



OPEN

## Bioinformatic analysis of the expression and prognostic value of chromobox family proteins in human breast cancer

Xiaomin Li<sup>1,3</sup>, Junhe Gou<sup>2</sup>, Hongjiang Li<sup>3</sup> & Xiaoqin Yang<sup>3</sup>✉

Chromobox (CBX) family proteins control chromatin structure and gene expression. However, the functions of CBXs in cancer progression, especially breast cancer, are inadequately studied. We assessed the significance of eight CBX proteins in breast cancer. We performed immunohistochemistry and bioinformatic analysis of data from Oncomine, GEPIA Dataset, bcGenExMiner, Kaplan–Meier Plotter, and cBioPortal. We compared mRNA and protein expression levels of eight CBX proteins between breast tumor and normal tissue. The expression difference of CBX7 was the greatest, and CBX7 was downregulated in breast cancer tissues compared with normal breast tissues. The expression of CBX2 was strongly associated with tumor stage. We further analyzed the association between the eight CBX proteins and the following clinicopathological features: menopause age, estrogen receptor (ER), progesterone receptor (PR) and HER-2 receptor status, nodal status, P53 status, triple-negative status, and the Scarff–Bloom–Richardson grade (SBR) and Nottingham prognostic index (NPI). Survival analysis in the Kaplan–Meier Plotter database showed that the eight CBX proteins were significantly associated with prognosis. Moreover, CBX genes in breast cancer patients had a high net alteration frequency of 57%. There were significant co-expression correlations between the following CBX protein pairs: CBX4 positively with CBX8, CBX6 positively with CBX7, and CBX2 negatively with CBX7. We also analyzed the Gene Ontology enrichment of the CBX proteins, including biological processes, cellular components, and molecular functions. CBX 1/2/3/5/8 may be oncogenes for breast cancer, whereas CBX 6 and 7 may be tumor suppressors for breast cancer. All eight CBX proteins may be predictive for prognosis. Clinical trials are needed to confirm the significance of the eight CBX proteins in breast cancer.

### Abbreviations

CBXs	Chromobox family proteins
ER	Estrogen receptor
PR	Progesterone receptor
TNBC	Triple-negative breast cancer
SBR	The Scarff–Bloom–Richardson grade
NPI	Nottingham prognostic index
HP1	Heterochromatin protein 1
OS	Overall survival
RFS	Relapse-free survival
PPS	Post-progression survival
DMFS	Distance metastasis-free survival

Chromobox (CBX) family proteins are the mammalian orthologs to the heterochromatin protein 1 (HP1) and Polycomb proteins that regulate heterochromatin, gene expression, and developmental programs<sup>1</sup>. The HP1 orthologs CBX1, CBX3, and CBX5 share a characteristic N-terminal chromodomain, a central hinge domain,

<sup>1</sup>West China School of Medicine, West China Hospital, Sichuan University, Chengdu, People's Republic of China. <sup>2</sup>Department of Pathology, West China Hospital, Sichuan University, Chengdu, People's Republic of China. <sup>3</sup>Department of Breast Surgery, West China Hospital, Sichuan University, No. 37 Guoxue Alley, Wuhou District, Chengdu City, Sichuan Province, People's Republic of China. ✉email: XiaoqinYang2019@outlook.com

and a C-terminal chromoshadow domain<sup>2</sup>. The Polycomb orthologs CBX2, CBX4, CBX6, CBX7, and CBX8 have a C-terminal polycomb repressor box that serves as a canonical component in Polycomb Repressive Complex 1<sup>3</sup>. Dysregulation of CBX family proteins is associated with tumorigenesis of many cancers, such as breast cancer<sup>4</sup>, pancreatic cancer<sup>5</sup>, thyroid cancer<sup>6</sup>, colorectal cancer<sup>7</sup>, lung cancer<sup>8</sup>, and ovarian cancer<sup>9</sup>.

Breast cancer is the most frequent malignancy and the leading cause of cancer death among women worldwide<sup>10</sup>. Although many experimental and clinical investigations have been conducted for novel and less toxic treatments, and the molecular basis of the pathogenesis of breast cancer has been studied extensively, patient survival rates still need improvement<sup>11</sup>. Biomarkers, such as ER, PR, and HER-2, have been used widely for breast cancer prognosis and as targets of endocrine therapy or targeted therapy<sup>12,13</sup>.

Due to tumor heterogeneity, there is a high demand for new biomarkers to improve individualized patient treatment and prediction of outcomes. Investigators reported that eight CBX family proteins (CBX 1–8) have important functions in breast cancer<sup>4,14–20</sup>. CBX2 promoted breast cancer cell proliferation; its overexpression caused upregulation of genes involved in cell cycle progression, and CBX2 overexpression was associated with poor 5-year survival<sup>14</sup>. Upregulation of CBX4 exerted an oncogenic effect on breast cancer by the Notch1 signaling pathway<sup>16</sup>. However, the activities of CBXs in the development of breast cancer, as tumor promoters or suppressors, require additional research.

In this study, our goal was to predict CBX family members' functional significance in breast cancer by bioinformatics analyses of databases. We examined diverse expression patterns, clinicopathological parameters, prognostic values, including overall survival (OS), disease-free survival (DFS), post-progression survival (PPS), and distance metastasis-free survival (DMFS), genetic alterations, and gene ontology. Our findings indicated that CBXs might have complex and distinct functions in breast cancer progression.

## Materials and methods

**ONCOMINE database.** ONCOMINE ([www.oncomine.org](http://www.oncomine.org))<sup>21</sup> is a cancer microarray database and data-mining platform facilitating discovery from genome-wide expression analysis. Using this database, we analyzed the mRNA expression of eight CBX family proteins in breast cancer compared with normal breast tissues. We chose the Breast Cancer vs. Normal Analysis about each individual CBX protein, and the threshold included expression fold change  $\geq 1.5$  between cancer and normal tissues,  $p$  value  $< 0.05$ , and gene rank  $\geq$  top 10%.

**GEPIA dataset.** GEPIA (Gene Expression Profiling Interactive Analysis; <https://gepia.cancer-pku.cn>)<sup>22</sup>, a tool based on TCGA and GTEx data, provides RNA expression data of 9,736 tumors and 8,587 normal samples. Using this database, we performed differential expression analysis and tumor stage analysis related to each CBX protein for patients with breast cancer. In the expression analysis, the threshold included expression fold change  $\geq 1.5$  between cancer and normal tissues,  $p$  value  $< 0.05$ .

**Immunohistochemistry.** We performed immunohistochemistry by using CBX2 (Abnova, monoclonal, mouse, ABN-MAB17287, 1/800, pH 6.0) and CBX7 (Invitrogen, polyclonal, rabbit, PA5-61801, 1/50, pH 7.2) antibodies in 40 pairs of paraffin-embedded invasive breast cancer tissues (IBCs) and tumor-adjacent normal tissues. These sample tissues were derived from 40 patients diagnosed with primary breast cancer in West China Hospital, Sichuan University, from 2018 to 2019. The Ethics Committee of West China Hospital, Sichuan University, approved this study, and all participants signed the written informed consent.

Sections of 3 mm were cut with a microtome from the paraffin-embedded tissue blocks of IBCs and normal tissues. Then, the sections were incubated with anti-CBX2 and anti-CBX7 antibody at 4°C overnight, covered with 3, 3'-diaminobenzidine, and mounted on slides with Vectashield (Vector Laboratories). Slides were observed by light microscopy. Control experiments without primary antibody demonstrated signal specificity.

All methods were carried out in accordance with relevant guidelines and regulations. The immunohistochemistry experiment was approved by the National Key Laboratory of Biotherapy of West China Hospital, Sichuan University.

**Breast cancer gene-expression miner v4.4 (bc-GenExMiner v4.4).** The Breast Cancer Gene-Expression Miner v4.4 (<https://bcgenex.centregauducheau.fr/BC-GEM/GEM-Accueil.php?js=1>)<sup>23,24</sup>, a DNA microarray and RNA-seq database, can be used to analyze prognosis based on gene expression. Using the RNA-seq data, we evaluated the association between mRNA expression of the eight CBX family proteins and clinicopathological parameters, such as menopause age, ER, PR, HER-2, nodal status, P53 status, basal-like and TNBC status, and the Nottingham prognostic index (NPI) and Scarff–Bloom–Richardson (SBR) grading. In addition, we performed the pairwise correlation analysis of the eight CBX proteins, and we analyzed their Gene Ontology enrichment, including biological processes, cellular components, and molecular functions. Data were last updated on December 9, 2019.

**The Kaplan–Meier Plotter.** The Kaplan–Meier Plotter ([www.kmplot.com](http://www.kmplot.com))<sup>25</sup> is a tool to draw survival plots with gene expression data and survival information from GEO, EGA and TCGA cancer microarray datasets. We evaluated the relevance of the mRNA expression level of eight CBX proteins to the clinical outcomes (OS, RFS, PPS and DMFS) of untreated breast cancer patients. This tool automatically calculates the best cutoff value, log-rank  $P$  value, hazard ratio (HR), and 95% confidence intervals (CIs).

**cBioPortal.** The cBio Cancer Genomics Portal (<https://cbioportal.org>)<sup>26,27</sup> is a resource for interactive exploration of multidimensional cancer genomics datasets. We analyzed the gene alteration frequency and co-expres-

Analysis Type by Cancer	CBX1		CBX2		CBX3		CBX4		CBX5		CBX6		CBX7		CBX8	
	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal	Cancer vs. Normal
Bladder Cancer	2		2		2		1				2		4			
Brain and CNS Cancer	3		3	1	14			2	5	1		13	1	11	1	
Breast Cancer	1		7	1	22		10		2		1	2	1	20	6	
Cervical Cancer	1			1	5				4					1		
Colorectal Cancer	6		10		24		18		10			4		12	6	
Esophageal Cancer	2	1			4						1			1		
Gastric Cancer	6		5		4		6				1			1		
Head and Neck Cancer	6	1	2		16		2		3		1			3		
Kidney Cancer	1	2	1		7		2	2	2	1		2		1	1	
Leukemia	1	3	5		1	1	3		5	4	1	3	1	8		
Liver Cancer	4				2			1	1		1			1		
Lung Cancer	13		3		12		2		8			1		7		
Lymphoma	1		1		5	3		5	6	2	8	1		1		
Melanoma					3			1	1					1		
Myeloma					1											1
Other Cancer	3	1	3		10		2	1	6	2		2		5	1	
Ovarian Cancer		1			2							1		5		
Pancreatic Cancer	2			1	2		1		3						1	
Prostate Cancer					4		5		1	3		2		4		
Sarcoma	11				11		2		10	1	2			10	2	
Significant Unique Analyses	62	9	42	4	150	3	53	12	67	13	16	32	3	95	18	1
Total Unique Analyses	442		367		455		421		451		392		351		337	



**Figure 1.** The mRNA expression of eight CBX proteins in various cancer types in OncoPrint. Red: overexpression or copy gain; Blue: underexpression or copy loss. Color intensity indicates the best rank of the gene in the analyses. The number in each cell is the number of analyses that met our thresholds.

sion of eight CBX family proteins using METABRIC data from 1904 breast cancer patients<sup>28</sup>. The mRNA expression z-score threshold was  $\pm 1.5$  between the unaltered and altered patients.

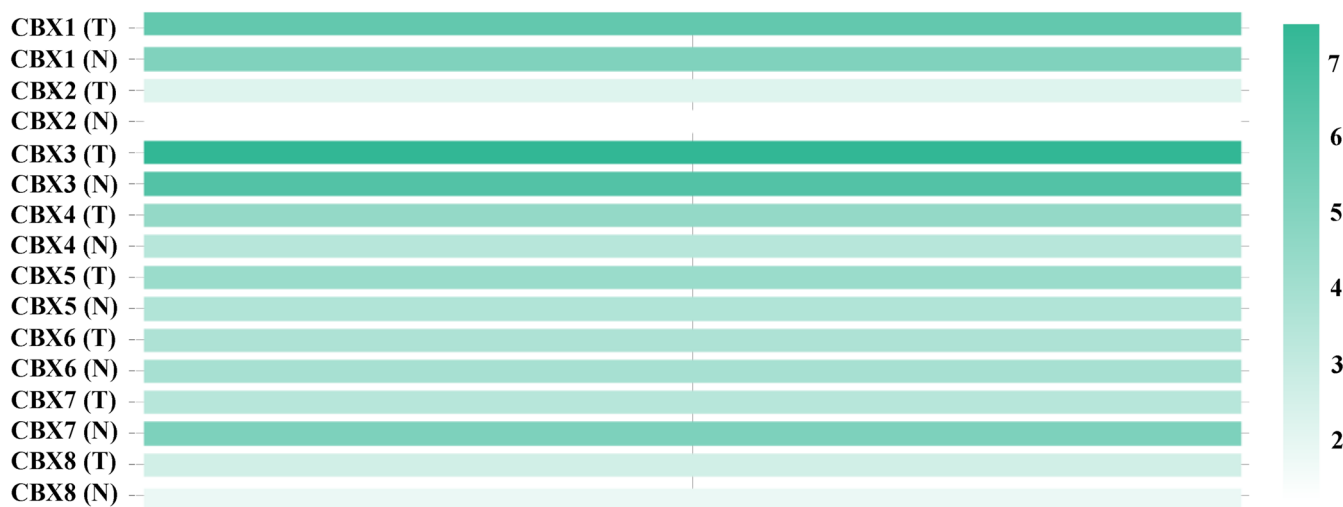
### Results

**The mRNA and protein expression of CBXs in breast cancer.** We used the OncoPrint and GEPIA databases to retrieve mRNA expression levels of the eight CBX proteins in breast cancer. OncoPrint analysis revealed the mRNA expression of the eight CBX proteins in 19 common types of cancer and their comparisons with normal tissues (Fig. 1). The following expression patterns were observed for breast cancer: Overexpressed, CBX1, one of 49 (1/49) analyses, CBX 2–8, 7/43, 22/53, 10/53, 2/53, 1/52, 1/41, and 6/42, respectively. Down-regulated, CBX2, CBX6, and CBX7, 1/43, 2/52, and 20/42, respectively.

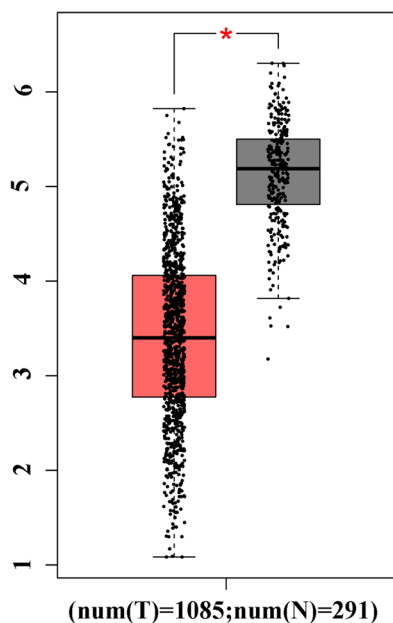
Figure 2A and Supplementary Figure 1A–B show the mRNA expression of eight CBX proteins in GEPIA. CBX7 was downregulated in tumor samples compared with the normal counterpart ( $P < 0.05$ ; Fig. 2B,C).

We performed immunohistochemistry to measure CBX2 and CBX7 protein expression (Fig. 3). We found that CBX2 protein was highly expressed in the breast cancer tissues compared with normal tissues, and expression of CBX7 protein in breast cancer tissues was lower than tumor-adjacent normal tissues.

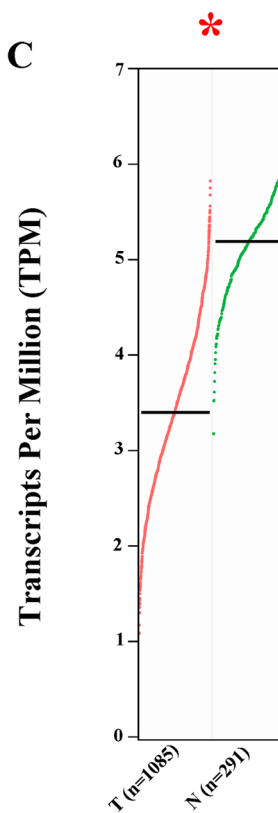
**A**



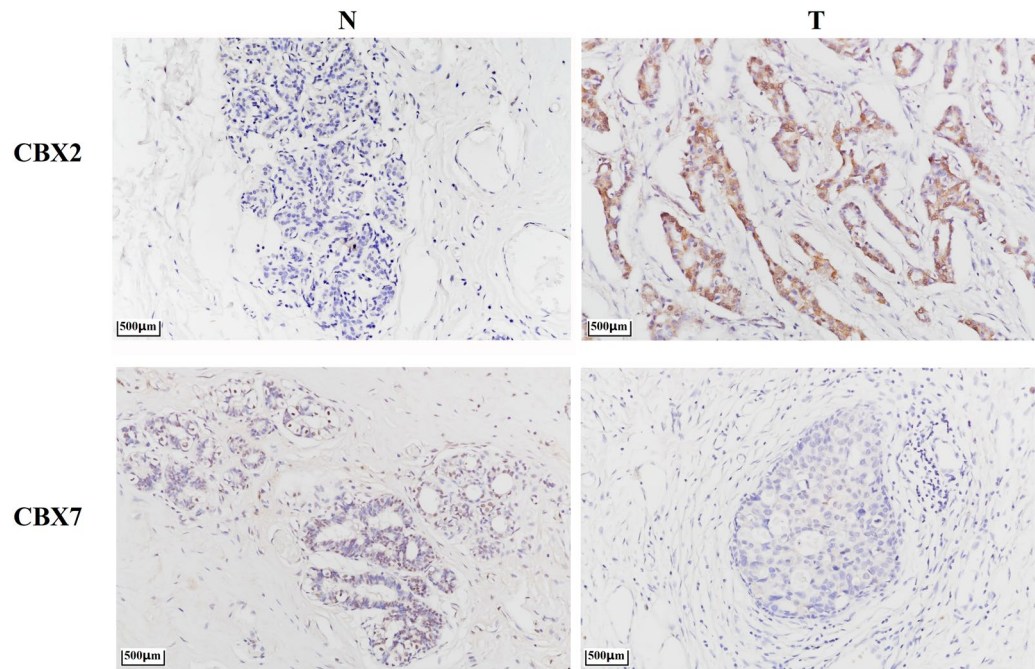
**B**



**C**



**Figure 2.** The mRNA expression of CBX proteins in breast tumors and normal tissues in GEPIA. **(A)** Eight CBX proteins. Color intensity indicates the mRNA expression of the gene in the tissue. **(B)** CBX7 mRNA expression on the box plot. **(C)** CBX7 mRNA expression profile; red: tumor tissue, Green: normal tissue; \* $P < 0.05$  and  $|\text{Log}_2(\text{fold-change})|$  cutoff = 1.5. We used a log scale to show mRNA expression level.



**Figure 3.** CBX2 and CBX7 protein expression in breast cancer and tumor-adjacent normal tissues. N: tumor-adjacent normal tissues; T: tumor tissues.

**Associations between CBXs and the clinicopathological parameters of patients with breast cancer.** Table 1 shows the clinicopathological parameters and associations derived from the analysis of 4712 breast cancer patients in the TCGA and GSE81540 RNA-Seq datasets in bc-GenExMiner v4.4. High CBX2 correlated with young menopause age ( $P < 0.0001$ ), whereas high CBX4 and CBX7 were associated with old advanced menopause age ( $P < 0.0001$ ). CBX2 was negatively associated with ER and PR expression and positively with HER-2 expression. However, CBX 4–7 were positively associated with ER and PR expression and negatively with HER-2 expression. Patients with high CBX 1–4 were more likely to be in positive nodal status, but patients with high CBX 6–7 tended to be in negative status. Except for CBX5 and CBX8, the other six CBX proteins were correlated with P53 status. Patients with high CBX 1–3 and low CBX 4–7 were more likely to be TNBC phenotype. Concerning the prognostic factors, Scarff–Bloom–Richardson grade (SBR) and Nottingham prognostic index (NPI) in breast cancer, high CBX 1–4 and CBX8 had high SBR and NPI; by contrast, high CBX 5–7 had low SBR and NPI. CBX2 was positively associated with clinical stages of patients (Fig. 4). Stage IV patients had higher CBX2 expression compared with other stages. Supplementary Figure 2 shows the Association associations between other CBX proteins and clinical stage.

**The prognostic value of CBXs.** Prognosis analysis by the Kaplan–Meier Plotter revealed that all eight CBX proteins had predictive value for the relapse-free survival of breast cancer patients (Supplementary Figure 3). Decreased CBX2 and increased CBX 4/6/7 mRNA levels were remarkably associated with longer overall survival (Fig. 5). Patients with high CBX 1/7/8 mRNA levels had longer post-progression survival than the low counterparts (Supplementary Figure 3). Moreover, decreased CBX 1/2/3/5 and increased CBX 6/7 mRNA levels were significantly correlated with longer distance metastasis-free survival (Supplementary Figure 3).

**Alterations and co-expression of CBXs.** Using cBioPortal, we analyzed genetic alterations of the eight CBX proteins and found a high alteration frequency (57%) in breast cancer patients (Fig. 6). Patients who had CBX4 alteration were the most cases in eight CBX family proteins, making up 15.45% of all cases involved, and their primary alteration type was mRNA high. Besides, there were co-expression correlations between the following CBX proteins: CBX4 positively with CBX8, CBX6 positively with CBX7, and CBX2 negatively with CBX7 (Fig. 7B). The bc-GenExMiner produced similar correlations (Fig. 7A,C–E).

**CBXs gene ontology enrichment.** Using bc-GenExMiner, we found the 50 (or fewer) genes most correlated to each CBX protein. Some genes were positively correlated with CBX protein, whereas some were negatively correlated with the protein. We performed gene ontology analysis of each CBX protein for biological processes (Supplementary Table 2 and Table 2; the most significant term), cellular components (Supplementary Table 3 and Table 3), and molecular functions (Supplementary Table 4 and Table 4).

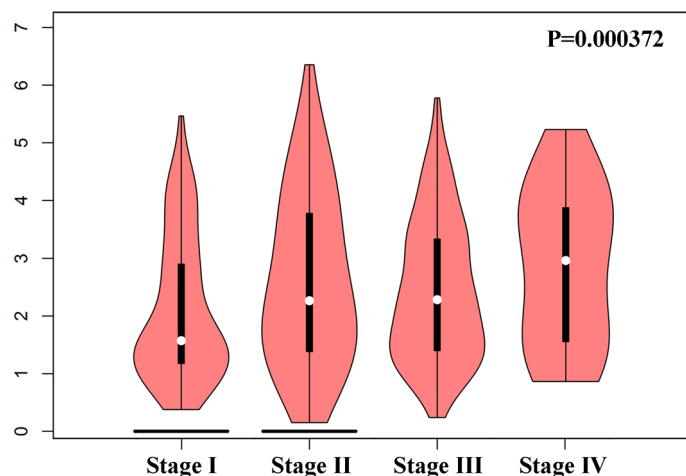
	CBX1	CBX2
<b>(A)</b>		
Age	≤ 51 = > 51 ( <i>P</i> = 0.1740)	≤ 51 > > 51 ( <i>P</i> < 0.0001)
ER	+ < - ( <i>P</i> = 0.0168)	+ < - ( <i>P</i> < 0.0001)
PR	+ = - ( <i>P</i> = 0.6980)	+ < - ( <i>P</i> < 0.0001)
HER-2	+ > - ( <i>P</i> = 0.0175)	+ > - ( <i>P</i> < 0.0001)
Nodal status	+ > - ( <i>P</i> = 0.0167)	+ > - ( <i>P</i> = 0.0079)
P53 status	Wild type < mutated ( <i>P</i> = 0.0007)	Wild type < mutated ( <i>P</i> < 0.0001)
TNBC status	TNBC > not TNBC ( <i>P</i> = 0.0341)	TNBC > not TNBC ( <i>P</i> < 0.0001)
SBR	SBR1 < SBR2 < SBR3 ( <i>P</i> < 0.0001)	SBR1 < SBR2 < SBR3 ( <i>P</i> < 0.0001)
NPI	NPI1 < NPI2 = NPI3 ( <i>P</i> < 0.0001)	NPI1 < NPI2 < NPI3 ( <i>P</i> < 0.0001)
	CBX3	CBX4
Age	≤ 51 = > 51 ( <i>P</i> = 0.0843)	≤ 51 < > 51 ( <i>P</i> < 0.0001)
ER	+ < - ( <i>P</i> = 0.0040)	+ > - ( <i>P</i> < 0.0001)
PR	+ = - ( <i>P</i> = 0.4357)	+ > - ( <i>P</i> < 0.0001)
HER-2	+ = - ( <i>P</i> = 0.1457)	+ < - ( <i>P</i> = 0.0005)
Nodal status	+ > - ( <i>P</i> = 0.0003)	+ > - ( <i>P</i> = 0.0130)
P53 status	Wild type < mutated ( <i>P</i> < 0.0001)	Wild type > mutated ( <i>P</i> < 0.0001)
TNBC status	TNBC > not TNBC ( <i>P</i> = 0.0115)	TNBC < not TNBC ( <i>P</i> < 0.0001)
SBR	SBR1 < SBR2 < SBR3 ( <i>P</i> < 0.0001)	SBR1 < SBR2 = SBR3 ( <i>P</i> = 0.0083)
NPI	NPI1 < NPI2 < NPI3 ( <i>P</i> < 0.0001)	NPI1 < NPI2 = NPI3 ( <i>P</i> = 0.0005)
	CBX5	CBX6
<b>(B)</b>		
Age	≤ 51 = > 51 ( <i>P</i> = 0.4356)	≤ 51 = > 51 ( <i>P</i> = 0.9725)
ER	+ > - ( <i>P</i> < 0.0001)	+ > - ( <i>P</i> < 0.0001)
PR	+ > - ( <i>P</i> < 0.0001)	+ > - ( <i>P</i> = 0.0002)
HER-2	+ < - ( <i>P</i> < 0.0001)	+ < - ( <i>P</i> < 0.0001)
Nodal status	+ = - ( <i>P</i> = 0.3564)	+ < - ( <i>P</i> < 0.0001)
P53 status	Wild type = mutated ( <i>P</i> = 0.8605)	Wild type > mutated ( <i>P</i> < 0.0001)
TNBC status	TNBC < not TNBC ( <i>P</i> = 0.0005)	TNBC < not TNBC ( <i>P</i> = 0.0009)
SBR	SBR1 > SBR2 > SBR3 ( <i>P</i> < 0.0001)	SBR1 > SBR2 > SBR3 ( <i>P</i> < 0.0001)
NPI	NPI1 > NPI2 > NPI3 ( <i>P</i> < 0.0001)	NPI1 > NPI2 > NPI3 ( <i>P</i> < 0.0001)
	CBX7	CBX8
Age	≤ 51 < > 51 ( <i>P</i> < 0.0001)	≤ 51 = > 51 ( <i>P</i> = 0.4638)
ER	+ > - ( <i>P</i> < 0.0001)	+ > - ( <i>P</i> = 0.0083)
PR	+ > - ( <i>P</i> < 0.0001)	+ = - ( <i>P</i> = 0.8690)
HER-2	+ < - ( <i>P</i> < 0.0001)	+ > - ( <i>P</i> = 0.0144)
Nodal status	+ < - ( <i>P</i> = 0.0174)	+ = - ( <i>P</i> = 0.0932)
P53 status	Wild type > mutated ( <i>P</i> < 0.0001)	Wild type = mutated ( <i>P</i> = 0.7531)
TNBC status	TNBC < not TNBC ( <i>P</i> < 0.0001)	TNBC = not TNBC ( <i>P</i> = 0.6027)
SBR	SBR1 > SBR2 > SBR3 ( <i>P</i> < 0.0001)	SBR1 < SBR2 < SBR3 ( <i>P</i> < 0.0001)
NPI	NPI1 > NPI2 > NPI3 ( <i>P</i> < 0.0001)	NPI1 < NPI2 < NPI3 ( <i>P</i> < 0.0001)

**Table 1.** Correlation between clinicopathological parameters and (A) CBX 1–4, (B) CBX 5–8. *CBXs* chromobox family proteins, *ER* estrogen receptor, *PR* progesterone receptor, *HER-2* human epidermal growth factor receptor 2, *P53* a tumor suppressor gene, *TNBC* triple-negative breast cancer, *SBR* the Scarff–Bloom–Richardson (SBR) grade, *NPI* the Nottingham prognostic index.

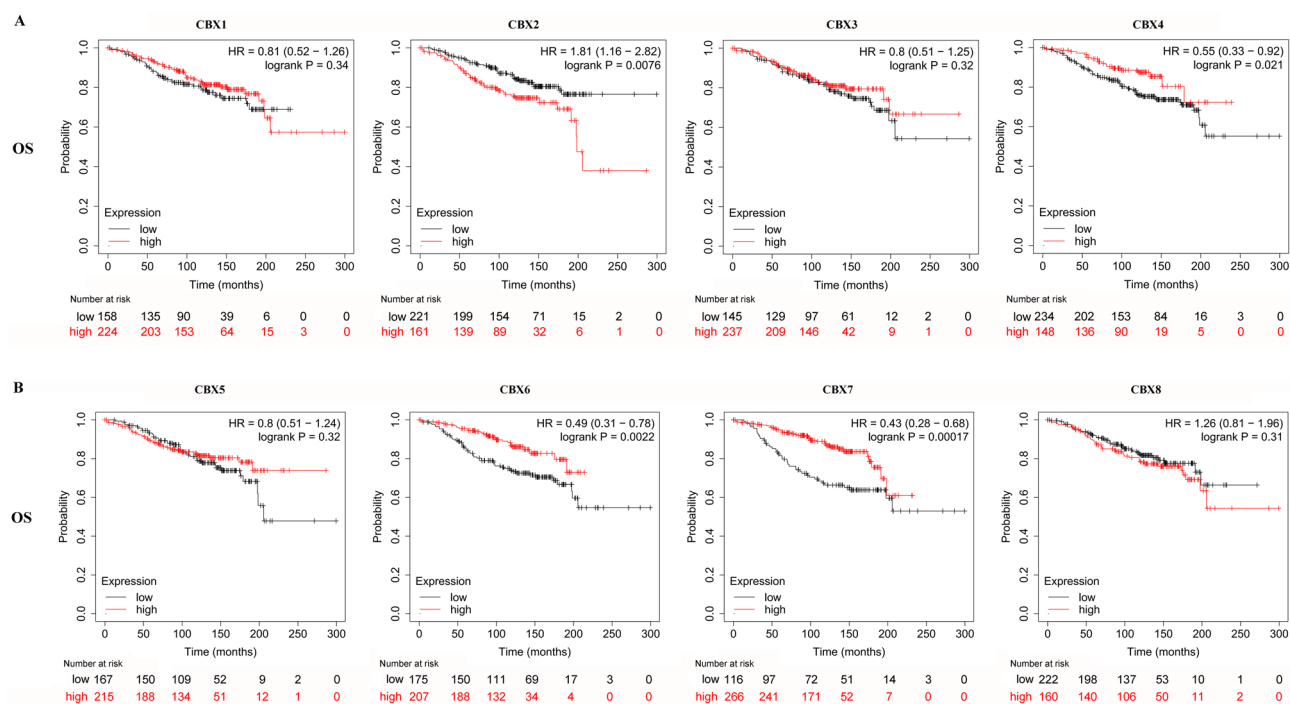
## Discussion

Dysregulation of CBX family proteins affects the development of multiple cancers, including breast cancer. For tumorigenesis and prognosis of breast cancer, despite the identification of the significant functions of some CBX family proteins, the complex and distinct activities of CBXs still require investigation. In this study, we used novel applications of bioinformatics to analyze four aspects of eight CBX proteins in breast cancer: expression pattern, clinicopathological parameters, prognostic value, and genetic alteration.

Human HP1 proteins, HP1 $\alpha$ /CBX5, HP1 $\beta$ /CBX1, and HP1 $\gamma$ /CBX3, correlated with proliferation, invasion, and metastasis by regulating gene expression in human breast cancer cells<sup>2,17,29</sup>. CBX5 is the most studied HP1 protein, and CBX3 is barely examined. Because of tumor heterogeneity, the expression of HP1 proteins differed in different breast cancer biospecimens. All three HP1 subtypes were positively correlated with the expression level of Ki-67<sup>17</sup>. We found that high CBX1 and CBX3 were associated with poor survival of breast cancer patients.



**Figure 4.** Association between CBX2 and clinical stages of breast cancer patients. The y axis:  $\log_2(\text{TPM} + 1)$  (TPM: transcript per million).

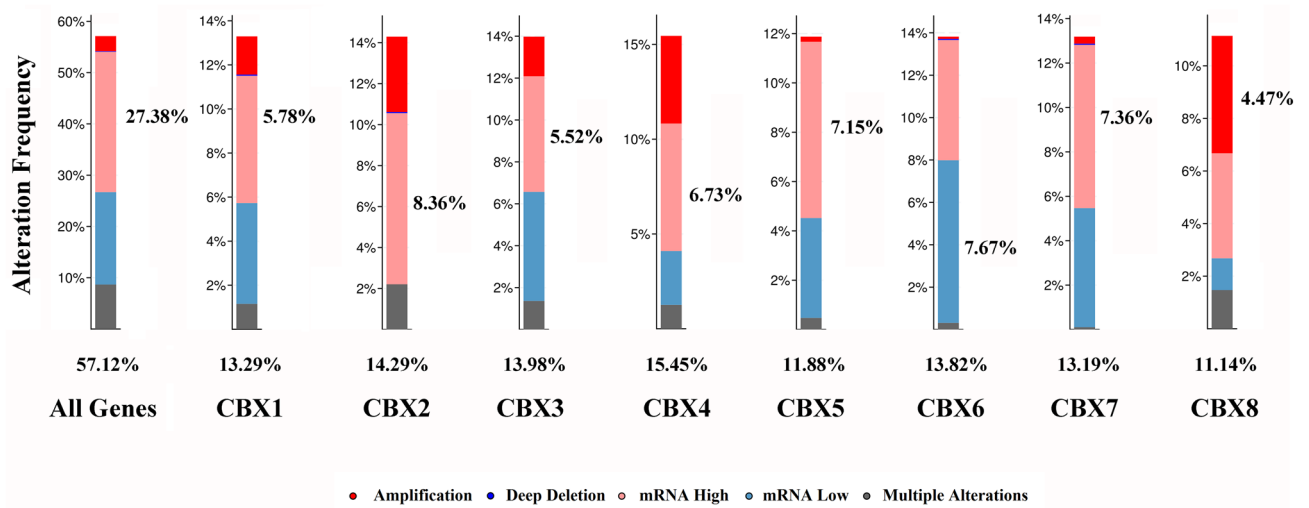


**Figure 5.** Prognostic values of CBX proteins for overall survival (OS). (A) CBX 1–4; (B) CBX 5–8.

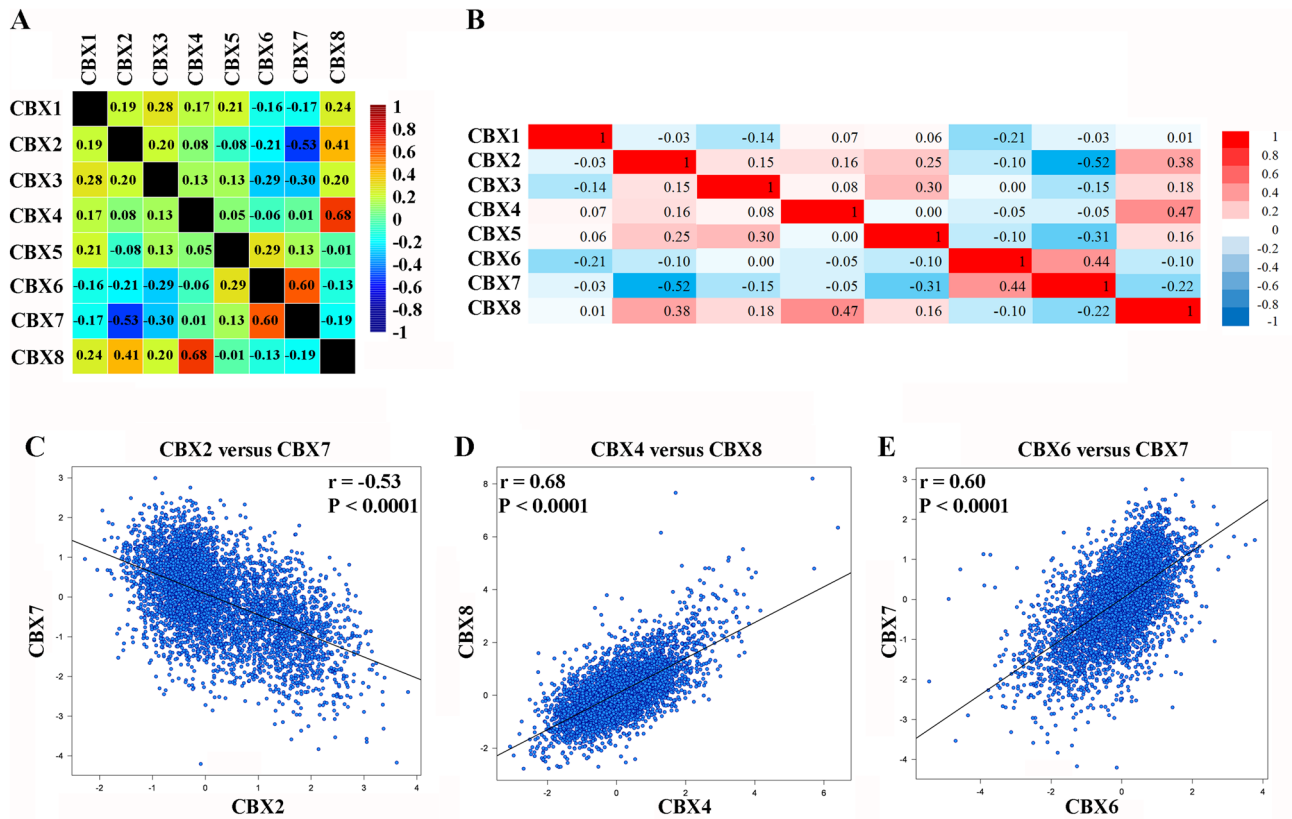
High CBX1 and CBX3 expression was associated with aggressive types of breast cancers (TNBC phenotype), and the patients were more likely to have had lymph node metastasis and P53 mutations. Therefore, CBX1 and CBX3 may function as oncogenes.

CBX5 was upregulated at the mRNA and protein levels in breast cancer cells compared with non-cancerous cells<sup>29,30</sup>. However, CBX5 was downregulated in highly invasive or metastatic breast cancer cell lines compared with weakly invasive or non-metastatic cells, which suggested that CBX5 is a metastatic suppressor in the invasion process<sup>29,31,32</sup>. The suppressor mechanism of CBX5 in invasion is unknown. We also found patients with high CBX5 tended to have less aggressive tumor subtypes (not TNBC phenotype). Prognosis analysis showed that high CBX5 was associated with shorter RFS and DMFS, which suggested that CBX5 functions as an oncogene.

CBX2, CBX4, CBX6, CBX7, and CBX8 are subunits of distinct polycomb repressive 1 complexes that have important functions in the development and progression of breast cancer. CBX2 was overexpressed in breast cancer, and high CBX2 expression was associated with lymph node metastasis, poor tumor differentiation, and high TNM stage<sup>15</sup>. Our results are consistent with those of Zheng et al. who found that CBX2 expression could affect OS and RFS of breast cancer patients independently<sup>15</sup>. Further, we found that patients with high CBX2 tended to have more aggressive tumor subtypes and P53 mutations. Moreover, CBX2 mRNA expression was negatively



**Figure 6.** The alteration frequency and mechanisms for CBX proteins.



**Figure 7.** Pearson's correlations for mRNA expression of pairwise combinations of CBX proteins in cBioPortal (B) and bc-GenExMiner (A, C–E). Tables (A), (B) include Pearson correlation coefficients, and p values of the coefficients are shown in Supplementary Table s1. The color scale interprets the correlation coefficient value. (C)–(E) show the correlation between CBX2 and CBX7, CBX4 and CBX8, and CBX6 and CBX7, respectively. *r* Pearson's correlation coefficient value; *P*; *P* value; *N* the number of patients.

correlated with CBX7. These results suggested that CBX2 may exert an oncogenic function in breast cancer. Zheng et al. revealed that CBX2 promotes breast tumorigenicity through the PI3K/AKT signaling pathway<sup>15</sup>. The oncogenic mechanism of CBX2 needs further explanation. CBX2 may be an oncogene and a potential therapeutic target for breast cancer. There are no available inhibitors of CBX2. CBX2 contains a chromodomain that binds H3K27me3 with high affinity; this property could be targeted pharmacologically. From the development perspective, a CBX2 antagonist would be a promising therapeutic agent for breast cancer. CBX2 was expressed



Genes	Significant terms	Description	p value	Associated genes
CBX1	GO:1900182	Positive regulation of protein localization to nucleus	4.45E-04	CDK5RAP3, KAT7
CBX2	GO:0045137	Development of primary sexual characteristics	1.14E-03	CBX2
CBX3	GO:1905007	Positive regulation of epithelial to mesenchymal transition involved in endocardial cushion formation	3.03E-05	ENG, TGFBR2
CBX4	GO:0042552	Myelination	6.37E-04	FAM126A, QKI
CBX5	GO:0070317	Negative regulation of G0 to G1 transition	4.45E-03	CBX5
CBX6	GO:0006325	Chromatin organization	1.69E-04	CBX6, CBX7
CBX7	GO:0051301	Cell division	7.24E-30	AURKA, BIRC5, BUB1, BUB1B, CCNA2, CCNB1, CCNE2, CDC20, CDCA2, CDCA3, CDCA5, CDCA8, CDK1, CENPE, CENPF, CENPW, CKS2, ERCC6L, FAM83D, KIF11, KIF14, KIF2C, KNSTRN, MAD2L1, MASTL, NCAPG, NEK2, OIP5, PRC1, PTTG1, SGO2, SPC25, TPX2, UBE2C
CBX8	GO:0070125	Mitochondrial translational elongation	2.59E-05	MRPL12, MRPL38, MRPL58, MRPS7

**Table 2.** Biological process. *CBXs* chromobox family proteins, *GO* gene ontology.

Genes	Significant terms	Description	p value	Associated genes
CBX1	GO:0005654	Nucleoplasm	1.97E-05	CBX1, DBF4B, EME1, KAT7, KPNB1, MRPL10, PHB, PNPO, PSMC3IP, SNF8, SRSF1, UBE2Z, UTP18
CBX2	GO:0000775	Chromosome, centromeric region	2.04E-03	CDCA5, CENPW
CBX3	GO:0005925	Focal adhesion	5.69E-07	ENG, FLNC, FLRT2, GSN, LIMS2, LRP1, STARD8, TNS1
CBX4	GO:1990907	Beta-catenin-TCF complex	7.55E-03	TLE4
CBX5	GO:0031618	Nuclear pericentric heterochromatin	1.46E-03	CBX5
CBX6	GO:0031519	PcG protein complex	1.23E-05	CBX6, CBX7
CBX7	GO:0030496	Midbody	3.45E-17	ANLN, ASPM, AURKA, BIRC5, CDCA8, CDK1, CENPE, CENPF, CEP55, ECT2, KIF14, KIF20A, KIF23, KIF4A, NEK2, PLK1, PRC1, RACGAP1
CBX8	GO:0035102	PRC1 complex	3.74E-06	CBX2, CBX4, CBX8

**Table 3.** Cellular component. *CBXs* chromobox family proteins, *GO* gene ontology.

Genes	Significant terms	Description	p value	Associated genes
CBX1	GO:0001850	Complement component C3a binding	1.33E-03	PHB
CBX2	GO:0033678	5'-3' DNA/RNA helicase activity	1.04E-03	PIF1
CBX3	GO:0034713	Type I transforming growth factor beta receptor binding	1.92E-04	ENG, TGFBR2
CBX4	GO:0051525	NFAT protein binding	3.75E-03	PPARA
CBX5	GO:0008486	Diphosphoinositol-polyphosphate diphosphatase activity	8.67E-04	NUDT4B
CBX6	GO:0004046	Aminoacylase activity	1.39E-03	ABHD14A-ACY1
CBX7	GO:0008017	Microtubule binding	8.43E-09	BIRC5, CENPE, FAM83D, KIF11, KIF14, KIF20A, KIF23, KIF2C, KIF4A, NUSAP1, PLK1, PRC1, RACGAP1
CBX8	GO:0035064	Methylated histone binding	2.48E-04	CBX2, CBX4, CBX8

**Table 4.** Molecular function. *CBXs* chromobox family proteins, *GO* gene ontology.

at low levels in most healthy adult tissues, so CBX2 inhibitors may have few side effects. In addition, CBX2 was overexpressed in breast cancer with poor prognosis, and CBX2 downregulation could inhibit breast tumorigenesis in vivo and vitro. Stage IV patients had higher CBX2 expression compared with other stages, and patients with high CBX2 were more likely to be in positive nodal status and TNBC phenotype. These findings suggested that CBX2 is associated with tumor progression and metastasis.

The mRNA and protein levels of CBX4 were higher in breast cancer tissues than in paired non-cancerous tissues, and high CBX4 expression was independently associated with shorter overall survival<sup>16</sup>. In addition, breast cancer patients with high CBX4 were more likely to have lymph node metastasis and higher clinical stages<sup>16</sup>. CBX4 exerted its oncogenic function through the Notch1 signaling pathway and circular RNA hsa\_circ\_0008039/miR-515-5p/CBX4 axis<sup>16,33</sup>. However, there is a contradiction between previous studies<sup>16,33</sup> and our survival measurements. By bioinformatic analysis, patients with high CBX4 had longer OS and RFS, which suggested that CBX4 exerts an anti-cancer effect. Large multicenter prospective studies are required to confirm our results.

Both CBX6 and CBX7 were downregulated in human breast cancer<sup>19,34</sup>. They inhibited breast progression through their pathways. CBX6 controlled a series of genes such as Bone Marrow Stromal cell antigen 2 (BST2) to regulate breast cancer<sup>35-37</sup>. CBX7 repressed breast tumorigenicity by suppressing the Wnt/b-catenin pathway<sup>38</sup>.

We found that the most significant difference between breast cancer and normal tissues was the mRNA expression of CBX7. Patients with low CBX6 or CBX7 were more likely to have lymph node metastasis and P53 mutations. These patients tended to have a more aggressive subtype (TNBC phenotype) with poor survival. Besides, CBX6 was positively correlated with CBX7 in breast cancer. These results suggested that CBX6 and CBX7 function as tumor suppressors in breast cancer.

CBX8 functioned in canonical and non-canonical ways to promote breast tumorigenesis<sup>39</sup>. First, polycomb repressive complex 1, the canonical CBX8-containing complex, promoted gene silencing by monoubiquitylation of H2AK119<sup>40</sup>. Second, the non-canonical CBX8 complex, in which CBX8 interacts with Wdr5, promoted the activation of genes in the Notch signaling pathway, regulating normal mammary gland development<sup>39,41</sup>. In addition, CBX8 regulated the p53/p21 WAF1 pathway by binding with SIRT1 to suppress premature senescence and growth arrest of breast cancer cells. We found that breast cancer patients with high CBX8 had shorter relapse-free survival compared with low CBX8. These findings suggested that CBX8 is a tumor promoter.

To date, a few investigators have studied the functional significance of CBX proteins in breast cancer. Our investigation consisted of just primarily bioinformatic analyses with some experimental data (i.e., immunohistochemistry). Therefore, extensive prospective clinical studies and other experiments are needed to validate our results. In addition, we need more research that compares CBX proteins with other prognostic markers.

We analyzed the expression, prognostic value, clinicopathological parameters, and Gene Ontology enrichment of eight CBX proteins using several large online databases. We identified the functional significance of these proteins in breast cancer.

Received: 27 December 2019; Accepted: 7 October 2020

Published online: 20 October 2020

## References

1. Kaustov, L. *et al.* Recognition and specificity determinants of the human cbx chromodomains. *J. Biol. Chem.* **286**(1), 521–529 (2011).
2. Lee, Y. H., Kuo, C. Y., Stark, J. M., Shih, H. M. & Ann, D. K. HP1 promotes tumor suppressor BRCA1 functions during the DNA damage response. *Nucleic Acids Res.* **41**(11), 5784–5798 (2013).
3. Klauke, K. *et al.* Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat. Cell. Biol.* **15**(4), 353–362 (2013).
4. Lee, S. H., Um, S. J. & Kim, E. J. CBX8 suppresses Sirtinol-induced premature senescence in human breast cancer cells via cooperation with SIRT1. *Cancer Lett.* **335**(2), 397–403 (2013).
5. Karamitopoulou, E. *et al.* Loss of the CBX7 protein expression correlates with a more aggressive phenotype in pancreatic cancer. *Eur. J. Cancer.* **46**(8), 1438–1444 (2010).
6. Pallante, P. *et al.* Loss of the CBX7 gene expression correlates with a highly malignant phenotype in thyroid cancer. *Cancer Res.* **68**(16), 6770–6778 (2008).
7. Wang, X. & Kang, T. Role of CBX4 in the colorectal carcinoma metastasis-response. *Cancer Res.* **77**(9), 2550–2551 (2017).
8. Chang, S. C. *et al.* CBX3/heterochromatin protein 1 gamma is significantly upregulated in patients with non-small cell lung cancer. *Asia Pac. J. Clin. Oncol.* **14**(5), e283–e288 (2018).
9. Wheeler, L. J. *et al.* CBX2 identified as driver of anoikis escape and dissemination in high grade serous ovarian cancer. *Oncogenesis.* **7**(11), 92 (2018).
10. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **68**(1), 7–30 (2018).
11. Singh, P. B. *et al.* A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants. *Nucleic Acids Res.* **19**(4), 789–794 (1991).
12. Johnston, S. R. Clinical efforts to combine endocrine agents with targeted therapies against epidermal growth factor receptor/human epidermal growth factor receptor 2 and mammalian target of rapamycin in breast cancer. *Clin. Cancer Res.* **12**(3 Pt 2), 1061s–1068s (2006).
13. Chavez-MacGregor, M. *et al.* Incorporating tumor characteristics to the American Joint Committee on Cancer Breast Cancer Staging System. *Oncologist* **22**(11), 1292–1300 (2017).
14. Pique, D. G., Montagna, C., Grealley, J. M. & Mar, J. C. A novel approach to modelling transcriptional heterogeneity identifies the oncogene candidate CBX2 in invasive breast carcinoma. *Br. J. Cancer.* **120**(7), 746–753 (2019).
15. Zheng, S. *et al.* Overexpression of CBX2 in breast cancer promotes tumor progression through the PI3K/AKT signaling pathway. *Am. J. Transl. Res.* **11**(3), 1668–1682 (2019).
16. Zeng, J. S. *et al.* CBX4 exhibits oncogenic activities in breast cancer via Notch1 signaling. *Int. J. Biochem. Cell Biol.* **95**, 1–8 (2018).
17. Lee, Y. H. *et al.* HP1 $\beta$  is a biomarker for breast cancer prognosis and PARP inhibitor therapy. *PLoS ONE* **10**(3), e0121207 (2015).
18. Crawford, N. T. *et al.* TBX2 interacts with heterochromatin protein 1 to recruit a novel repression complex to EGR1-targeted promoters to drive the proliferation of breast cancer cells. *Oncogene* **38**(31), 5971–5986 (2019).
19. Deng, H. *et al.* CBX6 is negatively regulated by EZH2 and plays a potential tumor suppressor role in breast cancer. *Sci. Rep.* **9**(1), 197 (2019).
20. Yan, L., Zhang, Z., Yin, X. & Li, Y. LncRNA NEAT1 facilitates cell proliferation, invasion and migration by regulating CBX7 and RTCB in breast cancer. *OncoTargets Ther.* **13**, 2449–2458 (2020).
21. Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**(1), 1–6 (2004).
22. Tang, Z. *et al.* GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**(W1), W98–W102 (2017).
23. Jezequel, P. *et al.* Bc-GenExMiner 3.0: new mining module computes breast cancer gene expression correlation analyses. *Database (Oxford)* **2013**, bas060 (2013).
24. Jezequel, P. *et al.* Bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. *Breast Cancer Res. Treat.* **131**(3), 765–775 (2012).
25. Nagy, A., Lanczky, A., Menyhart, O. & Gyorfy, B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci. Rep.* **8**(1), 9227 (2018).
26. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**(5), 401–404 (2012).
27. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**(269), pl1 (2013).

28. Pereira, B. *et al.* The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat. Commun.* **7**, 11479. <https://doi.org/10.1038/ncomms11479> (2016).
29. Vad-Nielsen, J. *et al.* Regulatory dissection of the CBX5 and hnRNPA1 bi-directional promoter in human breast cancer cells reveals novel transcript variants differentially associated with HP1 $\alpha$  down-regulation in metastatic cells. *BMC Cancer*. **16**, 32 (2016).
30. De Koning, L. *et al.* Heterochromatin protein 1 $\alpha$ : a hallmark of cell proliferation relevant to clinical oncology. *EMBO Mol. Med.* **1**(3), 178–191 (2009).
31. Kirschmann, D. A. *et al.* Down-regulation of HP1H $\alpha$  expression is associated with the metastatic phenotype in breast cancer. *Cancer Res.* **60**(13), 3359–3363 (2000).
32. Dialynas, G. K., Vitalini, M. W. & Wallrath, L. L. Linking heterochromatin protein 1 (HP1) to cancer progression. *Mutat. Res.* **647**(1–2), 13–20 (2008).
33. Huang, F. J., Dang, J. Q., Zhang, S. & Cheng, Z. Y. Circular RNA hsa\_circ\_0008039 promotes proliferation, migration and invasion of breast cancer cells through upregulating CBX4 via sponging miR-515-5p. *Eur. Rev. Med. Pharmacol. Sci.* **24**(4), 1887–1898 (2020).
34. Meseure, D. *et al.* Expression of ANRIL-polycomb complexes-CDKN2A/B/ARF genes in breast tumors: identification of a two-gene (EZH2/CBX7) signature with independent prognostic value. *Mol. Cancer Res.* **14**(7), 623–633 (2016).
35. Mahauad-Fernandez, W. D., Borchering, N. C., Zhang, W. & Okeoma, C. M. Bone marrow stromal antigen 2 (BST-2) DNA is demethylated in breast tumors and breast cancer cells. *PLoS ONE* **10**(4), e0123931 (2015).
36. Mahauad-Fernandez, W. D., DeMali, K. A., Olivier, A. K. & Okeoma, C. M. Bone marrow stromal antigen 2 expressed in cancer cells promotes mammary tumor growth and metastasis. *Breast Cancer Res.* **16**(6), 493 (2014).
37. Cai, D. *et al.* Up-regulation of bone marrow stromal protein 2 (BST2) in breast cancer with bone metastasis. *BMC Cancer*. **9**, 102 (2009).
38. Kim, H. Y., Park, J. H., Won, H. Y., Lee, J. Y. & Kong, G. CBX7 inhibits breast tumorigenicity through DKK-1-mediated suppression of the Wnt/ $\beta$ -catenin pathway. *FASEB J.* **29**(1), 300–313 (2015).
39. Chung, C. Y. *et al.* Cbx8 acts non-canonically with Wdr5 to promote mammary tumorigenesis. *Cell Rep.* **16**(2), 472–486 (2016).
40. Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**(4), 664–678 (2012).
41. Reedijk, M. Notch signaling and breast cancer. *Adv. Exp. Med. Biol.* **727**, 241–257 (2012).

## Acknowledgements

The authors thank AiMi Academic Services ([www.aimieditor.com](http://www.aimieditor.com)) for English language editing and review services.

## Author contributions

Xiaomin Li conceptualized, investigated and wrote the original draft. Junhe Gou and Hongjiang Li collected the experimental materials. Xiaoqin Yang reviewed and edited the manuscript. All authors read and approved the final manuscript.

## Funding

This study was sponsored by funds from the Key Scientific Research Project of Sichuan Provincial Science and Technology Department (2019YFS0377).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-74792-5>.

**Correspondence** and requests for materials should be addressed to X.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

© The Author(s) 2020