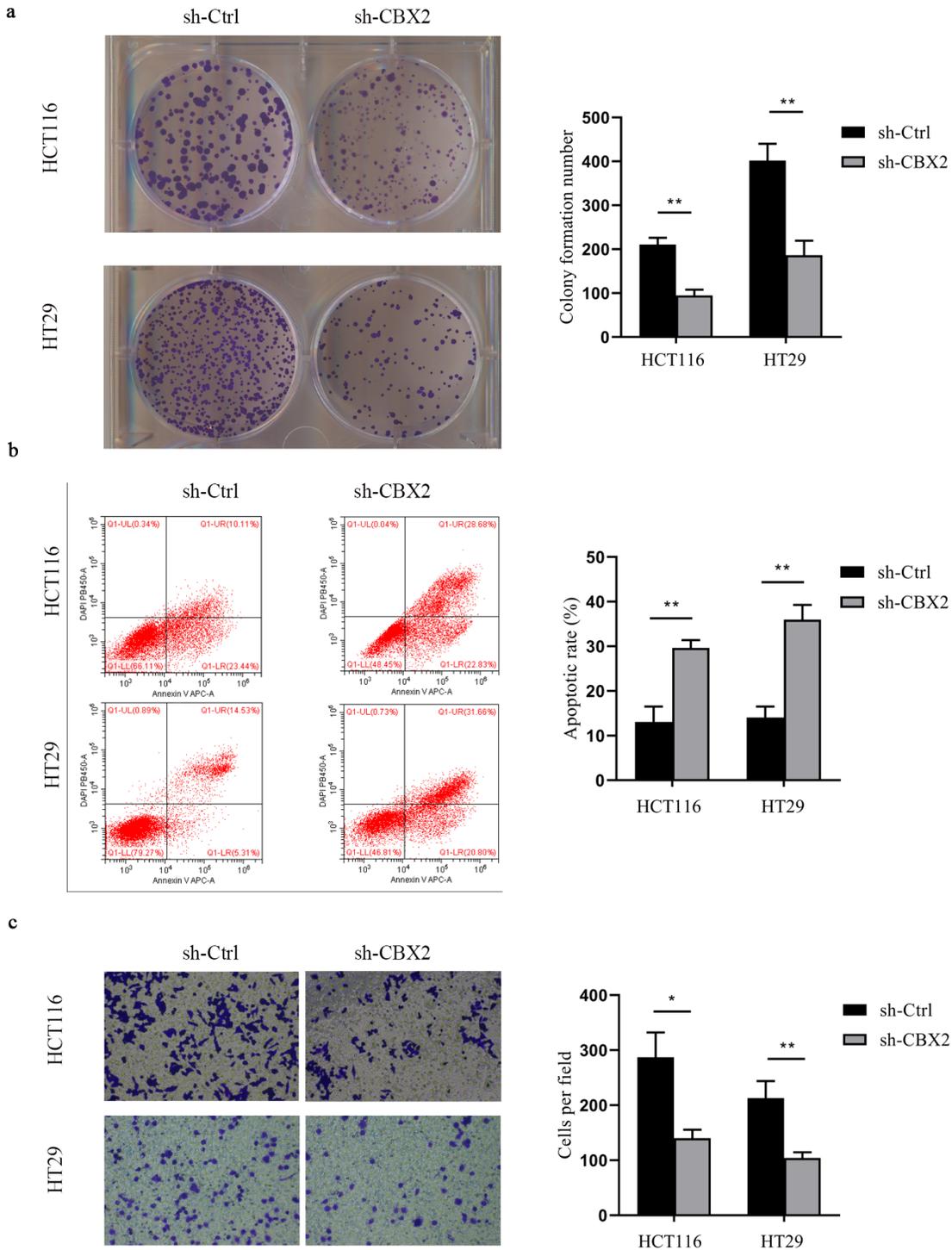


Figure 8

CBX2-knockdown inhibits CRC cell proliferation and invasion. a CBX2-knockdown suppressed cell proliferation, as determined by colony formation assays in HCT116 and HT29 cell lines (n=3 independent experiments $**P < 0.01$). b Apoptosis ratios of CBX2 in stably-transduced shCtrl and shCBX2 HCT116 and HT29 cell lines were detected by flow cytometric analysis (n=3 independent experiments $*P < 0.05$, $**P < 0.01$). c Knockdown of CBX2 significantly inhibited cell invasion in HCT116 and HT29 cell lines (n=3 independent experiments $*P < 0.05$, $**P < 0.01$).