Genotranscriptomic meta-analysis of the Polycomb gene CBX2 in human cancers: initial evidence of an oncogenic role

P-L Clermont 1,2, L Sun 3, F Crea 1,4, K L Thu 5, A Zhang 1, A Parolia 1,6, W L Lam 5 and C D Helgason * 1,4

1 Department of Experimental Therapeutics, BC Cancer Research Centre, 675 West 10th Avenue, Vancouver, BC V5Z 1L3, Canada; 2 Interdisciplinary Oncology Program, Faculty of Medicine, University of British Columbia, 675 West 10th Avenue, Vancouver, BC V5Z 1L3, Canada; 3 Mouse Cancer Genetics Program, National Cancer Institute, National Institutes of Health, 1050 Boyles Street, Frederick, MD 21702-1201, USA; 4 Department of Surgery, University of British Columbia, 950 West 10th Avenue, Vancouver, BC V5Z 1M9, Canada; 5 Genetics Unit, Department of Integrative Oncology, BC Cancer Research Centre, 675 West 10th Avenue, Vancouver, BC V5Z 1L3, Canada and 6 Honours Biotechnology Program, Department of Microbiology and Immunology, University of British Columbia, 2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada

Background: Polycomb group (PcG) proteins are histone modifiers known to transcriptionally silence key tumour suppressor genes in multiple human cancers. The chromobox proteins (CBX2, 4, 6, 7, and 8) are critical components of PcG-mediated repression. Four of them have been associated with tumour biology, but the role of CBX2 in cancer remains largely uncharacterised.

Methods: Addressing this issue, we conducted a comprehensive and unbiased genotranscriptomic meta-analysis of CBX2 in human cancers using the COSMIC and Oncomine databases.

Results: We discovered changes in gene expression that are suggestive of a widespread oncogenic role for CBX2. Our genetic analysis of 8013 tumours spanning 29 tissue types revealed no inactivating chromosomal aberrations and only 40 point mutations at the CBX2 locus. In contrast, the overall rate of CBX2 amplification averaged 10% in all combined neoplasms but exceeded 30% in ovarian, breast, and lung tumours. In addition, transcriptomic analyses revealed a strong tendency for increased CBX2 mRNA levels in many cancers compared with normal tissues, independently of CDKN2A/B silencing. Furthermore, CBX2 upregulation and amplification significantly correlated with metastatic progression and lower overall survival in many cancer types, particularly those of the breast.

Conclusions: Overall, we report that the molecular profile of CBX2 is suggestive of an oncogenic role. As CBX2 has never been studied in human neoplasms, our results provide the rationale to further investigate the function of CBX2 in the context of cancer cells.

Keywords: CBX2; polycomb; metastasis; prognosis; expression; breast; prostate; lung; meta-analysis

Recent advances in DNA sequencing technologies have allowed unprecedented insights into the genomic landscape of human tumours (Meyerson et al., 2010). A unifying theme is emerging: cancers display extensive inter- and intra-tumoral genetic heterogeneity, with very few neoplasms resulting from a single recurrent genomic aberration (Gerlinger et al., 2012; Swanton, 2012).

In parallel with those observations, many lines of evidence suggest that cancers may be driven by alterations that are epigenetic in nature, that is, they do not alter the primary DNA sequence (McDonald et al., 2011; Timp and Feinberg, 2013). As epigenetic modifications are reversible and can be pharmacologically targeted (Crea et al., 2011; Kim et al., 2013),
identifying the epigenetic drivers of cancer progression is thus a matter of utmost importance.

In the past decade, a number of chromatin regulators known to have key molecular functions in embryonic development have been identified as novel cancer-promoting genes (Varambally et al, 2002; You et al, 2009). These epigenetic factors are thought to retain cancer cells in an undifferentiated state, thus enhancing their metastatic potential and resistance to treatment (Ben-Porath et al, 2008; Ohm et al, 2007). The Polycomb group (PcG) family represents the typical example of such genes, as they can silence lineage-specific genes both in embryonic stem cells and multiple cancer types (Bracken and Helin, 2009). Polycomb group genes encode transcriptional repressors that assemble in a complex combinatorial manner to form two main Polycomb repressive complexes (PRC1 and PRC2) (Satijn et al, 1997; Kuzmichev et al, 2002).

In the classical model of PcG-mediated silencing, PRC2 trimethylates histone H3 at lysine 27 (H3K27me3) through the action of its catalytically active subunit EZH2 (Cao et al, 2002). H3K27me3 can be directly recognised by one of five chromodomain-containing proteins (CBX2, 4, 6, 7 and 8), which subsequently recruit PRC1 to chromatin by simultaneously interacting with the E3 ubiquitin ligase Ring1B through a C-terminal domain (Kaustov et al, 2011). This interaction brings Ring1B to H3K27me3 sites, where Ring1B ubiquitylates lysine 119 on histone H2A (H2AK119ub), further repressing transcription at target loci (van der Stoop et al, 2008). Interestingly, PRCs can also act independently of each other as recent studies have shown that PRC1 can silence genomic regions that are not marked by H3K27me3 (Tavares et al, 2012). Moreover, some CBX-containing PRC1 complexes are devoid of ubiquitin ligase activity despite retaining gene-silencing properties (Gao et al, 2012). Although PcG proteins may act through different pathways in different cancer types, PcG gain of function generally associates with an undifferentiated cellular state and aggressive clinical behaviour (Boyer et al, 2006).

Whereas EZH2 and BMI-1 have been the two most heavily investigated PcG genes in the context of human neoplasms (Bachmann et al, 2006; Cao et al, 2011), emerging evidence supports a critical role for the CBX proteins in cancer initiation and progression. All CBX family members share a conserved N-terminal chromodomain but display non-homologous sequences in their C-terminus, accounting for their non-redundant functions (Vincenz and Kerppola, 2008). Distinct CBX proteins interact with a wide range of macromolecules including DNA, non-coding RNAs, and numerous other proteins (Bernstein et al, 2006; Vandamme et al, 2011). Furthermore, individual CBX family members can be differentially expressed, undergo alternative splicing, harbour distinct post-translational modifications, and lie under the control of different microRNAs (O’Loghlen et al, 2012). All of these regulatory steps affect the function of individual CBX protein and allow for tremendous complexity in PRC1 activity and sequence specificity.

Individual CBX proteins are defined by distinct C-terminal sequences that underlie their specific properties. The structural differences between the CBX family members are also reflected in the diversity of molecular functions they can accomplish in the context of cancer cells. For example, CBX4 possesses SUMO activity and is thought to have important roles in cellular proliferation and DNA damage repair in some human tumours (Yang et al, 2011; Wang et al, 2013). Conversely, SNPs in CBX6 have been reported in genome-wide association studies (GWASs) of bladder cancer although their functional implications have not been fully elucidated (Rothman et al, 2010). CBX7 has been the most extensively investigated CBX family member and exhibits cancer type-specific activity in human tumours. Most studies report a widespread oncosuppressive function (Forzati et al, 2012a), notably in brain (Gargiulo et al, 2013), colon (Pallante et al, 2010), and lung cancers (Forzati et al, 2012b). However, other investigations have revealed an oncogenic role for CBX7 (Mohammad et al, 2009; Zhang et al, 2010) occurring mainly through silencing of CDKN2A and CDKN2B loci, which encode the cyclin-dependent kinase inhibitors p14ARF, p15INK4B, and p16INK4A (Bernard et al, 2005). Finally, CBX8 was found to be essential for MLL/AF9 leukemogenesis (Tan et al, 2011) in addition to its ability to silence the CDKN2A locus (Dietrich et al, 2007). Despite solid evidence implicating these four CBX family members in neoplastic transformation, no study has directly addressed the role of CBX2 in human cancers.

Emerging evidence supports a role for CBX2 in cellular proliferation. Loss of M33, the murine CBX2 ortholog, impairs progression through the S phase of cell cycle in an E2F-dependent mechanism, which is consistent with the decreased cellularity of hematopoietic organs in M33-KO mice (Coré et al, 2004). In addition, CBX2 directly binds the tumour-suppressive CDKN2A and CDKN2B loci in young and proliferative fibroblasts but progressively becomes absent from this chromosomal region as the cells undergo senescence and cell cycle arrest (Agherbi et al, 2009). Furthermore, experiments conducted in mice have demonstrated that unphosphorylated M33 is present in the cytoplasm of mature hepatocytes, but becomes phosphorylated and translocates to the nucleus in proliferative hepatocytes during the liver regeneration (Noguchi et al, 2002). Interestingly, studies have shown that CBX2 phosphorylation also increases its affinity to H3K27me3 in addition to inducing its nuclear translocation, which is consistent with cell cycle-dependent regulation of CBX2 activity (Hatano et al, 2010). Finally, Grau et al have demonstrated that CBX2 is the only human CBX family member able to induce chromatin compaction (Grau et al, 2011). Elegant studies based on electron microscopy have revealed that this feature of CBX2 has mediated a highly positively charged region located in its C-terminus not found in any other PcG member, thus suggesting a unique and crucial function for CBX2 in PcG-mediated repression.

On the basis of the critical role played by CBX2 in cellular proliferation and the emerging evidence implicating Polycomb-mediated silencing in tumour biology, we postulated that CBX2 may represent an important component in cancer initiation and progression. Given the lack of literature addressing the implications of CBX2 in a neoplastic context, we conducted a meta-analysis of CBX2 in human cancers at the genomic and transcriptomic level using publicly available databases. We report that CBX2 downregulation and inactivating genetic mutations represent extremely rare events in human tumours. Furthermore, we also provide the first evidence that CBX2 genomic amplification and mRNA upregulation predict metastatic progression and poor overall survival (OS) in multiple cancer types, notably those arising from the breast.

**MATERIALS AND METHODS**

**COSMIC database analysis.** The COSMIC database (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/) (Forbes et al, 2008) was used to extract all data relevant to our analysis of genetic alterations at the CBX2 locus. All data were collected after 20 October 2013, and a final revision of all acquired data was performed between 1 July and 21 July 2013. All data originating from the COSMIC database are specifically mentioned either in the main text, the figure legend, or both. All the raw data extracted from the COSMIC database can be found in Supplementary Information.

**Oncomine database analysis.** All transcriptomic data used in this manuscript were extracted from the Oncomine platform (www.oncomine.com) (Rhodes et al, 2004) after 20 October 2013, and a
final revision of all acquired data was performed between 1 July and 21 July 2013. Data were acquired in an unbiased manner by compiling all the Oncomine studies that showed significantly altered CBX2 expression at the threshold set for each individual analysis. Significant studies in which at least one analysed group was comprised of three patients or less were excluded. All data originating from the Oncomine database are specifically mentioned either in the main text, the figure legend, or both. All the raw data extracted from the Oncomine database can be found in Supplementary Information.

**Human Protein Atlas database analysis.** The image of CBX2 protein sequence and domains used in Supplementary Figure S1 was obtained from the Protein Atlas database (http://www.proteinatlas.org/) (Uhlen et al, 2010).

**Assessment of clinical correlations for genes within the minimal common region (MCR) of CBX2 amplification.** TCGA breast cancer copy number data were downloaded from Cancer Genomics Browser (https://genome-cancer.ucsc.edu/) and all tumours containing CBX2 gains were determined. We filtered the segments with gain (logR > 0.3) and at least 100 kb in size and mapped the MCR, defined by the sample with the smallest region containing CBX2. We then determined the Spearman copy number–expression correlation for genes mapping to the region assessed. Those with significant correlation ($P$ < 0.05 Spearman) were further analysed for survival associations using the Curtis Breast data set on Oncomine and all breast cancer samples from the KMplotter resource (http://kmplot.com/analysis/) (Mihaly et al, 2013).

**Statistical analysis.** Unless otherwise mentioned, all analyses were done using $P \leq 0.05$ as the significance threshold. GraphPad Prism software (version 6, GraphPad Software Inc., La Jolla, CA, USA) was used for all statistical analyses except for the multivariate analysis of variance (MANOVA) and the Cox proportional hazards (COXPH) regression, which were done using R statistical software (GNU, freely available at http://www.r-project.org/).

**RESULTS**

**Genomic analysis of the CBX2 locus.** As the first step of our systematic meta-analysis to study CBX2 expression in human cancers, we analysed genomic alterations at the CBX2 locus using the COSMIC database, a resource designed to store and display somatic mutation information and related details (Forbes et al, 2008). Strikingly, there was an extremely low frequency of alterations disrupting CBX2 function. In 8013 tumour samples spanning 29 tissue types, we did not find a single translocation, homozygous loss, insertion/deletion, or any other inactivating large-scale chromosomal abnormalities (Figure 1A). In total, only 40 point mutations were recorded at the CBX2 gene in these tumours (overall frequency = 0.5%). The CBX2 mutation frequency is thus considered very low, as high-frequency mutations are typically described as being over 20% and intermediate frequency as being between 2 and 20% (Lawrence et al, 2014). Further analysis revealed that of those 40 mutations, 2 were nonsense and 22 were missense, whereas 17 were silent mutations, which did not have any effect on CBX2 protein sequence (Figure 1B). These point mutations were distributed evenly across all tissues, with no tumour type having a marked increase in CBX2 mutation frequency (Supplementary Figure S1). The mutations were concentrated in regions where CBX2 is predicted to be highly disordered and hydrophilic, and within these regions their distribution was relatively homogenous (Supplementary Figure S2). In addition, no single residue within the CBX2 protein was mutated more than twice. Interestingly, there was a complete absence of point mutations in the CBX2 chromodomain, the region responsible for H3K27me3 binding (Supplementary Figure S2). Overall, the lack of large-scale genomic aberration and the non-synonymous mutation frequency of 0.3% make it unlikely that alterations at the CBX2 locus impact cancer cell phenotype.

The low frequency of genomic CBX2 disruption led us to investigate whether CBX2 copy number increases are favoured during oncogenesis. To address this question, we first analysed the amplification profile of the CBX2 locus in human neoplasms using the COSMIC database. In contrast with the rare nature of CBX2-inactivating mutations, we uncovered that CBX2 gene amplifications occur frequently in a number of tumour types. Overall, 714 out of 8013 samples from the COSMIC database had undergone copy number gain (CNG) at the CBX2 locus (overall frequency: 8.9%; see Figure 2 and Supplementary Table S1). Interestingly, the distribution of these amplifications was not homogenous across all tumour types. We observed five neoplasms in which the CBX2 CNG frequency ranged between 3 and 15%: those originating from the central nervous system, colon, endometrium, pancreas, and kidneys (Figure 2 and Supplementary Table S1). Furthermore, three cancer types harboured a frequency of CBX2 amplification > 30%: tumours of the ovaries (34.0%), breast (34.5%), and lungs (35.5%), suggesting that CBX2 copy number increases may provide a selective advantage to cancer cells.

**Transcriptomic analysis of CBX2 expression in human cancers.** As our genomic analysis revealed recurrent CNGs and very rare

---

**Figure 1.** Extremely rare occurrence of genetic mutations disrupting CBX2 function in human cancers (COSMIC database) (A) Percentage of specific inactivating genetic alterations at the CBX2 locus. (B) Distribution and frequency of CBX2 point mutations.
inactivating mutations at the CBX2 locus, we next investigated whether this trend would also be reflected at the mRNA level. Using the Oncomine database (Supplementary Table S2) (Rhodes et al., 2004), we identified a total of 25 studies that showed significant upregulation (FC > 2, P value < 0.001, top 10% over/underexpressed) in cancer compared with normal tissue (Figure 3 and Supplementary Table S3). Strikingly, not a single study reported downregulation of CBX2 using the same inclusion criteria (Figure 3), once again implying an important functional role in cancer cells. The total number of patients in the 25 studies showing CBX2 upregulation in cancer tissues is 3848 compared with 0 for CBX2 downregulation (Figure 3 and Supplementary Table S3). In the studies harbouring CBX2 overexpression, fold changes varied between 2.1 and 15, and the P values between 4.0E-3 and 3.6E-73 (Figure 3). The most represented cancer types in the CBX2-overexpressed studies were those originating from the colon (29.6%), breast (18.5%), stomach (14.8%), and lungs (11.1%). These results demonstrate a clear bias towards CBX2 upregulation and complement the genomic analysis that hinted towards a selective pressure to maintain CBX2 function.

Polycomb group complexes are known to repress the tumour-suppressive loci CDKN2A (encoding p14ARF and p16INK4A) and CDKN2B (encoding p15INK4B) in many human cancers (Aguiló et al., 2011). We thus sought to determine whether CBX2 overexpression correlated with silencing of the CDKN2A/B genes. Interestingly, we found that neither p14ARF nor p16INK4A were downregulated in any of the 25 studies with CBX2 overexpression (Supplementary Figure S3 and Supplementary Table S4). However, p15INK4B was found to be downregulated in 10 of those 25 studies (40%) using the same cut-off as for CBX2 (FC > 2, P value < 0.01, top 10% underexpressed). Further analysis revealed that 8 of the 10 studies with concomitant CBX2 upregulation and p15INK4B downregulation occurred specifically in the colorectal cancer (Supplementary Figure S3 and Supplementary Table S4). However, when using Spearman correlations to investigate the direct relationship between CBX2 and CDKN2A/B, not a single study showed a statistically significant correlation (Table 1, Spearman correlation of $R^2 > 0.1$, even in colorectal cancer). Taken together, analysis of CBX2 mRNA levels revealed lack of CBX2 downregulation contrasted by frequent CBX2 overexpression, an event that occurred independently of CDKN2A/B silencing.

**Clinical correlations of differential CBX2 expression.** Given the widespread CBX2 upregulation in cancerous compared with normal tissues, we investigated whether CBX2 expression was also correlated with clinical indicators of tumour progression. Metastasis was the first parameter we assessed. Using the Oncomine database (Rhodes et al., 2004), we found nine studies exhibiting significantly (FC > 1.5, P ≤ 0.05) increased CBX2 mRNA levels in metastatic compared with primary tumours (Table 2). Prostate cancer (PCA) was the most represented tumour type, accounting for three of the five most significant studies. Sarcoma and breast cancer followed with two studies each displaying CBX2 upregulation in metastatic disease. In contrast with the strong CBX2 upregulation observed in disseminated tumours, not a single study with CBX2 underexpression could be detected using the same parameters (Table 2), consistent with a selection against CBX2 loss of function in metastatic cells.

Next, we investigated whether differential CBX2 expression and genomic copy number were associated with poor clinical outcome in human malignancies. We queried the Oncomine database for studies in which CBX2 mRNA levels were significantly altered between alive and deceased patients at 1, 3, and 5 years after diagnosis using an FC > 1.2 and $P < 0.05$ as inclusion criteria (Supplementary Table S5). After 1 year, the number of studies with CBX2 overexpression (eight) was almost two-fold higher than those with CBX2 underexpression (five) (Figure 4A and Supplementary Table S5). Furthermore, there was a striking difference between the number of studies in which CBX2 mRNA was upregulated compared with those which displayed CBX2...
downregulation in patients who were deceased compared with those who were alive at 3 and 5 years post diagnosis (17/1 and 16/2, respectively, see Figure 4A and Supplementary Table S3).

To ensure that the number of studies showing CBX2 upregulation was higher than those with CBX2 downregulation across all three time points, we performed a chi-square test and found a statistically significant difference between the studies with high and low CBX2 (P < 0.00001, chi-square). Interestingly, some cancer types in particular showed a recurrent inverse correlation between high CBX2 mRNA levels and lower OS. Notably, tumours of the breast (five and five), central nervous system (two and two), and lungs (one and three) were among the most represented cancer types for this analysis at 3 and 5 years, respectively (Figure 4A). We also found three studies of head and neck tumours in which significantly decreased, not increased, levels of CBX2 led to lower OS. In addition, CBX2 CNG could also be associated with clinical outcome. Analysis of the Oncomine database revealed an increase in CBX2 copy number in metastatic compared with primary pancreatic cancer (Supplementary Figure S4, P < 0.001). Moreover, CBX2 CNG was also inversely correlated with OS in oligoastrocytoma and ovarian cancer at 3 years (Supplementary Figure S4, P = 0.02 for both), and in oligoastrocytoma at 5 years (Supplementary Figure S4, P = 0.02). Thus, our results demonstrate that both CBX2 gene amplification and mRNA upregulation may harbour prognostic significance.

To further assess the relationship between CBX2 and individual clinical variables, we conducted a MANOVA in one cohort for each of the breast, lung, and colorectal cancers from the Oncomine platform. All significant covariates (P < 0.05) in the MANOVA were further assessed using univariate analyses. We found that CBX2 was significantly associated with sex in the colon cancer, with higher expression observed in females (Supplementary Table S6, TCGA colorectal, P < 0.01, Mann–Whitney U-test). In the analysed lung cancer data set, subtype was the only clinical covariate associating with CBX2, with higher CBX2 expression in squamous cell carcinoma compared with lung adenocarcinoma (Supplementary Table S7, Bild Lung, P < 0.05, Mann–Whitney U-test). More specifically, higher CBX2 expression correlated with younger age, basal-like subtype, and higher grade, all of which are linked to poor patient prognosis. We therefore conducted a COXPH regression on the Curtis Breast data set to further explore the relationship between CBX2 and clinical outcome. Interestingly, we found that behind age, CBX2 was the second covariate most significantly associated with patient survival (Table 3, COXPH, P < 0.001), suggesting a role for CBX2 in promoting aggressive disease progression.

Finally, we investigated whether CBX2 is a potential driver of the recurrent 17q25.3 copy gains we observed. We used multiple criteria to investigate genes mapping to the MCR of amplification containing CBX2, including copy number expression association, expression in tumours compared with normals, and survival association, as these features could highlight the gene that is most likely to be driving the 17q25.3 amplicon. As CBX2 was recurrently amplified and differentially expressed in breast cancer, we investigate the clinical implications of the CBX2 amplicon in breast cancer. We first used breast cancer TCGA copy number data from the Cancer Genomics Browser (Zack et al, 2013; https://genome-cancer.ucsc.edu/) to identify all gained genomic segments containing CBX2. We then filtered the segments with gain (defined

| Table 1. Spearman correlation between CBX2 and CDKN2A/B in the data sets displaying CBX2 upregulation in malignant compared with normal tissues |
|-----------------|----------------|-----------------|-----------------|----------------|
| Data set        | Patients, n   | R² CDKN2A       | R² CDKN2B       |
| Brune lymphoma  | 42            | 0.00            | 0.05            |
| Cho gastric     | 71            | 0.02            | 0.07            |
| Crabtree uterus | 50            | 0.05            | 0.01            |
| Curtis breast   | 1992          | 0.01            | 0.00            |
| Dernico gastric | 38            | 0.02            | 0.00            |
| Hong colorectal | 70            | 0.04            | 0.00            |
| Hou lung        | 91            | 0.00            | 0.01            |
| Kerkola seminoma| 101           | 0.02            | 0.01            |
| Lee bladder     | 188           | 0.01            | 0.04            |
| Richardson breast 2 | 40       | 0.12            | 0.02            |
| Skrzypczak colorectal | 81   | 0.01            | 0.03            |
| Skrzypczak colorectal 2 | 20   | 0.03            | 0.01            |
| Sun brain       | 157           | 0.01            | 0.00            |
| TCGA breast     | 532           | 0.11            | 0.05            |
| TCGA colorectal | 215           | 0.00            | 0.00            |

Abbreviation: CDKN – cyclin-dependent kinase inhibitor.

| Table 2. List of studies with differential CBX2 expression between metastatic and primary tumours |
|-----------------|----------------|----------------|-----------------|----------------|
| Alteration      | Rank | Study name     | Cancer type     | Primary, n | Metastatic, n | Total, n | Fold change | P-value |
| Upregulation    | 1    | Chandran prostate | Prostate cancer | 10     | 21           | 31      | 7.8 | 1.7E-12 |
|                 | 2    | Bittner breast  | Breast cancer   | 327    | 9            | 336     | 1.9 | 3.2E-05 |
|                 | 3    | Varambally prostate | Prostate cancer | 7     | 6            | 13      | 9.2 | 1.3E-04 |
|                 | 4    | Segal sarcoma 2 | Sarcoma         | 46     | 4            | 50      | 13  | 4.4E-04 |
|                 | 5    | Grasso prostate | Prostate cancer | 13     | 17           | 30      | 1.9 | 4.0E-03 |
|                 | 6    | Riker melanoma | Melanoma        | 16     | 40           | 56      | 3.6 | 6.0E-03 |
|                 | 7    | Radvanyi breast | Breast cancer   | 41     | 4            | 45      | 3.1 | 9.0E-03 |
|                 | 8    | Bittner sarcoma | Sarcoma         | 42     | 10           | 52      | 1.7 | 1.1E-02 |
|                 | 9    | Liao liver     | Liver cancer    | 4      | 6            | 10      | 4.9 | 2.3E-02 |

Oncomine database; inclusion criteria: FC ≥ 1.5, P ≤ 0.05, top 10% under/overexpressed, Student’s t-test.
by having a log ratio > 0.3) and at least 100 kb in size and mapped the MCR, which is defined by the sample with the smallest region containing CBX2. We found that the MCR contained three genes: ENPP7, CBX2, and CBX8, present at 17q25.3 (Figure 5A and B). We next assessed the correlation between copy number and the expression of genes mapping to the region using Spearman correlation. ENPP7 had a negative correlation, indicating that its amplification did not result in increased expression. In contrast, a weak but significant positive correlation was observed for both CBX2 and CBX8 (Supplementary Table S9, P < 0.0001, Spearman).

To determine whether either of the two genes was associated with patient survival, we performed a Mantel–Cox log-rank test for both CBX2 and CBX8, comparing survival of patients with expression ranking in the top and bottom quartiles of expression. We report that elevated CBX2 levels, but not CBX8, significantly predicts lower OS in breast cancer (Figure 5C and D, CBX2 P < 0.0001; CBX8 P = 0.49, Mantel–Cox log-rank test). We also performed log-rank test on CBX2 and CBX8 using the KMplotter (http://kmplot.com/analysis/) (Mihaly et al., 2013). Once again, we found that CBX2 was the only gene within the MCR that was able to significantly predict lower OS (Figure 5E and F, P < 0.05, log-rank test). Overall, we demonstrate that CBX2 is frequently gained in breast cancer, which leads to its increased expression and associates with poor patient prognosis, suggesting that it is a candidate driver of the 17q25.3 amplicon.

**Figure 4. Differential CBX2 expression predicts OS.** (A) Number of oncomine studies with either significant CBX2 up- or downregulation (FC > 1.2, P < 0.05, Student’s t-test) in patients who are dead compared with alive at 1, 3, and 5 years post diagnosis (CBX2 high vs CBX2 low; P < 0.00001, chi-square test). (B) Sex-specific CBX2 expression in TCGA colon (P = 0.0015, Mann–Whitney U-test). (C) Subtype-specific CBX2 expression in Bild Lung (P = 0.03, Mann–Whitney U-test). (D) Subtype-specific CBX2 expression in Curtis Breast (P < 0.0001, Kruskal–Wallis test). (E) Grade-specific CBX2 expression in Curtis Breast (P < 0.0001, Kruskal–Wallis test).

| Curtis breast | Coef | s.e. (coef) | z | P-value |
|---------------|------|-------------|---|---------|
| Normal subtype | 0.2839 | 0.0133 | 2.76 | 0.0058 |
| Luminal A subtype | 0.2954 | 0.0128 | 2.32 | 0.0202 |
| Luminal B subtype | 0.2814 | 0.0131 | 2.19 | 0.0294 |
| Her2 subtype | 0.2784 | 0.0130 | 2.16 | 0.0303 |
| Basal subtype | 0.2772 | 0.0127 | 2.15 | 0.0313 |
| Grade 3 | 0.2813 | 0.0129 | 2.16 | 0.0307 |
| Grade 2 | 0.2839 | 0.0131 | 2.19 | 0.0301 |
| Grade 1 | 0.2839 | 0.0131 | 2.19 | 0.0301 |
| CBX2 | 0.2839 | 0.0131 | 2.19 | 0.0301 |

Abbreviations: CBX2 = chromobox protein 2; Coef = coefficient; exp = exponent.
In summary, genomic analysis has revealed that genetic alterations resulting in loss of function represent extremely infrequent events at the CBX2 locus. In contrast, recurrent amplifications were observed in multiple cancer types. In addition, transcriptomic analysis revealed a propensity for CBX2 upregulation in the following four independent scenarios: (1) cancer vs normal tissues; (2) metastatic vs primary tumours; (3) dead vs alive at 3 years; and (4) dead vs alive at 5 years (Supplementary Table S10). Surprisingly, with the exception of CDKN2B in the colorectal cancer, CBX2 overexpression was not correlated with silencing of the tumour-suppressive CDKN2A or CDKN2B loci.

**DISCUSSION**

Our results demonstrate a clinically relevant increase in CBX2 copy number and expression in human cancer. Given the very low frequency of CBX homozygous loss, point mutation, and under-expression, we believe CBX2 is likely to have an important functional role in tumour cells. In parallel, transcriptomic analysis of CBX2 expression revealed a strong bias towards CBX2 upregulation, although this alteration was observed at different stages among the various tumour types, indicating some context specificity in CBX2 activity and binding partners. Overall, breast cancer displayed the most significant associations with CBX2 alterations, exhibiting an overall frequency of CBX2 amplification exceeding 30%. We demonstrated that CBX2 copy number increases correlated with increased CBX2 expression, which is significantly associated with lower OS. Of all the genes present in the MCR containing CBX2, only CBX2 has prognostic relevance, suggesting that CBX2 displays a driving role in the progression of breast cancer. As CBX2 and CBX8 share a conserved chromodomain but largely differ in their C-terminus (Vincenz and Kerppola, 2008), we believe further investigation is required to elucidate the structural differences that may underlie the potential cancer-promoting role of CBX2.

Given that many breast cancers also harbour EZH2 gain of function (Granit et al, 2013; Kleer et al, 2003), an interesting possibility is that CBX2 upregulation represents a key adaptation necessary to perpetuate EZH2's oncogenic activity. This therefore implies an important role for the ability of CBX2 to interact with H3K27me3 and is consistent with our observation that missense mutations are excluded from the chromodomain of CBX2. Given that chromodomains act as protein–RNA binding modules...
Prostate cancer is another neoplasm in which EZH2 over-expression represents a key hallmark (Varambally et al, 2002; van Leenders et al, 2007), and interestingly it was the cancer type in which CBX2 upregulation most significantly correlated with metastatic progression. The frequent upregulation of many PcG family members (van Leenders et al, 2007), coupled with the prognostic significance of a Polycomb repressive signature in PCA (Yu et al, 2007), support the idea that PcG complexes undergo gain of function in metastatic disease. As opposed to breast cancer, we could not find any significant differential CBX2 expression in neoplastic tissues compared with normal ones, indicating that CBX2 upregulation likely represents a late event in PCA progression. Whereas the presence of metastases represents the most valuable indicator of PCA prognosis, we did not find any correlations between CBX2 expression and OS. We attribute this result to a technical limitation of the Oncomine database, which does not allow for an in-depth characterisation of subtype-specific CBX2 expression. For example, it is thus possible that CBX2 expression is markedly high in one molecularly or histologically characterised subtype while being low in another from a given neoplasm, something that could not be determined from our analysis. Nonetheless, by using the COSMIC and Oncomine databases we were able to analyse over 40,000 patient samples in a fully unbiased manner, which allowed us to observe a genotranscriptomic profile for CBX2 that was consistent with that of an oncogene.

Finally, the very low frequency of CBX2 mutation and under-expression suggest that the therapeutic inhibition of CBX2 might represent a valuable clinical strategy for the treatment of many human cancers. We report aberrant regulation of CBX2 in the breast, lung, colorectal, prostate, brain, and haematopoietic tumours, all of which rank among the 10 deadliest neoplastic diseases worldwide and are in dire need of novel therapeutics. Furthermore, epigenetic alterations are reversible and recent studies have highlighted the possibility of targeting histone readers with small-molecule inhibitors (Dawson et al, 2012; Tabet et al, 2013). Taken together, our work has identified a putative oncogenic role for CBX2 in numerous tumour types and has provided the rationale for the design of novel CBX2-targeting therapies.

ACKNOWLEDGEMENTS

This work was supported by Canadian Cancer Society Research Institute grant number 701097 (CDH) and a Kruger Graduate Fellowship (PLC).

REFERENCES

Agheri H, Gaussmann-Wenger A, Verthuy C, Chasson L, Serrano M, Djabali M (2009) Polycomb mediated epigenetic silencing and replication timing at the INK4a/ARF locus during senescence. PLoS One 4(5): e5622.

Agullo F, Zhou MM, Walsh MJ (2011) Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. Cancer Res 71(16): 5365–5369.

Akhtar A, Zink D, Becker PB (2000) Chromodomains are protein-RNA recognition modules. Nature 407(6802): 405–409.

Bachmann IM, Halvorsen OJ, Collett K, Stefansson IM, Straume O, Haavik SA, Salvesen HB, Otte AP, Akslen LA (2006) EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. J Clin Oncol 24(2): 268–273.

Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, Weinberg RA (2008) An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet 40(5): 499–507.

Bernard D, Martinez-Leal JF, Rizzo S, Martinez D, Hudson D, Visakorpi T, Peters G, Carnero A, Beach D, Gil J (2005) CBX7 controls the growth of normal and tumor-derived prostate cells by repressing the Ink4a/Arf locus. Oncogene 24(36): 5543–5551.

Bernstein E, Duncan EM, Masai O, Gil J, Heard E, Allis CD (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol 26(7): 2560–2569.

Boyer LA, Plath K, Zeiингinger J, Bramblik T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Bell GW, Otte AP, Vidal M, Gifford DK, Young RA, Jaenisch R (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441(7091): 349–353.

Bracken AP, Helin K (2009) Polycomb group proteins: navigators of lineage pathways led astray in cancer. Nat Rev Cancer 9(11): 773–784.

Cao L, Bombard J, Cintron K, Sheedy J, Weetall ML, Davis TW (2011) BMI1 is a novel target for drug discovery in cancer. J Cell Biochem 112(10): 2729–2741.

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298(5595): 1039–1043.

Core N, Joly F, Boned A, Djabali M (2004) Disruption of E2F signaling suppresses the INK4a-induced proliferative defect in M3-deficient mice. Oncogene 23(46): 7660–7668.

Crea F, Hunt EM, Mathews LA, Cabarcas SM, Sun L, Marquez VE, Danesi R, Farrar WL (2011) Pharmacologic disruption of polycomb repressive complex 2 inhibits tumorigenicity and tumor progression in prostate cancer. Mol Cancer 10: 40.

Dawson MA, Kouzardides T, Huntly BJ (2012) Targeting epigenetic readers in cancer. N Engl J Med 367(7): 647–657.

Dietrich N, Bracken AP, Trinh E, Schierling CK, Koseki H, Rapposilber J, Helin K, Hansen KH (2007) Bypass of senescence by the polycomb group protein protein CBX8 through direct binding to the INK4a-ARF locus. EMBO J 26(6): 1637–1648.

Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR (2008) The catalogue of somatic mutations in cancer (COSMIC). Curr Protoc Hum Genet. Chapter 10:Unit 10.11.

Forzati F, Federico A, Pallante P, Abbate A, Esposito F, Malapelle U, Sepe R, Palma G, Troncone G, Scarfo M, Arna C, Fedele M, Fusco A (2012a) CBX7 is a tumor suppressor in mice and humans. J Clin Invest 122(2): 612–623.

Forzati F, Federico A, Pallante P, Fedele M, Fusco A (2012b) Tumor suppressor activity of CBX7 in lung carcinogenesis. Cell Cycle 11(10): 1888–1891.

Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, Kluger Y, Reinberg D (2012) PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell 45(3): 344–356.

Gargiulo G, Cesaroni M, Serresi M, de Vries N, Hulsman D, Bruggeman SW, Lancini C, van Lohuizen M (2013) In vivo RNAi screen for Bmi1 targets
identifies TGF-β/BMP-ER stress pathways as key regulators of neural- and malignant glioma-stem cell homeostasis. Cancer Cell 23(5): 660–676.

Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Guil S, Soler M, Portela A, Carre`re J, Fonalleras E, Go´mez A, Villanueva A, Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D, Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, Ghosh D, Kim W, Bird GH, Neff T, Guo G, Kerenyi MA, Walensky LD, Orkin SH (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 366(10): 883–892.

Granit RZ, Gabai Y, Hadar T, Karamansh T, Liberman L, Waldhorn I, Gal-Viks I, Ruppin A, Maly B, Darash-Yahana M, Peretz T, Ben-Porath I (2013) EZH2 promotes a b-lineage identity in basal-like breast cancers. Oncogene 32(33): 3886–3895.

Grau DJ, Chapman BA, Garlick JD, Borowski M, Francis NJ, Kingston RE (2011) Complementation of chromatin by diverse Polycomb group proteins requires localized regions of high charge. Genes Dev 25(20): 2210–2221.

Guil S, Soler M, Portela A, Carrère J, Fonalleras E, Gómez A, Villanueva A, Esteller M (2012) Intrinsic RNAs mediate EZH2 regulation of epigenetic targets. Nat Struct Mol Biol 19(7): 664–670.

Hatano A, Matsutomo M, Higashinakagawa T, Nakayama KI (2010) Phosphorylation of the chromodomain changes the binding specificity of Cbx2 for methylated histone H3. Biochem Biophys Res Commun 397(1): 93–99.

Kaustov L, Ouyang H, Amaya M, Lemak A, Nady N, Duan S, Wasney GA, Li Z, Vedadi M, Schapira M, Min J, Arrowsmith CH (2011) Recognition and specificity determinants of the human cbx chromodomains. J Biol Chem 286(1): 521–529.

Kim W, Bird GH, Neff T, Guo G, Kerenyi MA, Walsenky LD, Orkin SH (2013) Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. Nat Chem Biol 9(10): 643–650.

Kleer CG, Cao Q, Varambally S, Chen M, Yang H, Vermeulen SH, Ahen KK, Witjes JA, Makkinje RR, Sulem P, Besebacher S, Stefansson K, Riboli E, Brennan P, Panico S, Navarro C, Allen NE, Bueno-de-Mesquita HB, Trichopoulos D, Caporaso N, Landi MT, Czernin F, Ljungberg B, Tjonneland A, Clavel-Chapelon F, Bishop DT, Teo MT, Knowles MA, Guerrero A, Polidoro S, Ricceri F, Sacerdoti C, Allione A, Cancel-Tassin G, Selinski S, Hengstler JG, Dietrich H, Fletcher T, Rudnai P, Gruzau E, Koppova K, Bolick SC, Godfrey A, Xu Z, Sanz-Velez JJ, D´Arcia-Prats M, Sanchez M, Valdivia G, Porrus S, Benhamou S, Hoover RN, Fraumeni JF, Silverman DT, Chanock SJ (2010) A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nat Genet 42(11): 978–984.

Satipn DP, Gunser MJ, van der Vlag J, Hamer KM, Schalk W, Almeja MJ, Saurin AJ, Freemont PS, van Driel R, Otte AP (1997) RING1 is associated with the polycomb group protein complex and acts as a transcriptional repressor. Mol Cell Biol 17(7): 4105–4113.

Senthilkumar R, Mishra RK (2009) Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. BMC Genomics 10: 549.

Simmons MN, Berglund RK, Jones JS (2011) A practical guide to prostate cancer diagnosis and management. Cleve Clin J Med 78(5): 321–331.

Swanton C (2012) Intratumor heterogeneity: evolution through space and time. Cancer Cell 22(2): 39–48.

Tabet S, Douglas SF, Daze KD, Garnett GA, Allen KJ, Abrioux EM, Quon TT, Tavares L, Dimitrova E, Oxley D, Webster J, Poot R, Demmers J, Rothman N, Garcia-Closas M, Chatterjee N, Malats N, Wu X, Figueroa JD, Chanock SJ (2010) A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nat Genet 42(11): 978–984.

Uhlen M, Oksvold P, Fagerberg L, Lundin P, Oda K, Persson M, Van den Houte M, Sjöblom T, Hallström OB, Mardin Y, van der Vliet J, Sjöblom T, Orlóci L, Madsen I, Marenco D, Ciliberti S, Fjellström P, Hallström OB, Oda K, Mardin Y, et al. (2013) Towards a knowledge-based Human Protein Atlas. Nat Rev Genet 14(2): 107–116.

Watkins DN, Herman JG, Baylin SB (2007) A stem cell-like chromatin microarray database and integrated data-mining platform. Nat Genet 39(9): 1097–1103.

Wulff JE, Hof F (2013) Synthetic trimethyllysine receptors that bind histone 3, trimethyllysine 27 (H3K27me3) and disrupt its interaction with the epigenetic reader protein CBX7. Bioorg Med Chem 21(22): 7004–7010.

Yamada K, Jones R, Koseki H, Nakayama M, Muntean AG, Maillard J, Hess JL (2011) CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. Cancer Cell 20(5): 563–575.

Zitvogel L, Rivoltini L, Aguzzi A (2005) Immunostimulatory DNA: objectives and strategies of cancer immunotherapy. Nat Rev Cancer 1: 46–54.

Zitvogel L, Rivoltini L, Apolloni M, Enninga B, Aguzzi A, Apolloni M, et al. (2003) Interferon-gamma as a key mediator of antitumor immunity. Cancer Cell 3(3): 223–233.

Zitvogel L, Rivoltini L, Apolloni M, Enninga B, Aguzzi A, Apolloni M, et al. (2003) Interferon-gamma as a key mediator of antitumor immunity. Cancer Cell 3(3): 223–233.

Zitvogel L, Rivoltini L, Apolloni M, Enninga B, Aguzzi A, Apolloni M, et al. (2003) Interferon-gamma as a key mediator of antitumor immunity. Cancer Cell 3(3): 223–233.
Chinnaiyan AM (2002) The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**(6907): 624–629.

Vincenz C, Kerppola TK (2008) Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proc Natl Acad Sci USA* **105**(43): 16572–16577.

Wang B, Tang J, Liao D, Wang G, Zhang M, Sang Y, Cao J, Wu Y, Zhang R, Li S, Ding W, Zhang G, Kang T (2013) Chromobox homolog 4 is correlated with prognosis and tumor cell growth in hepatocellular carcinoma. *Ann Surg Oncol* **20**(Suppl 3): 684–692.

Wu SC, Kallin EM, Zhang Y (2010) Role of H3K27 methylation in the regulation of lncRNA expression. *Cell Res* **20**(10): 1109–1116.

Yang L, Lin C, Liu W, Zhang J, Ohgi KA, Grinstein JD, Dorrestein PC, Rosenfeld MG (2011) ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147**(4): 773–788.

You JS, Kang JK, Seo DW, Park JH, Park JW, Lee JC, Jeon YJ, Cho EJ, Han JW (2009) Depletion of embryonic stem cell signature by histone deacetylase inhibitor in NCCIT cells: involvement of Nanog suppression. *Cancer Res* **69**(14): 5716–5725.

Yu J, Rhodes DR, Tomlins SA, Cao X, Chen G, Mehra R, Wang X, Ghosh D, Shah RB, Varambally S, Pienta KJ, Chinnaiyan AM (2007) A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Res* **67**(22): 10657–10663.

Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang CZ, Wala J, Mermel CH, Sougnez C, Gabriel SB, Hernandez B, Shen H, Laird PW, Getz G, Meyerson M, Beroukhim R (2013) Pan-cancer patterns of somatic copy number alteration. *Nat Genet* **45**(10): 1134–1140.

Zhang XW, Zhang L, Qin W, Yao XH, Zheng LZ, Liu X, Li J, Guo WJ (2010) Oncogenic role of the chromobox protein CBX7 in gastric cancer. *J Exp Clin Cancer Res* **29**: 114.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)