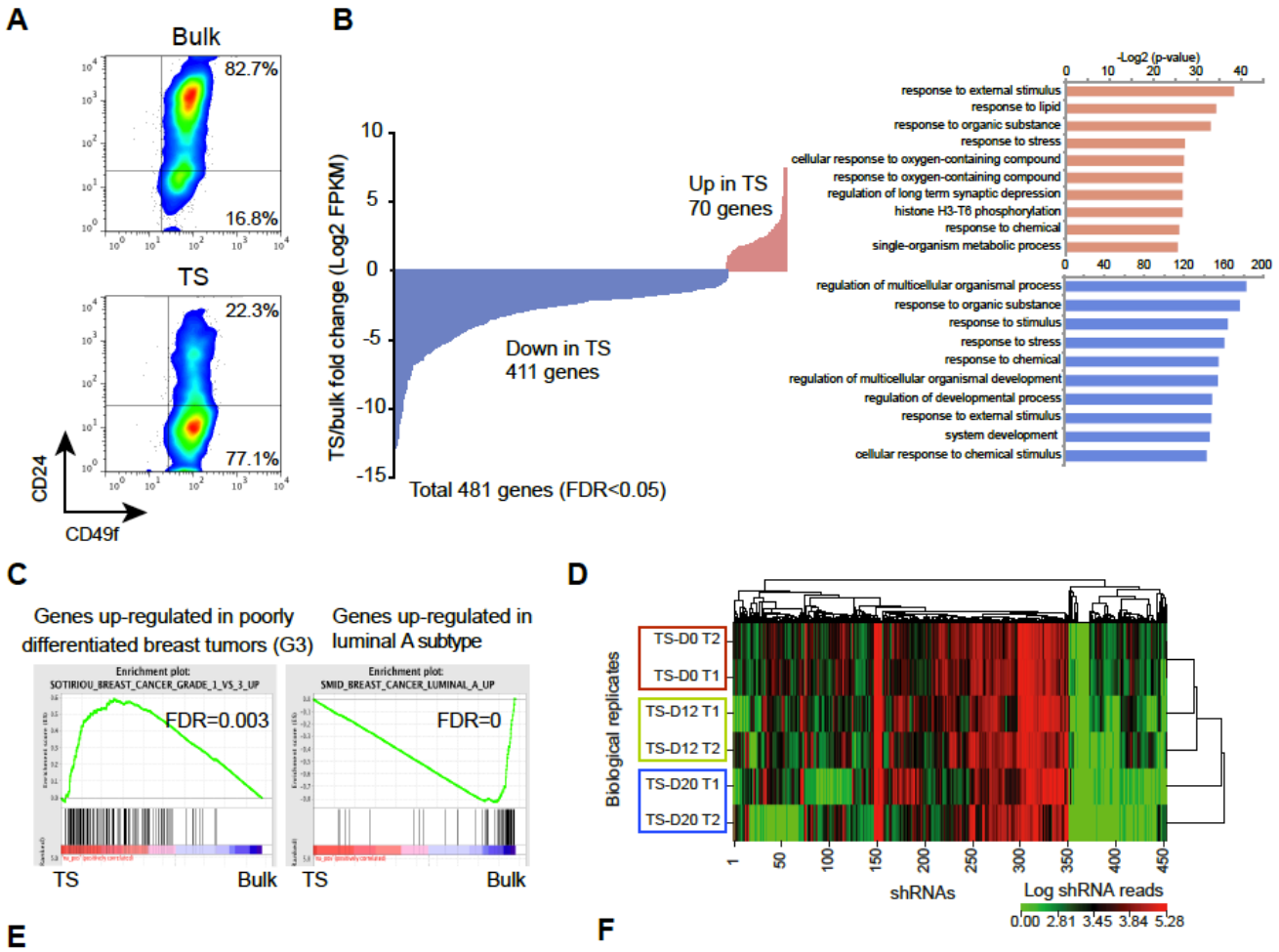
Supplemental Information

Cbx8 acts non-canonically with Wdr5 to promote mammary tumorigenesis

Chi-Yeh Chung, Zhen Sun, Gavriel Mullokandov, Almudena Bosch, Zulekha A. Qadeer, Esma Cihan, Zachary Rapp, Ramon Parsons, Julio A. Aguirre-Ghiso, Eduardo F. Farias, Brian D. Brown, Alexandre Gaspar-Maia, Emily Bernstein

SUPPLEMENTAL FIGURES



E

Selection Order	Criteria	shRNA Number	Percentage Total shRNA
1	> 500 average read at D0	411	91%
2	D12 and/or D20 significant read change (t test, p <0.1)	197	33%
3	TS specific read change over bulk	Depletion: 82 Enrichment: 12	Depletion: 18% Enrichment: 3%
4	Percentage of coherent individual shRNA changes ≥33% for a gene, genes with >2 shRNAs in library	10 genes	

F

Summary of TS RNAi screen top hits

Rank	Gene name	Total shRNA number	Percentage shRNA depleted in TS	Percentage TCGA alterations
1	*Srcap*	5	60%	15%
2	*Cbx8*	8	50%	11%
3	*Jarid2*	16	38%	9%
4	*Prdm2*	5	40%	8%
5	*Dnmt3a*	8	38%	8%
6	*Set*	8	50%	6%
7	*Cbx7*	5	60%	5%
8	*Dnmt3l*	8	38%	4%
9	*Hdac8*	5	40%	<1%
10	*Hdac6*	8	38%	<1%

Figure S1 (related to Figure 1). Mammary carcinoma TS RNAi screen identified key epigenetic factors. (A) FACS analysis of CD24 and CD49f of MMTV-Myc bulk and TS cells. (B) Representation of up (red) and down (blue) significantly altered genes (FDR<0.05) in day 20 TS compared to bulk MMTV-Myc cells (left). Gene ontology analyses are shown (right). (C) GSEA analysis of gene expression changes between TS and bulk MMTV-Myc cells using published gene sets associated with tumor grades and molecular subtypes (Smid et al., 2008; Sotiriou et al., 2006). (D) Representation of average

read counts of shRNAs during different time points of the TS screen (day 0, 12 and 20). Two individual tumors were used in the screen as biological replicates (T1 and T2). Unsupervised hierarchical clustering was used to compare the similarity between samples. (E) Hit selection criteria and statistics of the screen. (F) Summary of screen top ten hits. Hits were ranked by percentage of TCGA alterations.

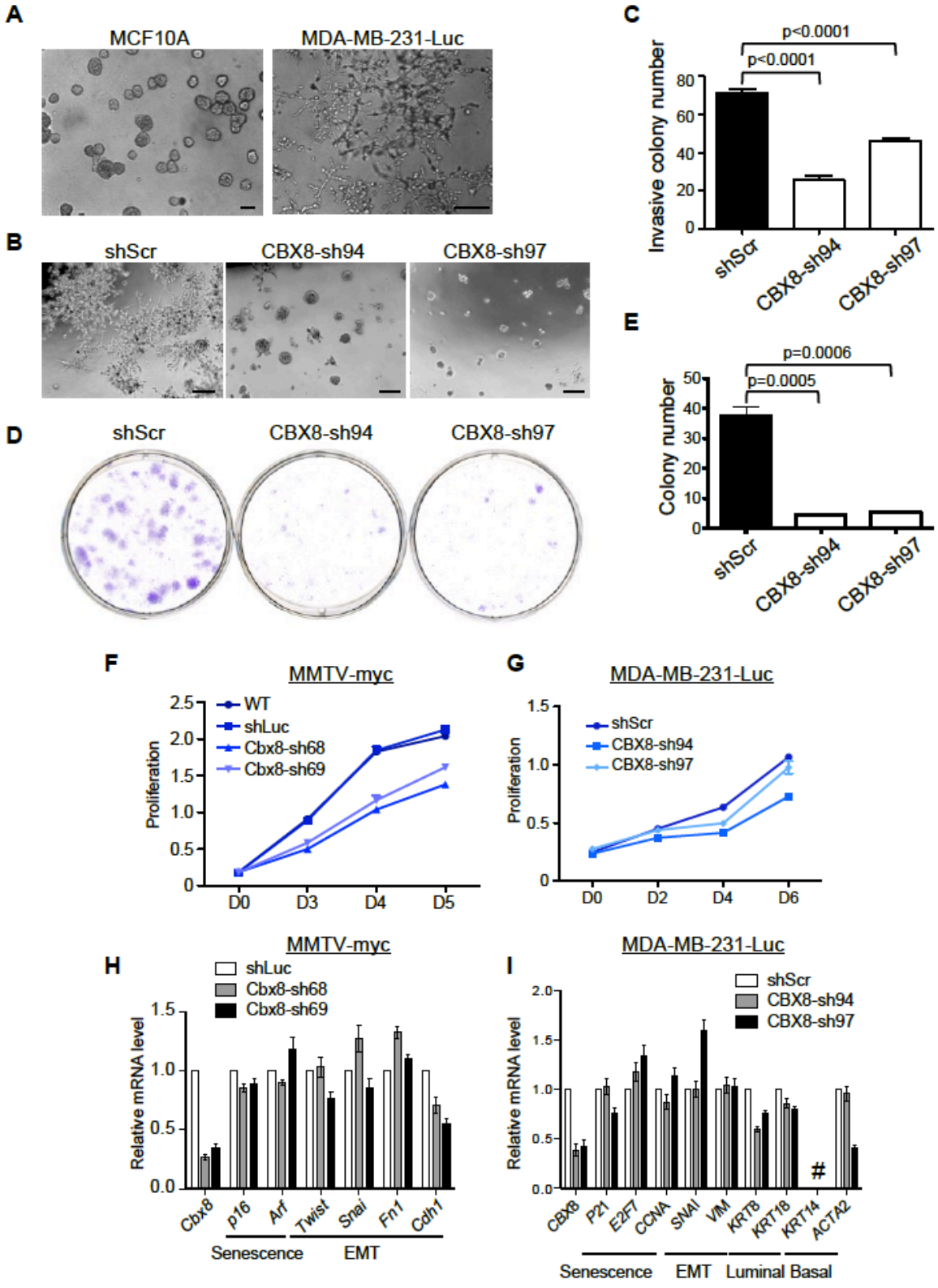


Figure S2 (related to Figure 2). Effects of CBX8 knock down on human breast cancer cells. (A) Representative bright field images of non-tumorigenic (MCF10A) and tumorigenic (MDA-MB-231-Luc) breast cells cultured in Matrigel. (B) Representative bright field images of control and CBX8 knockdown

MDA-MB-231-Luc cells cultured on Matrigel. (C) Quantification of invasive colony number from (B). Mean ± s.e.m (n = 6). (D) Representative wells showing clonogenic ability of control and CBX8 knock down MDA-MB-231-Luc cells. (E) Quantification of clonogenic ability from (D). Mean ± s.e.m (n = 3). (F) Growth curve of control and Cxb8 knockdown MMTV-Myc cells under bulk growing condition. Mean ± s.e.m (n = 5). (G) Growth curve of control and CBX8 knock down MDA-MB-231-Luc cells under bulk growing condition. Mean ± s.e.m (n = 5). (H) RT-qPCR analysis of control and Cbx8 knockdown MMTV-Myc cells with senescence and EMT markers. *Rpl7* was used for normalization. Mean ± s.e.m (n = 3). (I) RT-qPCR analysis of control and CBX8 knockdown MDA-MB-231-Luc cells with senescence, EMT, mammary luminal and basal markers. # = gene expression not detected. Note *p16/ARF* expression is undetectable in MDA-MB-231-Luc cells due to the loss of *INK4A/ARF* locus. *GAPDH* was used for normalization. Mean ± s.e.m (n = 3).

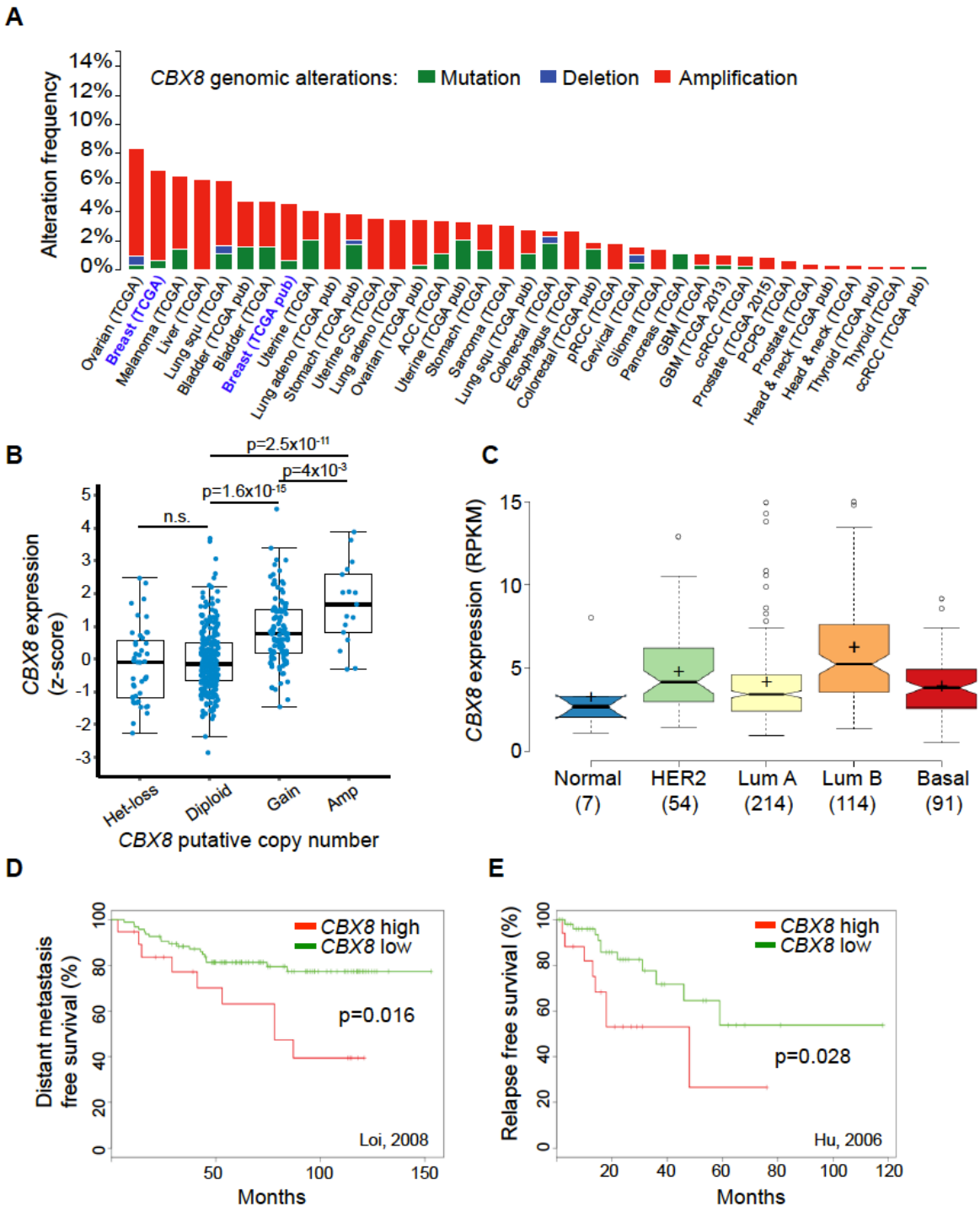


Figure S3 (related to Figure 3). CBX8 expression correlates with all subtypes and survival in breast cancer patients. (A) Genomic alteration status of CBX8 in different cancer types from published TCGA datasets (TCGA pub, Cancer Genome Atlas, 2012) and TCGA provisional (TCGA). Breast cancer datasets are highlighted in blue. (B) CBX8 putative copy number (GISTIC) and mRNA expression status from microarray (n = 457). n.s.=not significant. Boxes represent 25-75 percentiles. Bold lines represent the median. Whiskers represent 1.5x IQR. Blue dots represent individual data points. (C) RNA-Seq analysis of CBX8 expression in patients stratified by tumor molecular subtypes.

Crosses indicate sample mean. Patient numbers in each subtype are indicated. Box plot representations are the same as (B), except circles indicate outliers and notches indicate 95% CI. (D-E) Kaplan-Meier graph representing the probability of (D) distant metastasis free survival (Loi et al., 2008) and (E) relapse free survival of clinical data (Hu et al., 2006). Tumors were grouped by top 25% (CBX8 high) and lower 75% CBX8 (CBX8 low) expression. Logrank test p values are shown.

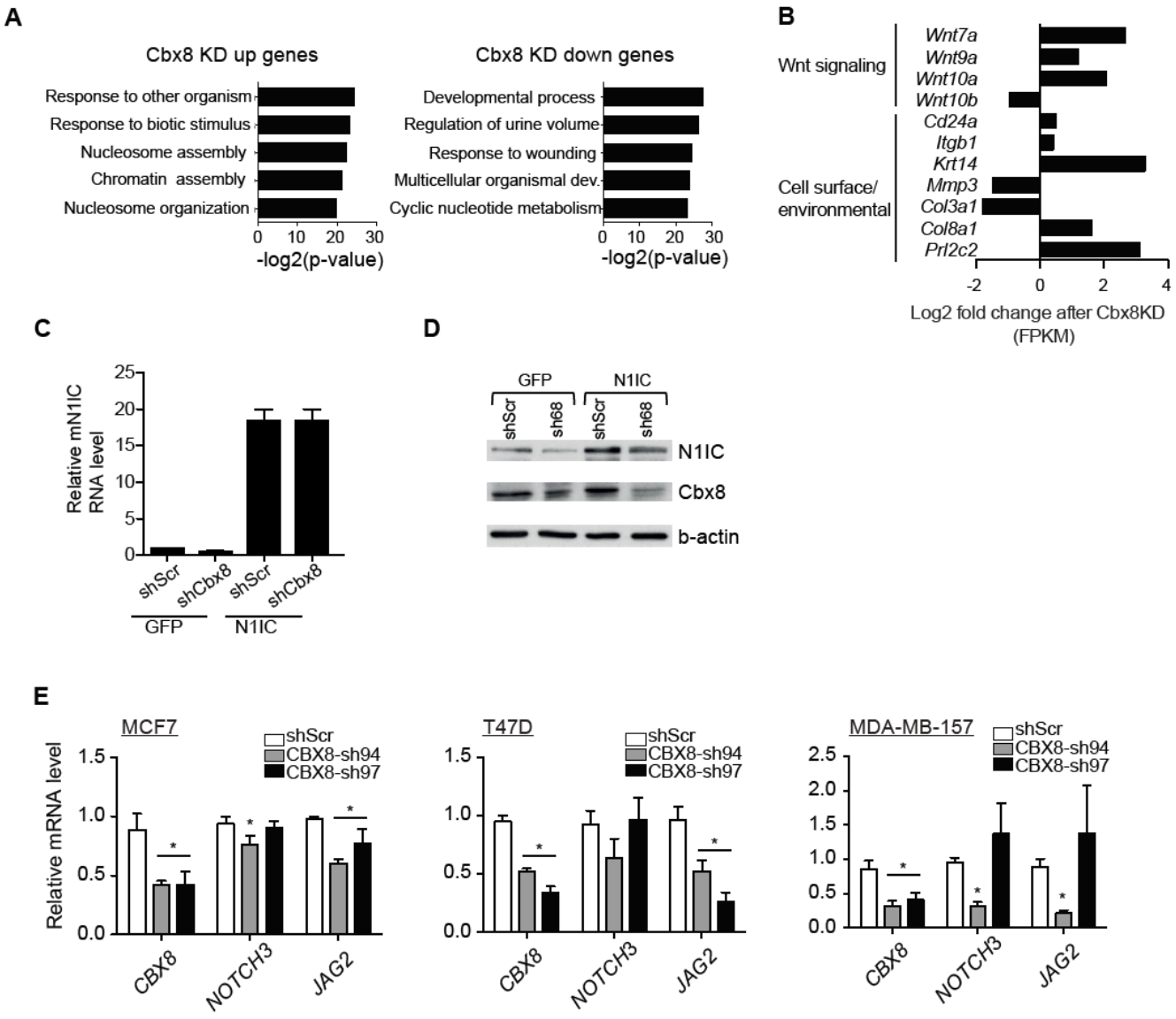


Figure S4 (related to Figure 4). The effect of Cbx8 knockdown on gene expression in breast cancer cells. (A) Gene ontology (GO) analysis of up- and down-regulated genes from RNA-seq analysis of Cbx8 knockdown MMTV-Myc cells in Figure 4A. (B) Change in expression of additional oncogenic players upon Cbx8 knockdown by RNA-seq. (C) Validation of Notch overexpression from rescue experiments. RT-qPCR analysis of mouse N1IC expression in the indicated samples in MMTV-Myc cells. *Rpl7* was used for normalization. Mean ± s.e.m (n = 3). (D) Immunoblotting of N1IC and Cbx8 in the indicated samples of MMTV-Myc whole cell lysate. B-actin used as loading control. (E) RT-qPCR analysis of NOTCH genes in control and CBX8 knockdown breast cancer cell lines. *GAPDH* was used for normalization. Mean ± s.e.m (n=3). *p<0.05 compared to shScr.

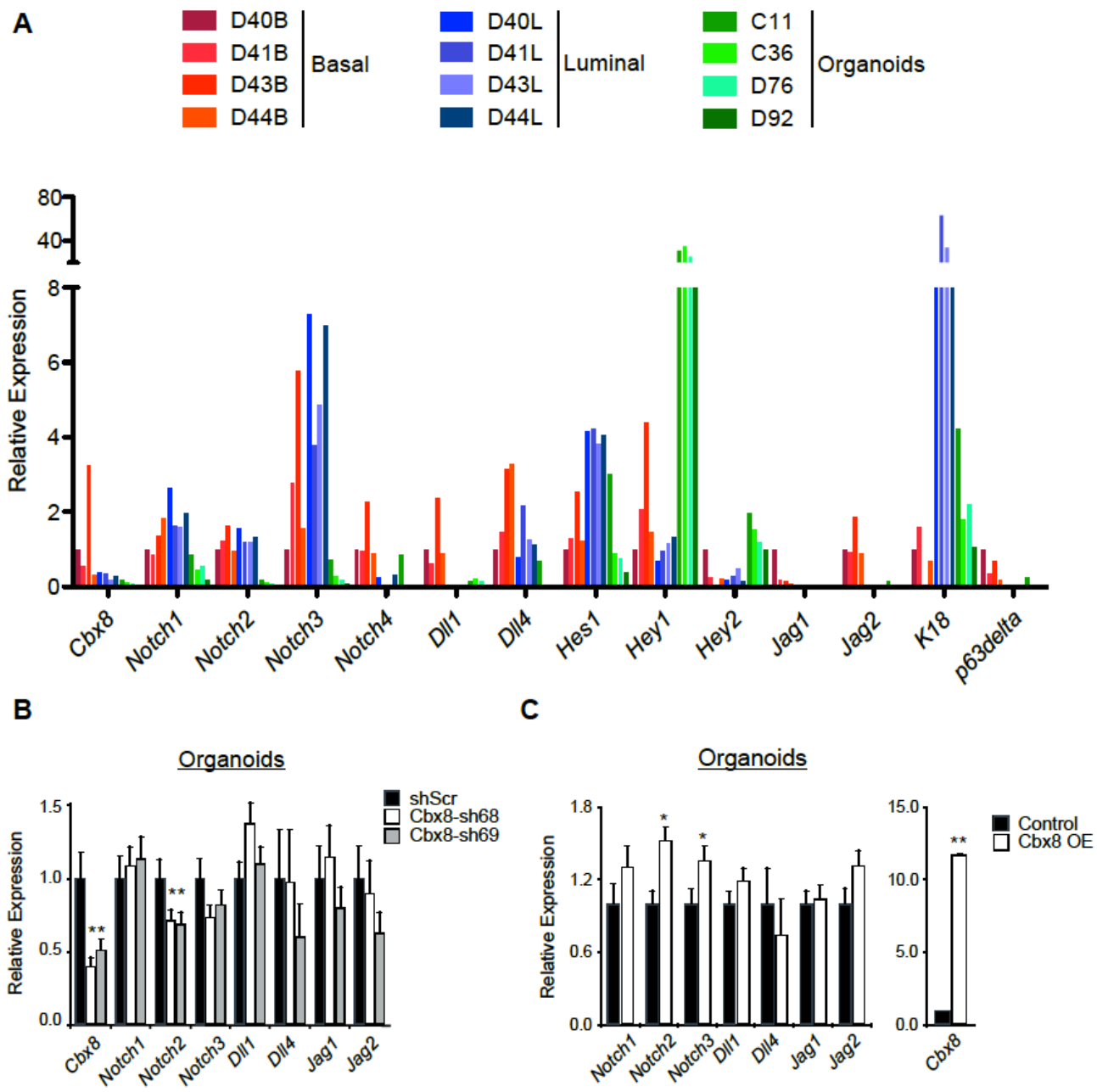
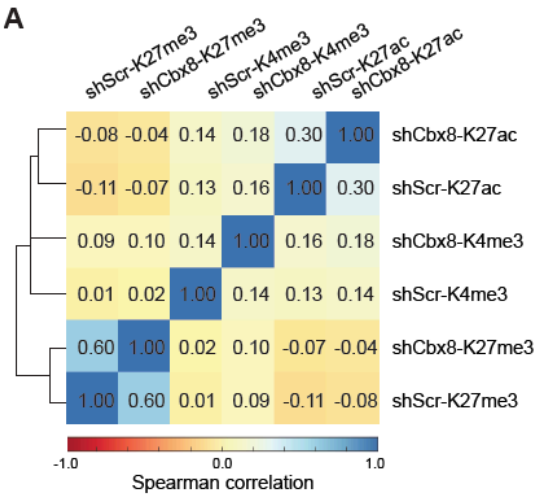


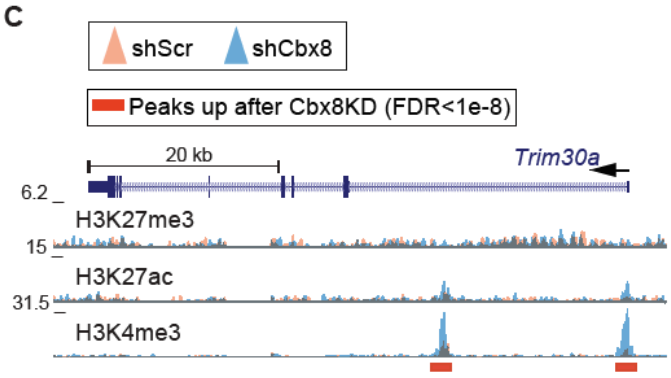
Figure S5 (related to Figure 5). Gene expression studies of Cbx8 and Notch in mouse mammary epithelial cells and organoids. (A) RT-qPCR analysis of Cbx8 and Notch-related genes from four FACS-sorted mammary glands with EpCAM low/CD49f high (basal), and EpCAM high/ CD49f low (luminal), and organoid cultures from four mammary glands. *Rpl7* was used for normalization. (B) RT-qPCR of Notch-related genes in control and Cbx8 knockdown organoids. *Rpl7* was used for normalization. Mean ± s.e.m (n = 3). * $p<0.05$ compared to control. (C) RT-qPCR of Notch-related genes in control and Cbx8 overexpressed organoids. *Rpl7* was used for normalization. Mean ± s.e.m (n = 3). * $p<0.05$, *$p<0.01$ compared to control.



B

Genes with histone modification changes upon Cbx8 KD in MMTV-Myc cells:

H3K4me3		H3K27ac		H3K27me3	
Down	Up	Down	Up	Down	Up
604	13	75	4	12	0



D

Genes with histone modification changes upon CBX8 KD in MDA-MB-231-Luc:

H3K4me3	
Down	Up
1,250	622

Figure S6 (related to Figure 6). The effect of Cbx8 knockdown on histone modifications. (A) Heatmap representing the correlation (Spearman) of histone modification ChIP signals between all samples (100bp bin) in MMTV-Myc cells. Unsupervised hierarchical clustering of samples is shown. (B) Gene numbers with significant histone modification changes at promoter (FDR<1X10^{-8} and fold change>1.8) upon Cbx8 knockdown in MMTV-Myc cells. (C) Overlay H3K27me3, H3K27ac and H3K4me3 profile (fold enrichment over input) on *Trim30a* in control and Cbx8 knockdown MMTV-Myc cells. Peaks significantly increased (FDR<1x10^{-8}) were underscored (red bars). (D) Gene numbers with significant H3K4me3 changes at promoter (FDR<2X10^{-3}) upon CBX8 knockdown in MDA-MB-231-Luc cells.

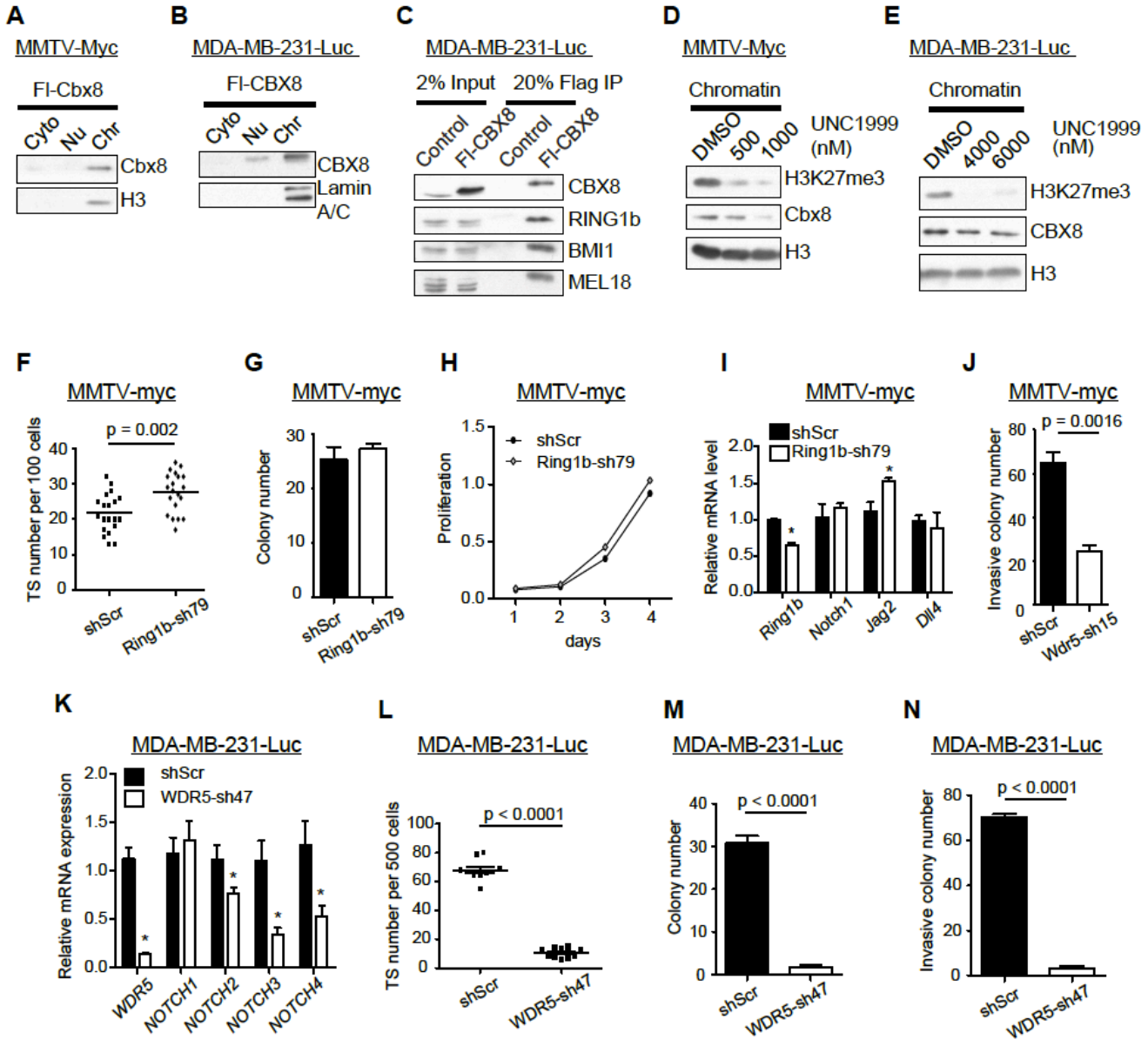


Figure S7 (related to Figure 7). Non-canonical PRC1 function of Cbx8 in breast cancer. (A) and (B) Immunoblotting of Cbx8 in cellular fractions of MMTV-Myc and MDA-MB-231-Luc cells. Cyto = cytoplasmic; Nu = nuclear chromatin free; Chr = chromatin. Histone H3 and Lamin A/C was used as control. (C) Co-immunoprecipitation of Flag-CBX8 with endogenous PRC1 members followed by immunoblotting in MDA-MB-231-Luc cells. (D) and (E) Chromatin immunoblotting of EZH1/2 inhibitor treated MMTV-Myc and MDA-MB-231-Luc cells. Histone H3 was used as loading control. (F) TS forming ability of control and Ring1b knockdown in MMTV-myc cells. Mean ± s.e.m (n = 20). (G) Clonogenic ability of control and Ring1b knockdown MMTV-myc cells. Mean ± s.e.m (n = 4). (H) Proliferation of control and Ring1b knockdown MMTV-Myc cells. Mean ± s.e.m (n = 6). (I) RT-qPCR of control and Ring1b knockdown MMTV-Myc cells. Mean ± s.e.m (n = 3). (J) Quantification of invasive colonies in control and Wdr5 knockdown MMTV-Myc cells cultured on Matrigel. Mean ± s.e.m (n = 3). (K) RT-qPCR analysis of control and WDR5 knockdown MDA-MB-231-Luc cells with WDR5 and

NOTCH genes. *GAPDH* was used for normalization. Mean ± s.e.m (n = 3). *p<0.05 compared to control. (L) TS forming ability of control and WDR5 knockdown MDA-MB-231-Luc cells. Mean ± s.e.m (n = 10). (M) Clonogenic ability of control and WDR5 knockdown MDA-MB-231-Luc cells. Mean ± s.e.m (n = 4). (N) Quantification of invasive colonies in control and WDR5 knockdown MDA-MB-231-Luc cells cultured on Matrigel. Mean ± s.e.m (n = 3).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Isolation of mammary epithelial cells (MEC)

Mammary glands from 129 wild type mice were minced and digested with 1.5mg/ml collagenase A (Roche) in DMEM/F12 medium at 37ºC for 2 hr. Digested samples were washed with PBS and spun down at 800rpm for 1 min. Cell were further digested with 0.05% trypsin for 5 min and 5mg/ml dispase (Stem Cell Technology) plus 100ug/ml DNase (Roche) for 5 min. Digested cells were then filtered through 40 mm cell strainers to obtain single cells. For separation of luminal and basal MEC subpopulations, single MECs were stained with antibodies against CD49f (PerCP-Cy5.5, BioLegend) and EpCAM-APC (BioLegend) after depletion with biotinylated antibodies against lineage markers (CD45, CD31, TER119, BP-1) using EasySep™ Mouse Mammary Stem Cell Enrichment Kit (Stemcell Technologies). Stained cells were sorted on FACSAria II sorter. Four different glands were used to isolate Basal (CD49f high/EpCAM low) and luminal (CD49f low/EpCAM high) populations.

Organoid Culture

MECs were prepared as above and cultured in Epicult-B medium (Stem Cell Technology) containing 3% Matrigel, 5% heat-inactivated FBS, 10 ng/ml EGF, 20 ng/ml bFGF, 4 mg/ml heparin, and 5 mM Y-27632 in ultra low attachment plates (Corning).

FACS

Cells were trypsinized and incubated at 1 $X10^7$ cells/ml in FACS buffer (DPBS and 2%BSA) at 4°C for 30 minutes using the following antibodies: anti-CD49f-FITC (mouse, BD #555735), anti-CD24-PE (mouse, BD #553262).

Constructs

pLKO.1 based lentiviral shRNAs (TRC Lentiviral shRNA, Open Biosystems) were used for all knockdown experiments. Sequences and clone IDs are provided in Table S6. GFP-tagged mouse N1IC

and human N2IC (MIGR1 vector) were provided by Iannis Aifantis (NYU). Gordon Peters (CRUK) provided Flag-tagged human CBX8 (pBabe-PURO). For overexpression experiments, mouse Cbx8 and Wdr5 were cloned by PCR with a self-cleaving peptide P2A and Neomycin resistance into a pLKO.1 vector, using NheI and BsrGI sites. Except for MIGR1-containing retroviruses, which was produced by transfecting Plate-E cells as described (Morita et al., 2000), all other viral constructs were produced by transfecting HEK293T cells. Lentiviral and retroviral transductions were performed using standard procedures.

Immunoblotting

Immunoblots were performed as described (Ratnakumar et al., 2012). Membranes were incubated with rabbit-anti-Cbx8 (1:2000, Bethyl, A300-882A), rabbit-anti-Notch1 (1:1000, Cell Signaling, #3608), rabbit-anti-Dll4 (1:1000, Abcam, ab7280), rabbit-anti-histone H3 (1:4000, Millipore, 05-928), mouse-anti-Ring1b (1:1000, Millipore, 05-1362), rabbit-anti-Mel18 (1:1000, Santa Cruz, SC-10744), mouse-anti-Bmi1 (1:1000, Millipore, 05-1321), rabbit-anti-histone H3K27me3 (1:1000, Millipore, 07-449), rabbit-anti-Wdr5 (1:1000, Abcam, ab22512), mouse-anti-LaminA/C (1:1000, Millipore, Mab3538) and mouse-anti-b actin (1:5000). Image quantification was performed using the gel analysis documentation of ImageJ.

Immunohistochemistry

Immunohistochemistry was performed as described (Kapoor et al., 2010). Mammary tumors were harvested and formalin-fixed and paraffin-embedded. 5 um tissue sections were incubated with Notch1 antibody (1:400, Cell Signaling, #3608). Images were taken on Nikon Eclipse Ci-L microscope with NIS Elements software (Nikon).

Matrigel culture

Morphogenesis analysis was performed using 3D cultures in Matrigel. Briefly, 3 x 10^3 MMTV-Myc and MDA-MB-231-Luc cells transfected with control and knockdown vectors were seeded on 2% Matrigel matrix (BD, #354234) with growth media as described (Debnath et al., 2003). Cells were cultured for 7 days and the morphology of the colonies was monitored daily by phase contrast microscopy and by confocal microscopy at the end of the experiment. The 3D cultures were processed and mounted in Prolong gold with DAPI (Molecular Probes, P36931) to visualize the nuclei and assess the induction of lumen formation by confocal microscopy as described (Debnath et al., 2003).

***In vitro* cellular assays**

Cells were plated in 96-well plates and proliferation was quantified by CellTiter 96 MTS solution (Promega) with 490nm plate reading at the indicated time points. For clonogenicity assays, 100-250 cells were plated in each well of 6-well plates for 7-10 days. When distinct colonies were formed, cells were stained with crystal violet and colonies were counted as described (Franken et al., 2006).

***In vivo* tumorigenesis**

For all experiments, the indicated number of cells were resuspended in 100µl of DMEM/F12 and injected into the mammary fat pad of female 8-12 week old (FvB/N) mice. The sites of injections were palpated every 48 h after injection and tumor size was measured by caliper twice a week. For bulk vs. TS experiments, 2 mice were injected with 2 injections per animal (left and right fat pad) for each sample (total injections = 4). All injections generated palpable tumors and were included for data analysis. For control vs. Cbx8 knockdown experiments, 8 mice were injected with 2 injections per animal for each sample (total injections = 16), and injections that did not generate tumors were regarded as unsuccessful injection and discarded for data analysis. The resulting sample numbers are n = 13 for shScr and n = 15 for shCbx8. Tumor size converted to tumor volume (mm^3) by: D X d^2 / 2. Mice were sacrificed after the indicated time points (when tumors reached 1000 mm^3) and tumors were

harvested for histological analyses. Mouse studies performed under ISMMS IACUC number LA09-00382.

cDNA generation and RT-qPCR

RNA was extracted with the QIAGEN RNeasy Mini Kit. 1 µg of total RNA was used to synthesize complementary DNA (cDNA) using Superscript II and Oligo dT primers (Invitrogen). RT-qPCR and gene expression analysis was carried out as described (Wu et al., 2013). Primer sequences are provided in Table S6.

Notch reporter assay

Notch-mCherry reporter was provided by Dimitris Placantonakis (NYU). pGreenFire1-Notch-puro reporter construct (System Biosciences) was modified by replacing the GFP and luciferase gene with mCherry. Lentivirus was produced by transfecting HEK293T cells using standard procedures. MMTV-Myc cells were infected with the Notch reporter and selected with 2µg/ml of puromycin for 4 days. The reporter cells were then infected with control or Cbx8 knockdown constructs. Three days later, cells were cultured as TS as described. TS were dissociated into single cells and subject to FACS analysis for mCherry fluorescence at day 15 of TS culture.

Cell fractionation

Cell fractionation was performed as previously described (Wysocka et al., 2001). Briefly, ~1 X 10^7 cells were collected, washed in PBS and resuspended in 1mL buffer A (10mM HEPES pH 7.9, 10mM KCl, 1.5mM MgCl2, 0.34M sucrose, 10% glycerol, 1mM DTT and 1 X protease inhibitor cocktail). Triton X-100 was added to 0.1% and the cells are incubated on ice for 10min. Nuclei were collected by centrifugation at 4000rpm at 4°C. The supernatant was taken as the cytosolic fraction. Nuclei were washed once with buffer A and then lysed for 30min in 'No Salt' buffer (3mM EDTA, 0.2mM EGTA,

1mM DTT and 1 X protease inhibitor cocktail) on ice. Chromatin was pelleted by centrifugation at 4000rpm at 4°C and supernatant was enriched in soluble nuclear proteins.

Co-immunoprecipitation

Co-immunoprecipitation was performed as described previously (Klauke et al., 2013). Briefly, whole cell extracts from MMTV-Myc and MDA-MB-231-Luc cells expressing FLAG-tagged mouse and human CBX8, respectively, were used for immunoprecipitation (IP). Whole cell extract was prepared by sonication (Bioruptor Plus, Diagenode) in buffer G (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.5% NP-40 and 1 X protease inhibitor cocktail). Anti-flag beads (EZView red anti-FLAG M2 resin, Sigma) were used to IP FLAG-tagged CBX8. After IP, beads were washed 3X in buffer G. For FLAG-CBX8 and WDR5 co-IP, beads were washed 2X in buffer G (250mM NaCl) and 1X in buffer G.

Ezh2 inhibitor treatment

MMTV-Myc and MDA-MB-231-Luc cells were treated with Ezh1/2 inhibitor UNC1999 at indicated concentrations for 7 days and chromatin was prepared as described (Konze et al., 2013).

Nuclear extraction and size exclusion chromatography

Nuclear extract was prepared as described (Abmayr et al., 2006). All procedures were performed at 4°C. Briefly, 1 X 10^8 – 1 X 10^9 MMTV-Myc cells were harvested, PBS washed and quickly resuspended in 5 X packed cell volume (pcv) of Dignam hypotonic buffer (10mM HEPES-KOH pH 7.9, 10mM KCl, 1.5mM MgCl2, 0.5mM DTT, 0.5mM PMSF and 1 X protease inhibitor cocktail). Cells were pelleted by centrifugation at 1850 g for 5 min. Cells were resuspended in 3 X pcv of Dignam hypotonic buffer and incubated on ice for 10 min to swell. Then cells were transferred to a Dounce homogenizer and homogenized slowly for 25 strokes with a loose pestle and 10 strokes with a tight pestle. Nuclei were pelleted by spinning 15 min at 3300 g. Nuclei were then broken and resuspended in 4 ml of Dignam high salt buffer (20 mM HEPES-KOH pH 7.9, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25%

glycerol, 0.5 mM DTT, 0.5 mM PMSF and 1 X protease inhibitor cocktail) and incubated 50 min on a rotation platform. Nuclear extracts were cleared by centrifugation at 20000 g for 1hr. Size exclusion chromatography was performed as described (Ding et al., 2012). In brief, nuclear extract was injected onto HiPrep 16/60 Sephacryl S-200 HR gel filtration column (GE Healthcare) controlled by AKTA Pure chromatography system (GE Healthcare). Samples were eluted at 1 ml/min and 2 ml fractions were collected. Fractions were subjected to immunoblotting analysis at equal loading with indicated antibodies. The column was calibrated with protein standard markers (Gel Filtration Calibration Kits, HMW, 28-4038-42, GE Healthcare) according to manufacturer's instructions.

RNA sequencing (RNA-seq)

Total RNA was extracted with the QIAGEN RNeasy Mini Kit and cDNA synthesis and sequencing was performed at the Mount Sinai Genomics Core Facility. In brief, RNA samples were quality checked (RIN score of 8 or higher) using the RNA 6000 Nano chip on a 2100 Bioanalyzer (Agilent). mRNA was then isolated using Sera-Mag oligo-dT beads (Thermo Fisher Scientific), and cDNA library was synthesized by SuperScript II (Invitrogen) with random primers. The cDNA library was quality checked and sequenced by Illumina Hi-Seq 2500 with 100nt single-end sequencing. RNA-Seq reads were mapped and analyses were performed as described (Trapnell et al., 2012). In brief, sequencing reads were first quality checked by FastQC (version 0.10.0) and then mapped to the reference genome (mouse mm9) by TopHat (version 1.4.0) (Trapnell et al., 2009) using the following parameters: maximum alignment number = 20, mismatch allowed = 2. Transcripts were assembled by Cufflinks program (version 1.3.0) (Trapnell et al., 2010) with quartile normalization and bias correction functions. Differential expression analysis was performed by Cuffdiff (version 0.0.7) program with the following parameters: library normalization method = geometric, dispersion estimation = blind, FDR = 0.05, minimum alignment count = 10, bias correction = on. Gene Ontology (GO) and canonical network analysis were conducted by Metacore (Thomson Reuters, https://portal.genego.com) using gene list and fold change data (Ashburner et al., 2000). Gene Set Enrichment Analysis (GSEA, Broad Institute,

http://www.broadinstitute.org/gsea/index.jsp) was performed using gene sets in the MSigDB database or published gene sets as indicated (Subramanian et al., 2005).

Gene expression and survival analysis of clinical data

Genomic alteration and copy number vs. expression analysis were extracted and plotted from cBioPortal for Cancer Genomics (http://www.cbioportal.org/public-portal/) (Gao et al., 2013). Normal vs. primary tumor expression, breast cancer subtype gene expression and survival data were extracted and plotted from ROCK-Breast Cancer Functional Genomics (http://rock.icr.ac.uk/index.jsp) (Sims et al., 2010). All patient data are from the 2012 TCGA breast cancer datasets unless otherwise noted (Cancer Genome Atlas, 2012).

ChIP-seq analysis

Sequencing reads were quality checked by FastQC (version 0.10.0) and NGS-QC generator (version 1.5.1) (Mendoza-Parra et al., 2013) prior to analysis. Sequence reads were then aligned to the NCBI Build 37 mouse genome (mm9) for MMTV-myc cells and human genome (hg19) for MDA-MB-231L cells with Bowtie (version 1.0.0) (Langmead et al., 2009) using the following parameters: seed length (l) = 70 bp, maximum mismatch (n) = 2, suppression (m) = 20, and reported alignments (k) = 1. Read difference between the control and Cbx8 knockdown samples of each ChIP were normalized by random selection of bed tags in the sample with higher reads to match that of the lower read sample (to normalize for read differences between samples). MACS2 (version 2.1.0) (Zhang et al., 2008) was used to generate Bedgraph files that show fold change enrichment of ChIP over input. Bedgraph files were then converted into BigWig files by BedClip program and uploaded onto UCSC genome browser for visualization and plotting. SICER-df (Zang et al., 2009) was used to reveal significantly changed peaks between the control and Cbx8 knockdown MMTV-Myc cells using the following parameters: window size = 50bp, gap size = 400bp, island calling = FDR<$5x10^{-55}$ (H3K4me3) and FDR<$1x10^{-4}$ (H3K27ac and H3K27me3), WT vs KD FDR<$1x10^{-8}$, WT vs KD fold change >1.8. For MDA-MB-231-Luc cells:

window size = 50 bp, gap size = 400 bp, island calling = FDR<$1x10^{-55}$, WT vs KD FDR<$2x10^{-3}$. Genes with significant histone modification changes were determined by intersecting significantly changed peaks of H3K4me3, H3K27ac and H3K27me3 ChIP (by SICER-df) to ±3 kb, ±5 kb and ±10 kb TSS of all Refseq genes, respectively, using Bedtools. Genes with significant changes after Cbx8 knockdown are summarized in Table S5. TSS analyses were performed using the SitePro tool from Cistrome (http://cistrome.org) (Liu et al., 2011). The genomic positions around TSS were generated using RefSeq gene annotation downloaded from UCSC Genome Browser (https://genome.ucsc.edu). Only the longest isoform of each gene was used to prevent double plotting of the same genomic region. Heatmaps showing the enrichment of ChIP samples around TSS of all genes were generated with 100 bp window without k-means clustering using computeMatrix and Heatmapper function from deepTools (http://deeptools.ie-freiburg.mpg.de) (Ramirez et al., 2014). Hierarchical clustering and correlation heatmap between each ChIP samples were generated with 100 bp window and Spearman correlation using bamCorrelate function from deepTools program. Histone modification snapshots were generated using UCSC Genome Browser.

Statistical analyses

Statistical inferences on tumor growth were analyzed with two-way ANOVA by Prism. Spearman correlation and significance between Cbx8 and Notch gene expression were analyzed by the rcorr function of Hmisc package in R statistical environment (R Foundation for Statistical Computing, http://www.R-project.org). Other statistical analyses were performed with two-tailed t-test assuming equal variances with Prism, Excel or R statistics.

SUPPLEMENTAL TABLES

Table S1. Gene expression data and Gene Ontology analyses of RNA-seq experiments in MMTV-Myc cells (related to Figure S1 and Figure 4).

Table S2. Target genes and shRNA sequences for the shRNA library used in tumorsphere screen (related to Figure 1).

Table S3. Raw data of bulk and tumorsphere shRNA screen in MMTV-Myc cells (related to Figure 1).

Table S4. Network analysis of Cbx8KD down genes in MMTV-Myc cells by Metacore (related to Figure 4).

Table S5. List of genes with significant histone mark change at promoter upon Cbx8KD (related to Figure 6 and Figure S6).

Table S6. List of shRNA sequences and RT-qPCR primers (related to experimental procedures).

shRNA sequences used for knockdown studies

Human		
Name	Mature antisense sequence (5'-3')	Clone ID
CBX8-sh94	TTTCACGAGGTATTCCATGCG	TRCN0000021894
CBX8-sh97	TTACTTTCCTTAATGGTGACG	TRCN0000021897
WDR5-sh47	GCCTCCTCTCTGAAGATGATT	TRCN0000118047
Mouse		
Name	Mature antisense sequence (5'-3')	Clone ID
Cbx8-sh68	AAAGCACTAAGCCATTTGAGC	TRCN0000093068
Cbx8-sh69	TTTCACGAGATATTCCATGCG	TRCN0000093069
Ring1b-sh79	CCCATCCAACTCTTATGGAAA	TRCN0000040579
Wdr5-sh15	GCAGCGTTAGAGAACGACAAA	TRCN0000034415

Primer sequences for RT-qPCR

Human	
Name	Sequence (5'-3')
CBX8-F	ATGGAGCTTTCAGCGGTGG
CBX8-R	ATGCGTCCTTTCCGTATGCG
NOTCH1-F	GACAGCCTCAACGGGTACAA
NOTCH1-R	CACACGTAGCCACTGGTCAT
NOTCH2-F	ACAGATGCGAGTGTGTCCCA
NOTCH2-R	GATTTCATACCCCGAGTGCCT
NOTCH3-F	GCCAAGCGGCTAAAGGTAGA
NOTCH3-R	TGAGTCCACTGACGGCAATC

NOTCH4-F	CTTGGAGAAGGGGCTGTGGAAT
NOTCH4-R	CTGGCAGGTCCCTTGTCCC
GAPDH-F	TTTGTCAAGCTCATTTCCTGG
GAPDH-R	TGATGGTACATGACAAGGTGC
JAG1-F	GTCTCAACGGGGGAACTTGT
JAG1-R	GCGTGCTCAGCAATTTCACA
JAG2-F	TGGGACTGGGACAACGATAC
JAG2-R	AGTGGCGCTGTAGTAGTTCTC
DLL1-F	GATTCTCCTGATGACCTCGCA
DLL1-R	TCCGTAGTAGTGTTCGTCACA
DLL4-F	CCATGCAAGAATGGGGCAAC
DLL4-R	GCCATCCTCCTGGTCCTTAC
HES1-F	TGAGCCAGCTGAAAACACTG
HES1-R	GTGCGCACCTCGGTATTAAC
HES5-F	CAAAGAGAAAAACCGACTGCGG
HES5-R	CGAGTAGCCTTCGCTGTAGT
HEY1-F	ATCTGCTAAGCTAGAAAAAGCCG
HEY1-R	GTGCGCGTCAAAGTAACCT
HEY2-F	GCAACAGGGGGTAAAGGCTA
HEY2-R	GCGCAACTTCTGTTAGGCAC
WDR5-F	AATTCAGCCCGAATGGAGAGT
WDR5-R	AGGCTACATCGGATATTCCCAG
P21-F	CAGGGGACAGCAGAGGAAGA
P21-R	TTAGGGCTTCCTCTTGGAGAA
E2F7-F	AGGCAGCCCAGACTAGATTTT
E2F7-R	GCTGGCAGCACTAATGAGCA
CCNA-F	GCGTTCACCATTCATGTGGA
CCNA-R	CAGGGCATCTTCACGCTCTATT
SNAI1-F	TCGGAAGCCTAACTACAGCGA
SNAI1-R	AGATGAGCATTGGCAGCGAG
VIM-F	GAGAGGAAGCCGAAAACACCC
VIM-R	TGGACATGCTGTTCCTGAATC
KRT14-F	TGAGCCGCATTCTGAACGAG
KRT14-R	GATGACTGCGATCCAGAGGA
ACTA2-F	GTGTTGCCCCTGAAGAGCAT
ACTA2-R	GCTGGGACATTGAAAGTCTCA
KRT8-F	TCCTCAGGCAGCTATATGAAGAG
KRT8-R	GGTTGGCAATATCCTCGTACTGT
KRT18-F	TCGCAAATACTGTGGACAATGC
KRT18-R	GCAGTCGTGTGATATTGGTGT
Mouse	
Name	Sequence (5’-3’)
Cbx8-F	ATTCGCAAAGGACGCATGGAA
Cbx8-R	CCTCGCTTTTTGGGGCCATA
Rpl7-F	AGCGGATTGCCTTGACAGAT
Rpl7-R	AACTTGAAGGGCCACAGGAA
Notch1-F	TCCTGAAGAACGGAGCCAAC
Notch1-R	CCAGCAACACTTTGGCAGTC

Jag2-F	CGGTGTCGCACCCTCTG
Jag2-R	TCGTCAATGTTTTCATGGCAGT
Dll4-F	AGCTGGAAGTGGACTGTGGT
Dll4-R	TAGAGTCCCTGGGAGAGCAA
Ring1b-F	GAGTTACAACGAACACCTCAGG
Ring1b-R	CAATCCGCGCAAAACCGATG
Wdr5-F	TTTGAAGATTTGGGACGTGAGTT
Wdr5-R	ATGGGCAGGCAAAGTCTTGAG
P16-F	GTGTGCATGACGTGCGGG
P16-R	GCAGTTCGAATCTGCACCGTAG
Arf-F	GCTCTGGCTTTCGTGAACATG
Arf-R	TCGAATCTGCACCGTAGTTGAG
Twist-F	GGACAAGCTGAGCAAGATTCA
Twist-R	CGGAGAAGGCGTAGCTGAG
Snai-F	CACACGCTGCCTTGTGTCT
Snai-R	GGTCAGCAAAAGCACGGTT
Fn1-F	ATGTGGACCCCTCCTGATAGT
Fn1-R	GCCCAGTGATTTCAGCAAAGG
Cdh1-F	CAGGTCTCCTCATGGCTTTGC
Cdh1-R	CTTCCGAAAAGAAGGCTGTCC
Hey1-F	TGAGCTGAGAAGGCTGGTAC
Hey1-R	ACCCCAAACTCCGATAGTCC
Hey2-F	TGAGAAGACTAGTGCCAACAGC
Hey2-R	TGGGCATCAAAGTAGCCTTTA
Notch4-F	GAGGACCTGGTTGAAGAATTGATC
Notch4-R	TGCAGTTTTTCCCCTTTTATCC
Dll1-F	CATGAACAACCTAGCCAATTGC
Dll1-R	GCCCCAATGATGCTAACAGAA
Jag1-F	ACACAGGGATTGCCCACTTC
Jag1-R	AGCCAAAGCCATAGTAGTGGTCAT
K18-F	GTCATACTGGGCAGGATGT
K18-R	CAAGATCGAAGACCTGAGG
p63deltaN-F	CCTGGAAAACAATGCCCAGAC
p63deltaN-R	GAGGAGCCGTTCTGAATCTGC

SUPPLEMENTAL REFERENCES

Abmayr, S. M., Yao, T., Parmely, T., and Workman, J. L. (2006). Preparation of nuclear and cytoplasmic extracts from mammalian cells. Curr Protoc Pharmacol *Chapter 12*, Unit12 13.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics *25*, 25-29.

Cancer Genome Atlas, N. (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61-70.

Debnath, J., Muthuswamy, S. K., and Brugge, J. S. (2003). Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. Methods *30*, 256-268.

Ding, J., Xu, H., Faiola, F., Ma'ayan, A., and Wang, J. (2012). Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res *22*, 155-167.

Franken, N. A., Rodermond, H. M., Stap, J., Haveman, J., and van Bree, C. (2006). Clonogenic assay of cells in vitro. Nature protocols *1*, 2315-2319.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E.*, et al.* (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling *6*, pl1.

Hu, Z., Fan, C., Oh, D. S., Marron, J. S., He, X., Qaqish, B. F., Livasy, C., Carey, L. A., Reynolds, E., Dressler, L.*, et al.* (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC genomics *7*, 96.

Kapoor, A., Goldberg, M. S., Cumberland, L. K., Ratnakumar, K., Segura, M. F., Emanuel, P. O., Menendez, S., Vardabasso, C., Leroy, G., Vidal, C. I.*, et al.* (2010). The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. Nature *468*, 1105-1109.

Klauke, K., Radulovic, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K.*, et al.* (2013). Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. Nature cell biology *15*, 353-362.

Konze, K. D., Ma, A., Li, F., Barsyte-Lovejoy, D., Parton, T., Macnevin, C. J., Liu, F., Gao, C., Huang, X. P., Kuznetsova, E.*, et al.* (2013). An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1. ACS chemical biology *8*, 1324-1334.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Liu, T., Ortiz, J. A., Taing, L., Meyer, C. A., Lee, B., Zhang, Y., Shin, H., Wong, S. S., Ma, J., Lei, Y.*, et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome biology *12*, R83.

Loi, S., Haibe-Kains, B., Desmedt, C., Wirapati, P., Lallemand, F., Tutt, A. M., Gillet, C., Ellis, P., Ryder, K., Reid, J. F*., et al.* (2008). Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. BMC genomics *9*, 239.

Mendoza-Parra, M. A., Van Gool, W., Mohamed Saleem, M. A., Ceschin, D. G., and Gronemeyer, H. (2013). A quality control system for profiles obtained by ChIP sequencing. Nucleic acids research *41*, e196.

Morita, S., Kojima, T., and Kitamura, T. (2000). Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene therapy *7*, 1063-1066.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic acids research *42*, W187-191.

Ratnakumar, K., Duarte, L. F., LeRoy, G., Hasson, D., Smeets, D., Vardabasso, C., Bonisch, C., Zeng, T., Xiang, B., Zhang, D. Y*., et al.* (2012). ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression. Genes & development *26*, 433-438.

Sims, D., Bursteinas, B., Gao, Q., Jain, E., MacKay, A., Mitsopoulos, C., and Zvelebil, M. (2010). ROCK: a breast cancer functional genomics resource. Breast cancer research and treatment *124*, 567-572.

Smid, M., Wang, Y., Zhang, Y., Sieuwerts, A. M., Yu, J., Klijn, J. G., Foekens, J. A., and Martens, J. W. (2008). Subtypes of breast cancer show preferential site of relapse. Cancer research *68*, 3108-3114.

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B*., et al.* (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. Journal of the National Cancer Institute *98*, 262-272.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols *7*, 562-578.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Wu, H. A., Balsbaugh, J. L., Chandler, H., Georgilis, A., Zullow, H., Shabanowitz, J., Hunt, D. F., Gil, J., Peters, G., and Bernstein, E. (2013). Mitogen-activated protein kinase signaling mediates phosphorylation of polycomb ortholog Cbx7. The Journal of biological chemistry *288*, 36398-36408.

Wysocka, J., Reilly, P. T., and Herr, W. (2001). Loss of HCF-1 chromatin association precedes temperature-induced growth arrest of tsBN67 cells. Molecular and cellular biology *21*, 3820-3829.

Zang, C., Schones, D. E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952-1958.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.