Supplementary Materials for

Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning

Mei Sheng Lau, Matthew G. Schwartz, Sharmistha Kundu, Andrej J. Savol, Peggy I. Wang, Sharon K. Marr, Daniel J. Grau , Patrick Schorderet, Ruslan I. Sadreyev, Clifford J. Tabin, Robert E. Kingston*

*Correspondence to: kingston@molbio.mgh.harvard.edu

This PDF file includes:

Materials and Methods

All animal procedures were performed according to NIH guidelines and approved by the Committee on Animal Care at Massachusetts General Hospital and Harvard University.

Cell culture

Cbx2$^{-/-}$ and its derivative doxycycline-inducible mouse embryonic stem cell (mESC) lines, V6.5, CJ7 and J1 mESC lines were cultured on irradiated or mitomycin-C treated mouse embryonic fibroblast cells in mESC media, containing DMEM supplemented with 15% FBS, 1000U/mL LIF, non-essential amino acids, GlutaMax, penicillin/streptomycin , 2-β-mercaptoethanol . E14 mESCs were culture on 0.2% gelatin-coated plate in mESC media. For doxycycline induction, cells were treated with appropriate amounts of doxycycline for 48 hours before collected for experiments. The concentrations of doxycycline used for each cell line were as follow: WT.1- 2.5ng/mL; WT.2- 1.5ng/mL; KRA.1- 1.5ng/mL; KRA.1 with higher dox- 2.5mg/mL; KRA.2- 2.0ng/mL; DEA- 1.5ng/mL.

Generation of doxycycline-inducible mESC lines

Cell lines were generated using Tet-On® 3G Inducible Expression System (EF1alpha Version) and Tet-On® 3G Vector Set (Bicistronic Version) (Clontech 631167, 631161) according to the manufacturer's protocol. Briefly, *Cbx2*$^{-/-}$ cells were transfected with the pEF1alpha-Tet3G plasmid and puromycin linear selection marker and selected for puromycin resistance. To test their responsiveness to doxycycline, individual clones were transfected with pTRE3G-Luc and assayed for luciferase activity. A responsive clone was chosen and transfected with the pTRE3G plasmid containing a Kozak sequence followed by the coding sequence for HA-tagged CBX2^{WT}, CBX2^{23KRA} or CBX2^{DEA} in multiple-cloning site 1 (MCS1), and the coding sequence for GFP in MCS2, along with linear hygromycin selection marker. Hygromycin-resistant clones were treated with doxycycline and screened for GFP expression by imaging and for expression of CBX2 by Western blot using anti-HA antibody (see Table S3).

Cbx2$^{-/-}$ mESCs were a gift from Haruhiko Koseki, RIKEN Center for Integrative Medical Sciences. The doxycycline-inducible mESC lines were generated with technical assistance from the MGH Gene Targeting Facility.

Western blot

Whole cells or whole embryos were lysed in RIPA buffer (50mM Tris-HCl pH 8.0, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP40, 150mM NaCl), separated on NuPAGE® 4-12% Bis-Tris gels (ThermoFisher Scientific), transferred onto nitrocellulose membranes and blotted for PRC1 components using the antibodies listed in the Table S3. TBP or H3 protein levels were used as loading controls. Relative protein levels were quantified using ImageJ.

Cross-linked Chromatin Immunoprecipitation (ChIP)

ChIP experiments were performed using the antibodies listed in Table S3 using standard protocols. Cells were cross-linked in 1% formaldehyde for 10 minutes at room temperature. 3x10^6 mESCs were used for each ChIP experiment. ChIP enrichments were assessed via quantitative PCR analysis (ChIP-qPCR) or high throughput sequencing (ChIP-seq).

For ChIP-qPCR, the sequences of the primers are listed in Table S5. Statistical significance was determined using multiple t-test without correction for multiple testing (GraphPad Prism 6, GraphPad Software).

For ChIP-seq, the enriched DNA and input DNA were made into sequencing libraries and sequenced on Illumina HiSeq 2000, resulting in approximately 30 million 50bp reads per sample. Reads were aligned against the mm9 reference genome using BWA. Alignments were filtered for uniquely mapped reads and alignment duplicates were removed. Input-normalized coverage tracks were generated using SPP (*18*). To determine regions of ChIP-seq tag enrichment, tag counts in 1kb-sized windows that overlap by 200bp were analyzed across entire chromosomes. Statistical significance of enrichment of ChIP vs. input was estimated using negative binomial distribution, with the estimate of the mean based on the tag counts in input, and the size parameter(s) selected based on manual inspection of resulting peak calls. Adjacent enriched windows separated by 1kb or less were merged to generate a list of regions with significant enrichment.

To compare the ChIP-seq signal intensities between $CBX2^{WT}$ and $CBX2^{23KRA}$ (fig. S4B), we first identified peaks from the $CBX2^{WT}$ dataset that overlapped at least 1bp with a region that is ± 3kb of a transcription start site. We obtained 502 peaks with peak widths that ranged from 1kb to 52.2kb with a median of 4.8kb. The input –normalized signal densities at these regions from the $CBX2^{WT}$ and $CBX2^{23KRA}$ ChIP-seq datasets were calculated and plotted. The best-fit line through the density distribution plot was computed using a first-order call to the MATLAB polyfit function. The correlation coefficient is Pearson's *r*. See Table S1 for the list of genes with $CBX2^{WT}$ or $CBX2^{23KRA}$ binding and their signal intensities.

Native ChIP (NChIP)

NChIP were performed as described (*19*) with the following modifications: Two whole E10.5 embryos were used for each ChIP experiment. Embryos were lysed in dounce buffer (10mM Tris-HCl pH 7.5, 4mM MgCl2, 1mM CaCl2, supplemented with 1 × EDTA free- protease inhibitor cocktail). The lysate was treated with 2U/µl MNase (NEB) for chromatin digestion for 7.5 minutes at 37°C , and quenched with 10mM EDTA. To release the chromatin from nuclei, Triton –X and sodium deoxycholate were added to the sample to 1% final concentration and incubated on ice for 20 minutes. The sample was diluted 4-fold with NChIP immunoprecipitation buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 15 mM NaCl, 0.1% Triton X-100, supplemented with 1 × EDTA-free protease inhibitor cocktail) and incubated at 4 °C with rotation for 1 hour. The debris was pelleted by centrifugation and soluble chromatin was recovered in the supernatant. The chromatin was pre-cleared using 20µl of protein A Dynabeads (Life Technologies) and immunoprecipitated with 4µg-20µl antibody–bead complex overnight at 4 °C . Immunoprecipitated complexes were washed four times using low salt wash buffer (20mM Tris-HCl pH 8.0, 2mM EDTA, 150mM NaCl, 1% Triton-X, 0.1% SDS,

supplemented with 1 × EDTA-free protease inhibitor cocktail). Protein–DNA complexes were eluted in ChIP elution buffer (100 mM NaHCO3 and 1% SDS) for 1-1.5 hours at 65 °C. The eluted material was treated with RNaseA and Proteinase K before DNA purification by phenol-chloroform and ethanol-precipitation. The DNA pellet was resuspended in nuclease –free water for ChIP-qPCR analyses.

Co-immunoprecipitation

To prepare nuclear lysate, 8x10^6 mESCs were suspended in 800 µl RSB buffer (10mM Tris-HCl pH7.5, 10mM NaCl, 3mM $MgCl_2$), and incubated on ice for 15 min. Cells were collected by centrifugation for 3 min at 300g, resuspended in 1mL RSB buffer, and lysed using a tight homogenizer. The homogenate was centrifuged for 2 min at 9,000g to pellet nuclei. The nuclei were washed three times with RSB buffer, and resuspended in 225 µl of buffer C (20mM HEPES pH7.9, 25% glycerol, 0.42M NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA, 0.5mM DTT, 1x EDTA-free protease inhibitor cocktail). The suspension was incubated with rotation for 30 min at 4 °C, and then centrifuged at 9,000g for 30 min at 4 °C . The supernatant was collected as the nuclear extract.

Nuclear extract was mixed with an equal volume of buffer C' (20mM HEPES pH7.9, 0.2% NP40, 0.42M NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA, 0.5mM DTT, 1x EDTA-free protease inhibitor cocktail) and incubated with rotation with 4 µg of antibody (Table S3) at 4 °C for 1 hour. 50 µl of protein A Dynabeads was added and incubated with rotation at 4 °C for 1 hour. The antigen-antibody complex was washed three times with washing buffer (20mM HEPES pH7.9, 12.5% glycerol, 0.15M NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA, 0.5mM DTT, 0.1% NP-40, 1x EDTA-free protease inhibitor cocktail). The proteins were eluted by heating at 70°C in LDS for 10 minutes.

Gene expression analyses

RNA-seq: Total RNA was extracted from mESCs with Trizol, DNaseI-treated and purified using Direct-zol™ RNA MiniPrep (Zymo). mRNA was purified from the RNA and made into cDNA libraries using TruSeq total RNA kit (Illumina). Sequencing libraries were prepared and sequenced as the ChIP-seq libraries above. The reads were aligned to the mouse transcriptome (mm9 RefSeq annotation) using TopHat (*20*). Gene expression counts were calculated using HTSeq v.0.6.0 (*21*). Calculation of expression values and differential expression analyses were performed using edgeR (*22*), with a filter for genes with average read counts of $\log_2 CPM \geq 1$ from all biological replicates. The differential expression is a comparison of the +dox (with induction of CBX2 expression) and –dox (no CBX2 expression) states for each mESC line. Two or three biological replicates were used for each mESC line. Genes with fold change (FC) of expression value equal or larger than 1.3 and $p < 0.10$ were defined as differentially expressed. See Table S2 for the list of all differentially expressed genes in each mESC line. Read-normalized coverage tracks were generated using SPP.

To identify candidate CBX2 target genes for gene expression analysis, we took the top 285 genes from each of the $CBX2^{WT}$ and $CBX2^{23KRA}$ ChIP-seq datasets, identified the genes that were found in both lists and designated these as candidate targets of CBX2. The differential expression values of these genes ($\log_2$(FC[+dox/-dox])) were used to generate the clustered heatmap (Fig. 1C) using Euclidean dissimilarity measure and complete linkage clustering algorithm implemented in the R function heatmap.2 with

default arguments (distfun = dist, hclustfun=hclust). Note that the final gene list is relatively short as many of the genes were not expressed (log_2CPM<1) in mESCs.

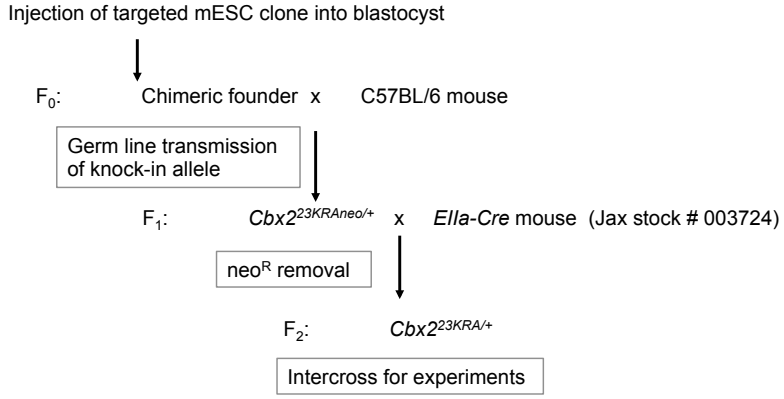
To quantify the enrichment of PRC1 targets among differentially expressed genes, we performed one-tailed Fisher's exact test on the number of PRC1 targets among differentially expressed genes compared to the number of PRC1 targets among all genes in the reference Ensembl genome annotation. PRC1 targets were identified as described in *(23)*.

Reverse-transcription quantitative PCR (RT-qPCR): Total RNA was prepared as above and reverse transcribed using SuperScript® VILO™ cDNA Synthesis Kit (Life Technologies). qPCR was performed using the Taqman probes indicated in Table S4. B2M levels were used as the internal control for cDNA content. The expression level of a gene in each +dox sample is presented as a relative value to its expression level in the corresponding no dox sample. Statistical significance was determined using multiple t-test without correction for multiple testing (GraphPad Prism 6, GraphPad Software).

Generation of *Cbx2*^{23KRA} mutant mice

The *Cbx2*^{23KRA} targeting construct was assembled from the following components into a pUC19 vector backbone: gBlock encoding the K-to-A and R-to-A mutations (Integrated DNA Technologies, IDT); homologous arms that were PCR-amplified from a BAC clone (RP23-272P15); thymidine kinase selection cassette that was PCR-amplified from the plasmid PL253 (ATCC, PTA-4998); loxP-flanked neomycin selection cassette that was PCR-amplified from a standard plasmid. We verified the construct by sequencing.

The steps to obtain *Cbx2*^{23KRA} mutant mice are summarized in the schematic below:



The Genome Modification Facility at Harvard University performed the electroporation of the *Cbx2*^{23KRA} targeting construct into V6.5 (C57BL/6 X 129S4/SvJae) mESCs, and the injection of targeted mESC clone into host blastocysts (C57BL/6 x BDF1).

The genotyping PCRs were described in fig. S7.

The *Cbx2*$^{23KRA/+}$ F2 mice were intercrossed to generate animals for experimentation. F3 progeny were genotyped using a melt curve analysis approach (see below and fig. S8C), and also screened for the presence of neo^R. Only mice with complete neo^R removal were used for further breeding and experimentation.

Production of sgRNA and donor DNA for CRISPR-mediated gene targeting

sgRNAs (sgRNA-L: GCCTTCTGTTCTGGGGGTAGAGG; sgRNA-R:GAGACCCGTCAGCCTGGCCAAGG) were designed using Zinc Finger Targeter (ZiFiT) (*24*), synthesized as DNA oligonucleotides (IDT), annealed, phosphorylated and cloned into pX335 (Addgene plasmid # 42335) by the Genome Engineering Production Group, Department of Genetics, Harvard Medical School. The T7 promoter was added to the sgRNA template by PCR-amplification using primers CLO94/95 (Table S5). The T7-sgRNA PCR product was gel purified and used as the template for in vitro transcription using MEGAshortscript T7 kit (Life Technologies). sgRNAs were purified using MEGAclear kit (Life Technologies) and eluted in RNase-free water.

Donor DNA was assembled from the following components into a pUC19 vector backbone: gBlock encoding the K-to-A and R-to-A mutations (IDT) and homologous arms that were PCR-amplified from a BAC clone (RP23-272P15) using primer pairs CLO84/85 and CLO86/87 (see Table S5 for primer sequences). The assembled donor was purified using UltraClean Endotoxin Removal (MO Bio) and eluted in nuclease-free water.

Cas9 nickase mRNA was purchased from SystemBiosciences (CAS504A-1).

Generation of *Cbx2*^{13KRA} mutant mice

The single-guide RNA (sgRNAs) pair, Cas9 nickase mRNA, and plasmid donor for homologous recombination were injected into mouse zygotes (C57BL/6 inbred strain) by the Genome Modification Facility at Harvard University.

Mice born from the injected zygotes were screened for homologous recombination events via PCR-RFLP (restriction fragment length polymorphism) (fig. S8B). Based on these results, three mice were identified to have homologous recombination events in their genomes and designated as founders. Exon 5 of the *Cbx2* gene from these animals were cloned and sequenced so that they can be characterized in detailed, e.g. to identify small internal insertions or deletions that could not be detected by PCR-RFLP.

The founders were crossed to C57BL/6 mice to test for germ line transmission. *Cbx2*$^{13KRA/+}$ progeny were intercrossed to produce animals for experimentation. Animals from subsequent generations were genotyped using the melt curve analysis approach.

Genotyping by PCR

Genomic DNA was isolated from mouse toe or tail tip tissues using NucleoSpin Tissue (Machery-Nagel), and genotyped by PCR amplification using primers indicated in fig. S7A, C, and fig. S8A, followed by gel electrophoresis analyses of the products. Primer sequences are listed in Table S5. For RFLP analysis, the PCR products were purified, treated with HindIII and XmaI for at least 3 hours at 37 °C, heat inactivated, and analyzed by gel electrophoresis.

Genotyping by melt curve analysis

Exon 5 of *Cbx2* is amplified by real-time PCR using the primer pairs gDet23F/gDet24 (fig. S8C) and iTaq Universal SYBR Green Supermix (BioRad #1725120), with the cycling parameters: 95 °C for 10min, followed by 40 cycles of 95 °C for 15s and 60 °C for 1min. The melt curve profiles were generated by increasing the temperature from 75.0 to 95.0 °C in 0.1 °C increments, and analyzed using Bio-

Rad's CFX Manager™ software. The PCR products from wild-type and *Cbx2* mutant alleles give different profiles due to differences in the nucleotide sequence. We verified the results from this method by cloning and sequencing the PCR products.

Protein Purification for Nucleosome Remodeling and Inhibition Assays

Flag-tagged CBX2 variants and untagged Ring1b were expressed and purified as previously described *(10)*. Briefly, the coding sequences were inserted into pFastBac1. Baculovirus was generated according to the Bac-to-Bac® system procedures (Thermofisher). The proteins were co-expressed by co-infecting Sf9 cells. Sf9 nuclear extracts were generated using the Dignam protocol, flash-frozen, and stored at -80°C until needed. To purify proteins, nuclear extracts were thawed in cold water and then incubated with M2-agarose (Sigma Aldrich A2220) at 4 °C with gentle rocking for 4 hours. The M2-agarose was collected by centrifuging at 250 g for 5 min, then washed in batch, first with BC buffer (20mM HEPES 7.5, 0.2mM EDTA, 20% glycerol, 1mM DTT) containing 500mM KCl, then 300mM KCl. Proteins were eluted in BC buffer plus 300mM KCl containing 0.4 mg/ml flag peptide and quantified using standard techniques.

Nucleosome Remodeling and Inhibition Assays

A plasmid containing 10 5S nucleosome positioning sequences and a unique centrally located HhaI restriction enzyme site (pG5E4) has been previously described (*25*). The 2.5 kb Asp718/ClaI G5E4 fragment was digested, purified, and end labeled with Cy5. Chromatin arrays were assembled using HeLa core octamers and a salt-gradient assembly protocol, and were typically 80-85% assembled, as determined by levels of background cutting with HhaI (*26*). Reactions were done essentially as described (*10*). PcG proteins were incubated with 0.5-3nM nucleosomes for 30 min at 30°C, and then 100ng of purified human SWI/SNF and 8 units of HhaI were added to a volume of 20µl. The reactions were allowed to proceed for 1 hour at 30 °C , and then 5µl of stop buffer was added and reactions incubated an additional 30 min at 55 °C . Final reaction buffer conditions were as follows: 16 mM HEPES 7.5, 80mM KCl, 2mM $MgCl_2$, 1mM $MnCl_2$, 2mM ATP, 2.5 µg BSA, 8% glycerol, and 2 mM DTT. Stop buffer consists of 10mM Tris 7.5, 70mM EDTA, 1% SDS, 0.1% Orange G, and 1.5 mg/ml Proteinase K. After completion, reactions were separated on 1% agarose TAE gels and scanned and quantified using a Typhoon phosphoimager and ImageQuant TL software (GE Healthcare). Graphs were generated using KaleidaGraph graphing software (Synergy).

Skeleton analyses

Skeletal whole mounts of P15-17 mice were prepared as previously described (*15*). Carcasses were skinned, eviscerated, incubated in 96% ethanol for 5 days, and transferred to acetone for 2 days. Staining was performed in 0.005% alizarin red S, 0.015% alcian blue 8GS in 5% acetic acid, 5% water, and 90% ethanol for 3 days at 37°C or 5 days at room temperature. Samples were rinsed with water and cleared by sequential incubation in 1% KOH, 0.8% KOH and 20% glycerol, 0.5% KOH and 50% glycerol, and 0.2% KOH and 80% glycerol. Cleared skeletons were stored in 100% glycerol.

Author Notes

M.S.L and R.E.K. designed the project. M.G.S. performed skeletal analyses; S.K performed ChIP for qPCR and ChIP-seq from wild-type mESCs; S.K.M performed mouse genotyping; D.J.G. performed chromatin remodeling assays; M.S.L. conducted all other experiments. P.S. advised on mouse work. A.J.S, P.I.W and R.I.S. performed bioinformatics analyses. M.S.L., M.G.S., P.S, C.J.T and R.E.K. wrote the manuscript.

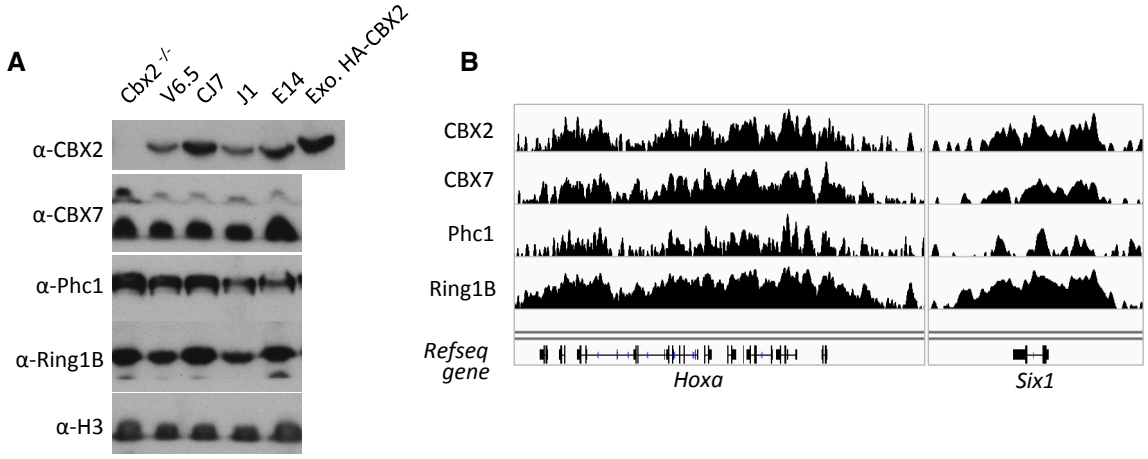


Fig. S1. CBX2 is expressed and binds to PRC1 target genes in mESCs.

(A) Western blot for CBX2 in four common mESC lines. Lysate from *Cbx2*$^{-/-}$ mESCs and exogenous HA-tagged CBX2 were used as negative and positive controls, respectively, for the specificity of the CBX2 antibody. The presence of other PRC1 components, CBX7, Phc1 and Ring1B, are also shown for these cells. H3 is used as loading control.

(B) ChIP-seq tracks for CBX2 and other PRC1 subunits at two known PRC1 target regions.

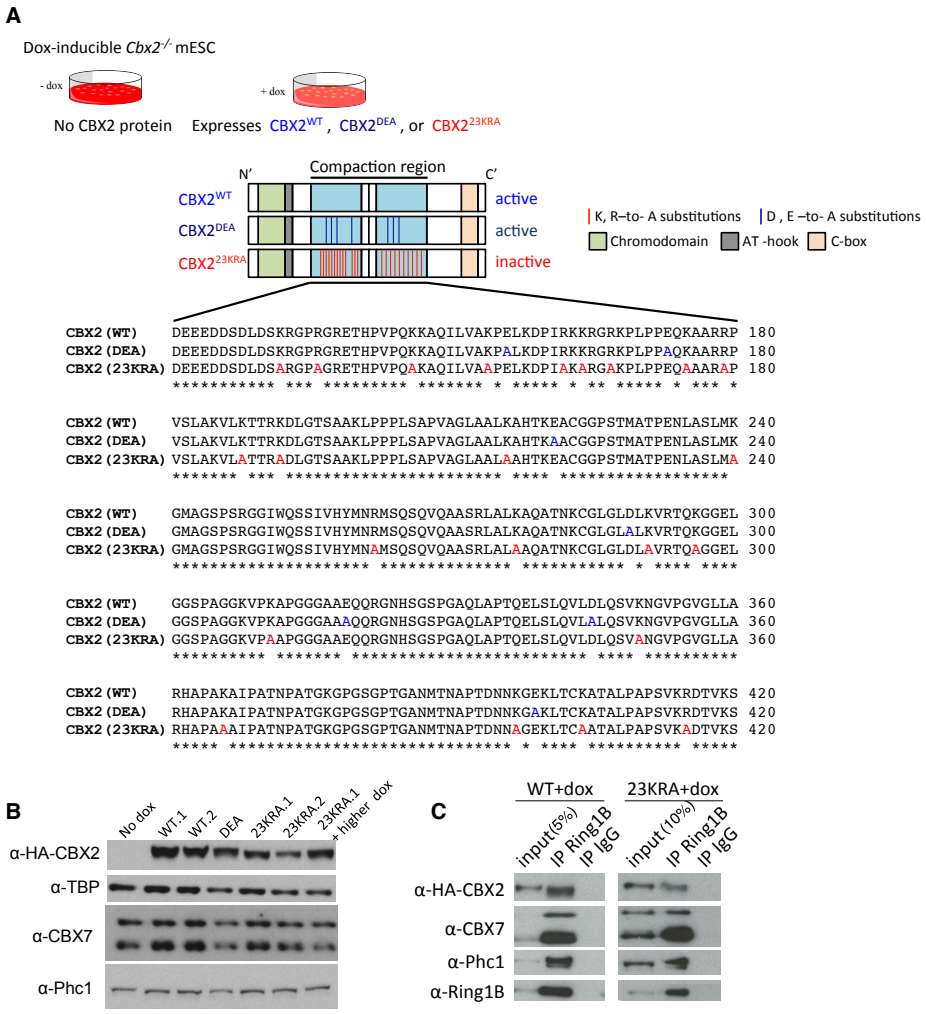


Fig. S2. mESC lines expressing wild-type or variant CBX2.

(A) Schematic illustration of the doxycycline-inducible system used to express CBX2^{WT}, CBX2^{DEA} and CBX2^{23KRA} in *Cbx2*$^{-/-}$ mESCs, and the amino acid sequence of the compaction regions of the CBX2 variants. Amino acid substitutions in CBX2^{DEA} and CBX2^{23KRA} are highlighted in blue and red, respectively.

(B) Western blots of CBX7 and Phc1 in mESC lines expressing wild-type or variant CBX2. The blots for CBX2 and TBP are the same as those shown in Fig. 1B, and are included here for reference.

(C) Western blots of CBX2, CBX7 and Phc1 following immunoprecipitation of Ring1B from CBX2^{WT}- and CBX2^{23KRA}-expressing mESCs.

A

WT.1	Number of genes	
	D.E	Not D.E.
Non-PRC1 targets	101	33, 831
PRC1 targets	23	640
P-value	1.10e-16	

23KRA.1	Number of genes	
	D.E	Not D.E.
Non-PRC1 targets	23	33,909
PRC1 targets	2	661
P-value	0.082	

WT.2	Number of genes	
	D.E	Not D.E.
Non-PRC1 targets	54	33,878
PRC1 targets	27	636
P-value	7.74e-27	

23KRA.2	Number of genes	
	D.E	Not D.E.
Non-PRC1 targets	37	33,895
PRC1 targets	1	662
P-value	0.520	

DEA	Number of genes	
	D.E.	Not D.E.
Non-PRC1 targets	77	33,855
PRC1 targets	23	640
P-value	6.61e-19	

D.E. = Differentially expressed

B

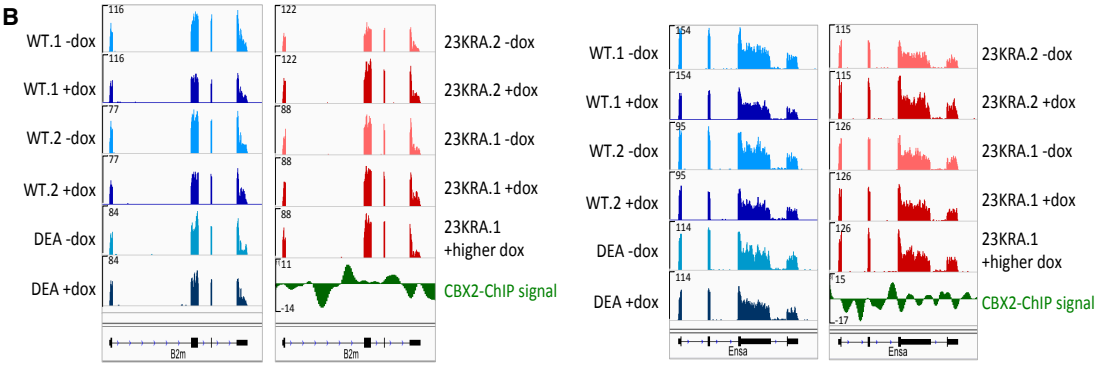


Fig. S3. Expression of CBX2^{WT} and CBX2^{DEA}, but not CBX2^{23KRA}, leads to selective repression of PRC1 target genes.

(A) Contingency tables and results of one-tailed Fisher's exact tests for the enrichment of PRC1 targets among differentially expressed genes in each cell line.

(B) RNA-seq tracks of non-PRC1 targets before and after induction of CBX2 expression using doxycycline (-dox and +dox, respectively). y-axis indicates normalized read counts.

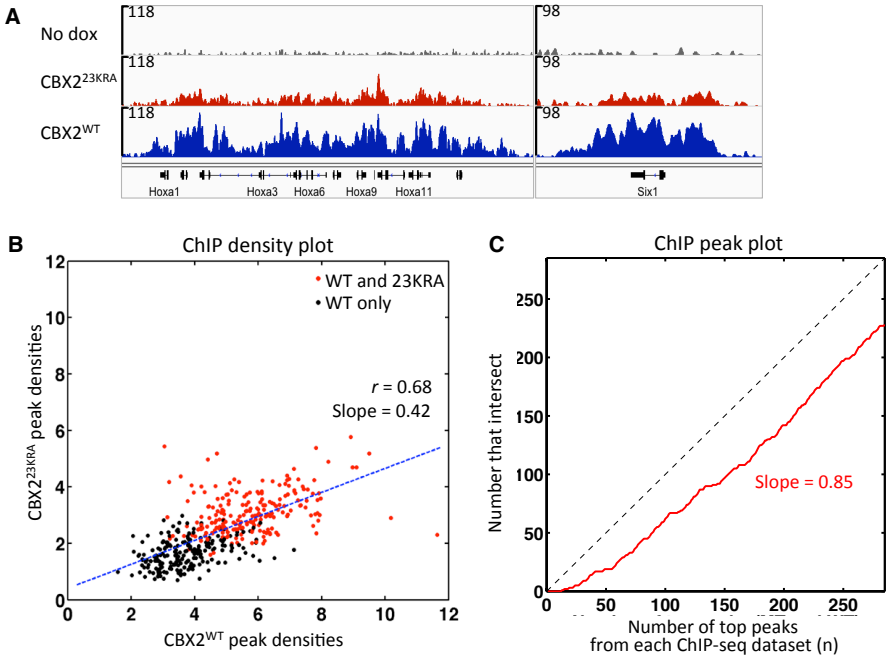


Fig. S4. CBX2 binding in CBX2^{23KRA}- and CBX2^{WT}-expressing mESCs.

(A) CBX2 ChIP-seq tracks in CBX2^{23KRA}- and CBX2^{WT} –expressing mESCs at known PRC1 target genes. y-axis indicates input-normalized read counts.

(B) Plot of CBX2 ChIP-seq signal densities, as determined from ChIP-seq analyses in CBX2^{23KRA}- and CBX2^{WT} –expressing mESCs, at every genomic region where a peak is called in the CBX2^{WT} ChIP-seq dataset. Regions where a peak is also called in the CBX2^{23KRA} ChIP-seq dataset are highlighted in red. r = correlation efficiency between the CBX2^{WT} and CBX2^{23KRA} density values. The CBX2^{23KRA} peak intensities are on average 42% of the CBX2^{WT} peak intensities.

(C) Plot of the number of genes that intersect when comparing the top n genes from the CBX2^{WT} and CBX2^{23KRA} ChIP-seq datasets. The average overlap is 85%.

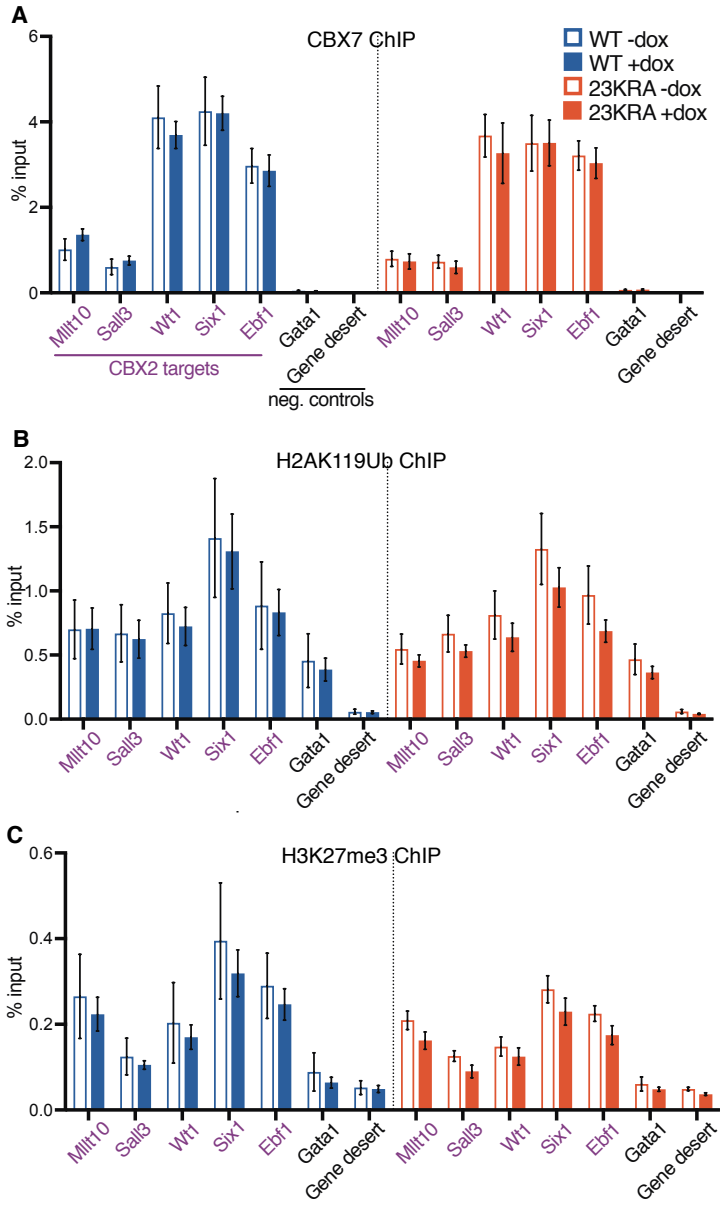


Fig. S5. Occupancy of PRC1 and its associated histone modifications on chromatin are similar in CBX2^{WT}- and CBX2^{23KRA}-expressing mESCs.

ChIP-qPCR analyses for (A) CBX7, (B) H2AK119Ub, and (C) H3K27me3 levels. Mean ± SD, n≥3 biological replicates. No statistically significant differences were detected between –dox and +dox pairs, or between WT+dox and 23KRA+dox samples in any of the three analyses.

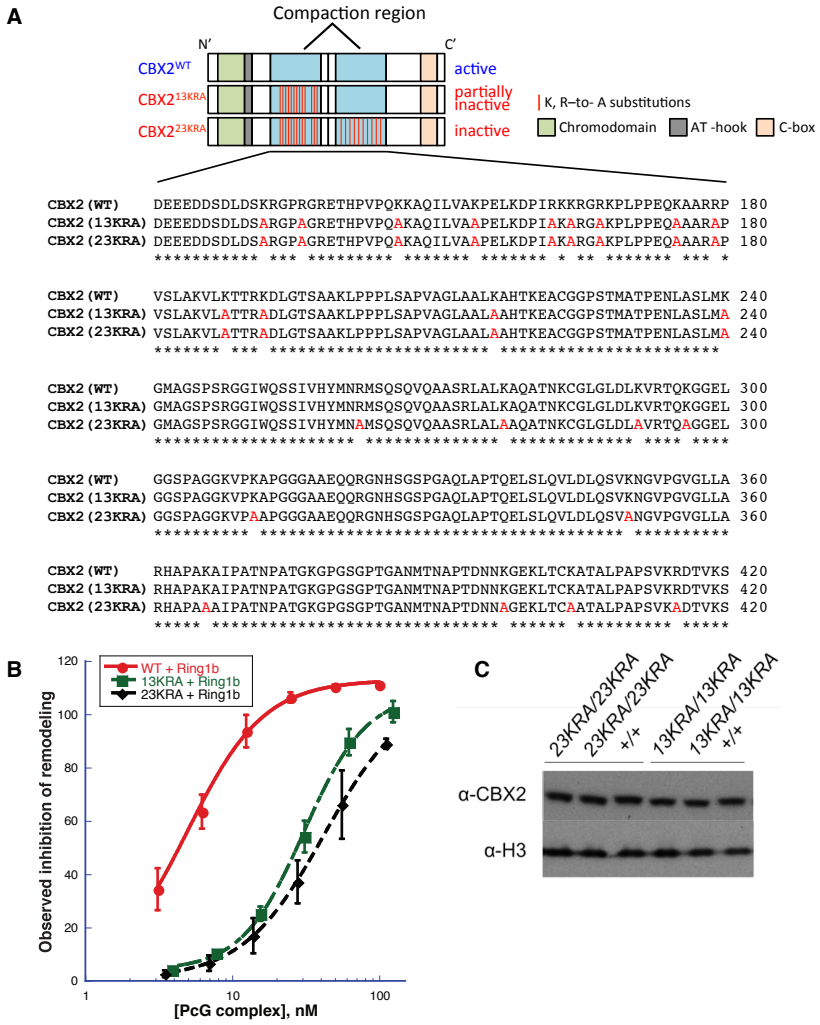
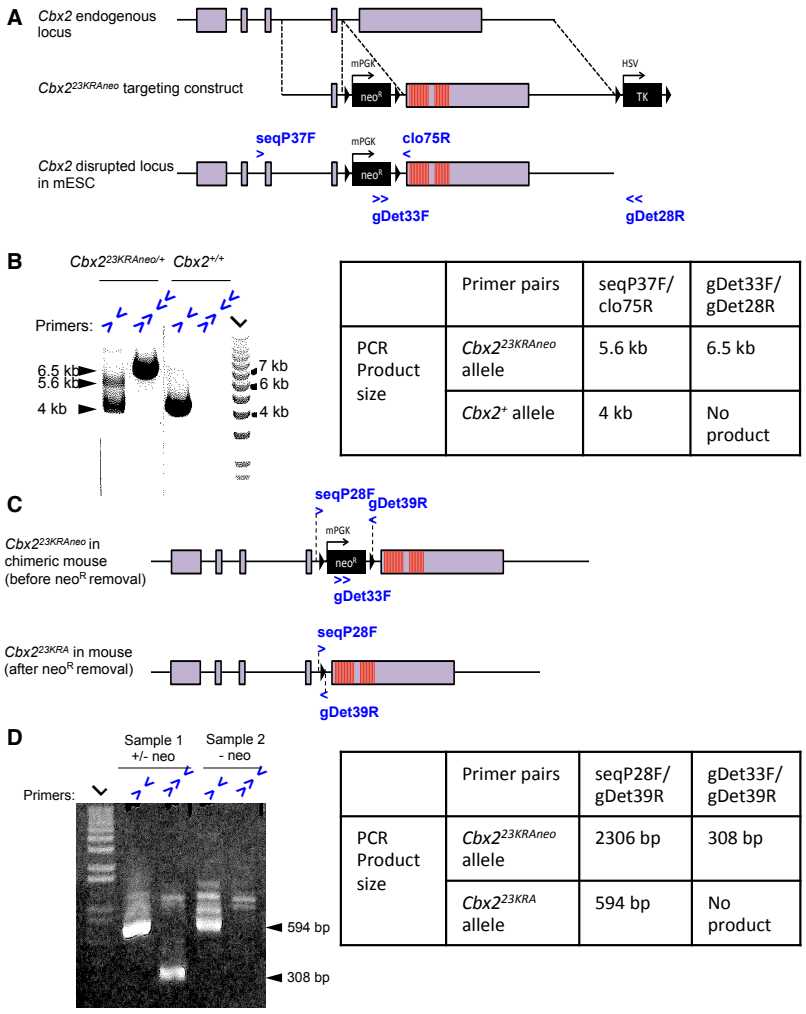


Fig. S6. *Cbx2*^{13KRA} and *Cbx2*^{23KRA} alleles in mutant mice.

(A) Schematic illustration of the gene products of the *Cbx2* mutant alleles in mice, and their amino acid sequences in the compaction region. Amino acid substitutions are highlighted in red.

(B) Inhibition of chromatin remodeling by CBX2 variants in complex with Ring1B. Mean ± SD, n=3 experimental replicates. The ability of CBX2 to inhibit chromatin remodeling by SWI/SNF on a reconstituted chromatin template correlates well with its chromatin compaction activity (*10*).

(C) Western blot of CBX2 in lysates from E10.5 embryos. H3 levels are used as loading controls. Samples with matched genetic backgrounds are marked with horizontal lines.



	Primer pairs	seqP37F/ clo75R	gDet33F/ gDet28R
PCR Product size	*Cbx2*^{23KRAneo} allele	5.6 kb	6.5 kb
	Cbx2^{+} allele	4 kb	No product

	Primer pairs	seqP28F/ gDet39R	gDet33F/ gDet39R
PCR Product size	*Cbx2*^{23KRAneo} allele	2306 bp	308 bp
	Cbx2^{23KRA} allele	594 bp	No product

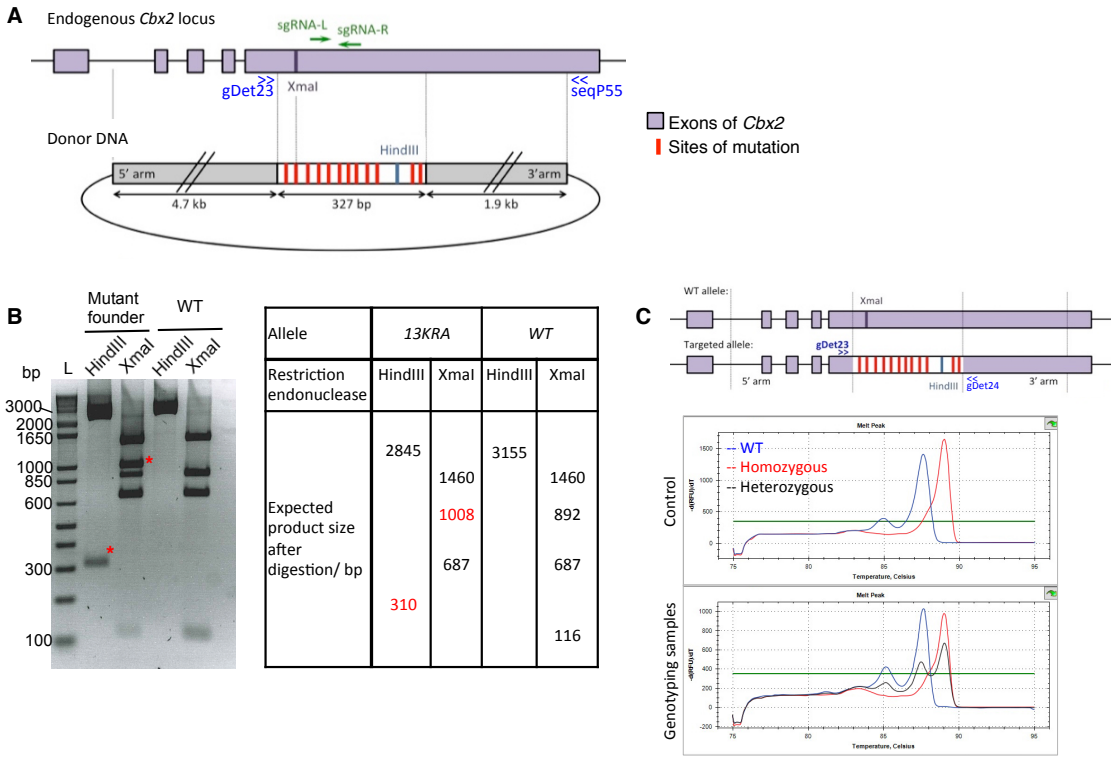
Fig. S7. Generation of *Cbx2*^{23KRA} mutant mice.

(A) Strategy for knocking-in *Cbx2*^{23KRA} in mESCs. The endogenous locus, targeting construct, and targeted locus after homologous recombination are shown. Blue arrowheads indicate the binding sites of the primers used to screen for successful targeting events.

(B) DNA gel image of genotyping PCR results. The table presents the expected sizes of the PCR products from the primers shown in (A).

(C) The *Cbx2*^{23KRA} allele before and after neomycin cassette (neo^{R}) removal. Blue arrowheads indicate the binding locations of the primers used to identify successful CRE-*lox* excision events.

(D) DNA gel image of genotyping PCR results. The table presents the expected sizes of the PCR products from the primers shown in (C). For sample 1 on the gel image, the presence of a 594bp and a 308bp band indicates that CRE-mediated neo^{R} removal is incomplete. For sample 2, only the 594bp band is present, indicating neo^{R} removal is complete. Only *Cbx2*$^{23KRA/+}$ mice with complete neo^{R} removal were used for further breeding and experimentation.



Allele	*13KRA*		*WT*	
Restriction endonuclease	HindIII	XmaI	HindIII	XmaI
Expected product size after digestion/ bp	2845	1460	3155	1460
		1008		892
		687		687
	310			116

Fig. S8. Generation of *Cbx2*^{13KRA} mutant mice.

(A) Schematics of the endogenous *Cbx2* locus and the DNA donor used to knock-in the *Cbx2*^{13KRA} mutations via CRISPR-mediated targeting in zygotes. The recognition sites of the sgRNA pair, the sites encoding the mutations, and the primer and restriction sites for RFLP (restriction fragment length polymorphism) analysis are indicated.
(B) DNA gel image of genotyping PCR-RFLP results. Fragments unique to the *Cbx2*^{13KRA} and wild-type (WT) alleles are marked with red asterisks and indicated in red in the table. The mutant founder has both *Cbx2*^{13KRA} and wild-type alleles.
(C) Genotyping melt curve analysis. The binding sites of the primers used in this analysis are shown in the schematics. Melt curves from control plasmids, wild-type, *Cbx2*$^{13KRA/+}$ and *Cbx2*$^{13KRA/13KRA}$ animals are shown. The profiles for *Cbx2*^{23KRA} alleles are similar to the *Cbx2*^{13KRA} profiles.

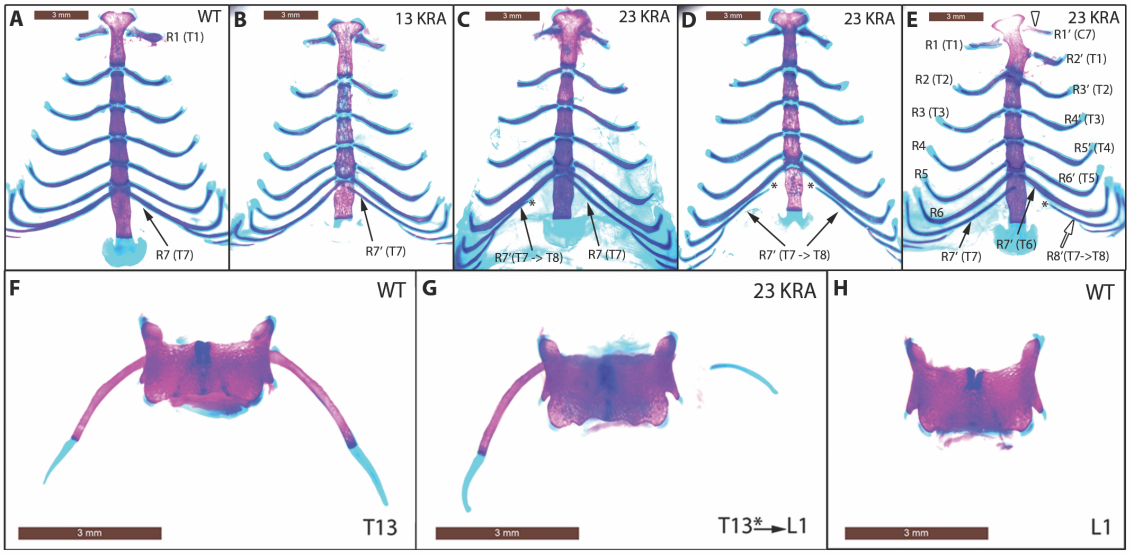


Fig. S9. T7-to-T8 and T13-to-L1 transformations in *Cbx2*^{23KRA} and *Cbx2*^{13KRA} homozygous mutants.

(A-E) Ventral views of the ribs and sternum showing T7-to-T8 transformations that are characterized by changes in the vertebrosternal ribs in *Cbx2*^{23KRA} (C to E) and *Cbx2*^{13KRA} (B) homozygous mutants as compared to wild-type (A). Black arrows mark normal (R7) or transformed (R7') T7 ribs. Partial T7-to-T8 transformations are characterized by R7' ribs that articulate to the xiphisternum (right side of B and left side of E). Complete T7-to-T8 transformations are characterized by the conversion of a true rib to a false rib and are marked with an asterisk (left side of C and both sides of D). The *Cbx2*^{23KRA} mutant in (E) has extensive anterior-to-posterior transformations on the right side, characterized by a complete misalignment of the ribs on the left and right sides. A white arrowhead marks the presence of an ectopic C7 rib (R1') and an asterisk marks a complete T7-to-T8 transformation at R8'. R8' is indicated with a white arrow and is labeled R8' (T7->T8) due to the shift of all the ribs on the right side in (E) as a consequences of the ectopic C7 rib.

(F-H) Dorsal views of disarticulated T13 and L1 vertebrae. T13-to-L1 transformation is characterized by degeneration of floating fib on the right (G).

Scale bar = 3mm.

A

Outcome	Expected #	Observed #		
+/+	42	38		
23KRA/+	84	90		
23KRA/23KRA	42	40	P value:	0.64
TOTAL	168	168		

Outcome	Expected #	Observed #		
+/+	54.75	59		
13KRA/+	109.5	109		
13KRA/13KRA	54.75	51		
TOTAL	219	219	P value:	0.74

B

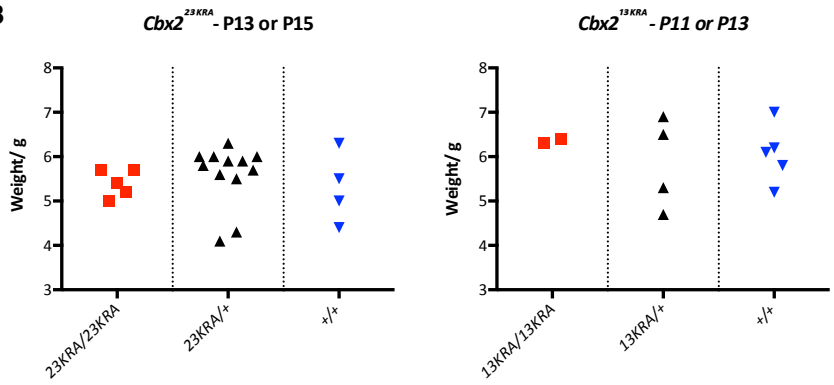


Fig. S10. *Cbx2*^{23KRA} and *Cbx2*^{13KRA} mutations do not result in reduced weight or lethality.

(A) Chi-square test on the number of progeny from heterozygous intercrosses of the mutants.

(B) Weight of the homozygous mutants in comparison to their heterozygous and wild-type littermates at the indicated age.

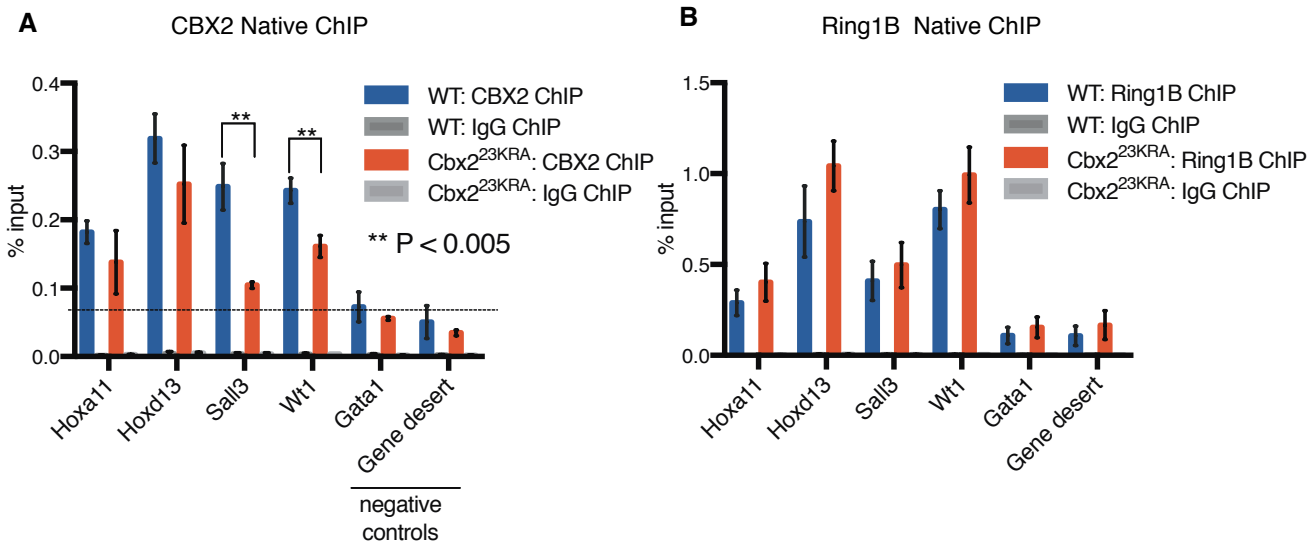


Fig. S11. PRC1 occupancy on chromatin is similar in wild-type and *Cbx2*^{23KRA} homozygous mice.

Native ChIP-qPCR analyses for CBX2 (A) and Ring1B (B) in wild-type and *Cbx2*^{23KRA} homozygous mutant embryos. Mean ±SD, n=3 biological replicates. ** P< 0.005, t-test. In (A), the % input for CBX2^{23KRA} is on average 66% that for CBX2^{WT}, compared to the average of 42% measured in ChIP with cross-linking (fig. S4B). The lower signal for CBX2^{23KRA} compared to CBX2^{WT} in native ChIP is indicative of CBX2^{23KRA} having lower binding affinity, which is consistent with its compromised ability to compact chromatin. Compaction occurs through CBX2 forming contacts with multiple nucleosomes; therefore, disrupting its compaction ability is likely to reduce its residence time on chromatin.

Table S3. Antibody list.

Details of the antibodies and respective amounts used for various applications.

Antibodies	Company	Catalog Number	ChIP	IP	Western-blot
CBX2	Bethyl Laboratories Inc	A302-524A	4ug		
CBX2 (C-18)	Santa-Cruz	sc-19297			1/500
CBX7	Abcam Inc	ab21873	4ug		
CBX7	EMD Millipore	07-981			1/2000
CBX7 (G-3)	Santa-Cruz	sc-376274			1/500
CBX8	Bethyl Laboratories Inc	A300-882A			1/1000
H2AK119Ub	Cell Signaling Technology	8240S	4ug		
H3	Abcam Inc	ab1791			1/10,000
H3K27me3	EMD Millipore	07-449	4ug		
HA	Covance Research Inc	MMS-101P			1/2000
HA	Abcam Inc	ab9110			1/2000
Phc1	Active Motif	39723			1/2000
Rabbit IgG	Abcam Inc	ab37415		4ug	
Ring1B	Bethyl Laboratories Inc	A302-869A	4ug	4ug	
Ring1B	Active Motif	39663			1/2000
TBP	Abcam Inc	ab818			1/2000

Table S4. Taqman probe list.

Gene	Assay ID
Sall3	Mm01265835_m1
Wt1	Mm01337048_m1
Six1	Mm00808212_m1
Ebf1	Mm01288946_m1
Mllt10	Mm00487708_m1
B2m	Mm00437762_m1
Hprt	Mm00446968_m1
Pou5f1	Mm00658129_gH
Nanog	Mm02384862_g1
Actb	Mm00607939_s1

Table S5. Primer list.

ChIP primers	
Mllt10_F	CAATGCCTCGTCACAGCTAATA
Mllt10_R	GCCTCCTGTGAAAGTGAAAGA
Sall3_F	CAGAGATGGAACTGGTGCTTTA
Sall3_R	GCCCAGAGTACAGTGTGTTAG
Wt1_F	GGAAGTGTTGGGAATCTCTCTT
Wt1_R	CTCCAGGGATTTGGTTCGTATC
Six1_F	CACTGTCCCGAGAAACATAGAG
Six1_R	GTATGAGGAGAGGATAGGGATAGG
Ebf1_F	CTCCTGACTTCTGTGTGGTTT
Ebf1_R	CTGAGTCCTGGTTACACATAGC
Gata6-F	GCTCCTTTCCCAGAGCGTTGAAT
Gata6-R	CCCTCCTTCCAAATTAAGCCC
Hoxa11-F	GGAAGTAAGCAGAAAGATACAGGGAAGG
Hoxa11-R	TTTGTCAATAATCCGCGCTGTCCG
Hoxd13-F	TGGGCTATGGCTACCACTTC
Hoxd13-R	GACACTTCCTTGGCTCTTGC
Sox17-F	TTACTTGTGGCATTTGTGGCTGGC
Sox17-R	CAGCAGTGTGAGTGGGCCATATTT
Gene desert_F	CCACTCTTCTTATAGGACCCTTTG
Gene desert_R	CCTGTCTACCTGTTCTTTACATTCT
Gata1_F	GATCACCCTGAACTCGTCATAC
Gata1_R	TTTGGGAATCAAGACTGACCTG

Cloning and genotyping primers	
CLO75	GGAGTCTAGGTCGCTGTCGT
CLO84	GTGAATTCGAGCTCGGTACCCGCTCTGCGGGGCTAACCGC
CLO85	AGAGGACGAACTGCTGGATTTGGA
CLO86	CATGAGCCAGAGTCAGGTTCAGGC
CLO87	TTGCATGCCTGCAGGTCGACTCAGCACTAGACTTATCTCC
CLO94	TTAATACGACTCACTATAGG
CLO95	AAAAGCACCGACTCGGTGCC
gDet23	GAACCAGATGCGCCATCCAAATC
gDet24	TCATGTAGTGTACGATGGAGC
gDet28	GGTGGTCAGACTAGGAACACAGAAAT
gDet33	CCGATTCGCAGCGCATCGCCTTCTATCGCC
gDet39	GAATCAACGCCATAACTTCG
seqP28	CCCCTATTGGAGCTAGAGCA
seqP37	AGTTTGCCTGTGGGTTTGAG
seqP55	AAAGGCACCAAATTGGGATT

References:

18. P. V. Kharchenko, M. Y. Tolstorukov, P. J. Park, Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology*. **26**, 1351–1359 (2008).
19. J. Brind'Amour *et al.*, An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Comms*. **6**, 6033 (2015).
20. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. **25**, 1105–1111 (2009).
21. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data (2014).
22. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2009).
23. S. Kundu *et al.,* Polycomb Repressive Complex 1 generates discrete compacted domains that change during differentiation. *Mol. Cell*. **65**, p432–446.e5 (2017).
24. J. D. Sander *et al.*, ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Research*. **38**, W462–8 (2010).
25. D. J. Steger, A. Eberharter, S. John, P. A. Grant, J. L. Workman, Purified histone acetyltransferase complexes stimulate HIV-1 transcription from preassembled nucleosomal arrays. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12924–12929 (1998).
26. J. Côté, R. T. Utley, J. L. Workman, in *Methods in Molecular Genetics* (Elsevier, 1995), vol. 6, pp. 108–128.