

Regulation of cancer epigenomes with a histone-binding synthetic transcription factor

David B. Nyer, Daniel Vargas, Caroline Hom, and Karmella A. Haynes

School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287

ABSTRACT

Chromatin proteins have expanded the mammalian synthetic biology toolbox by enabling control of active and silenced states at endogenous genes. Others have reported synthetic proteins that bind DNA and regulate genes by altering chromatin marks, such as histone modifications. Previously we reported the first synthetic transcriptional activator, the “Polycomb-based transcription factor” (PcTF), that reads histone modifications through a protein-protein interaction between the PCD motif and trimethylated lysine 27 of histone H3 (H3K27me3). Here, we describe the genome-wide behavior of PcTF. Transcriptome and chromatin profiling revealed PcTF-sensitive promoter regions marked by proximal PcTF and distal H3K27me3 binding. These results illuminate a mechanism in which PcTF interactions bridge epigenetic marks with the transcription initiation complex. In three cancer-derived human cell lines tested here, many PcTF-sensitive genes encode developmental regulators and tumor suppressors. Thus, PcTF represents a powerful new fusion-protein-based method for cancer research and treatment where silencing marks are translated into direct gene activation.

Keywords: Synthetic biology, transcription factor, epigenetics, histone methylation, Polycomb

Introduction

Proteins from the gene regulatory complex known as chromatin mediate stable, epigenetic expression states that persist over multiple cell divisions in metazoan tissues. Harnessing the potent gene-regulating functions of chromatin proteins has become a high-priority for cancer therapy and tissue engineering. The “histone code” model of chromatin function^{1,2} has strongly influenced work in epigenetic engineering and drug development^{3,4}. According to this model, biochemical marks are written onto DNA-bound histone proteins and these marks are read when effector proteins physically interact with the modified histones. Then the effector proteins enhance or inhibit transcription initiation. Much of our current understanding about chromatin-mediated gene regulation comes from deconstructive methods such as genetic mutations and RNA interference (for examples see^{5,6,7}). Constructive approaches, where synthetic systems are built from chromatin components, are gaining recognition as an important and powerful research method^{8,9} as well as a powerful application for biomedical engineering.

We constructed a synthetic transcription factor based on the natural effector protein CBX8. In many human cancers, lysine 27 on histone H3 (H3K27) becomes marked by trimethylation at abnormally high levels near tumor suppressor loci. CBX8-containing complexes accumulate at H3K27me3 and repress gene transcription^{10,11}. A central step in the Polycomb pathway is the

specific interaction between the H3K27me3 mark and a hydrophobic binding pocket within the Polycomb chromodomain (PCD) motif of CBX isoforms^{12, 13}. We used this interaction to design a Polycomb-based transcription factor (PcTF) that has VP64 (tetrameric VP16) and a visible red fluorescent tag (mCherry) fused to the C-terminus of the 60 amino acid PCD. The C-terminal VP64 domain allows PcTF to stimulate activation at repressed H3K27me3-associated genes (Fig. 1a). In previous work, we demonstrated H3K27me3-dependent PcTF activity at a model locus (*UASTk-luciferase*). PcTF also activated endogenous silenced loci in an osteosarcoma cell line (U-2 OS)¹⁴. Low-resolution protein mapping experiments at four loci identified two promoter regions that were co-occupied by PcTF and H3K27me3. Further investigation is needed to understand and accurately predict targets of the PcTF transcription activator. Here, we report genome-wide analyses that substantially advance our understanding of PcTF function. PcTF-stimulated gene activation in three different cancer cell types identified a large cohort of universally upregulated genes with a signature H3K27me3 enrichment profile. PcTF is enriched at transcription start sites within the nucleosome-free region of promoters and this profile depends upon the methyl-histone-binding PCD domain. We present a model where PcTF bridges distal H3K27me3 with endogenous transcription factors near the transcription start site. These findings provide significant progress towards predicting targets of PcTF based on distributions of epigenetic marks.

RESULTS

Common Polycomb-silenced loci are activated in PcTF-expressing cells.

We measured PcTF-mediated activation of known Polycomb targets in three different cell lines, osteosarcoma (U-2 OS), neuroblastoma (SK-N-SH), and leukemia (K562). We selected a panel of 13 genes for which Polycomb-mediated silencing is supported by genetic and pharmacologic disruption studies^{15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25} and protein mapping studies^{26, 27, 28, 29} in human cancers and stem cells: *ATF3*, *CADM1*, *CASZ1*, *CDH1*, *CLU*, *HK2*, *HOXA6*, *HOXB4*, *IRF8*, *LAMB3*, *RUNX3*, *TRIM22*, and *WT1*. We transfected U2-OS, SK-N-SH, and K562 cells with a PcTF-expressing plasmid (Fig. 1a) via Lipofectamine LTX. We used quantitative reverse transcription PCR (qRT-PCR) to measure mRNA from mock-transfected (vehicle only) and plasmid-transfected cells (Fig. 1b). Primers and dye-conjugated hydrolysis probes are described in Table S1. In all three cell lines, over half of the genes were activated two-fold or higher compared to mock-transfected controls. Up-regulation persisted for several genes as PcTF transcript levels decreased over time. Three of the Polycomb target genes were upregulated in all three cell types: *ATF3*, *CASZ1*, and *IRF8*.

We used data from chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) to detect H3K27me3 enrichment at Polycomb target genes. In U-2 OS, ChIP was performed as previously described using an antibody against H3K27me3¹⁴. DNA was purified from the immunoprecipitated chromatin, analyzed by next generation deep sequencing, and reads were aligned to the hg19 human genome consensus (Feb. 2009 GRCh37). Regions of enrichment compared to non-immunoprecipitated bulk chromatin (input) were calculated using the Hotspot

algorithm³⁰. We used shared and public data to determine H3K27me3 enrichment in SK-N-SH and K562 cells. *ATF3*, *CASZ1*, and *IRF8* are marked by H3K27me3 (Fig. S1) and become upregulated upon PcTF expression in all three cell types tested here. *HOXA6*, *HK2*, and *TRIM22* are enriched for H3K27me3 in one or two cell types, and only become up-regulated when H3K27me3 is present (Fig. S1). These results are consistent with a mechanism where PcTF accumulates at H3K27me3 sites and induces gene transcription.

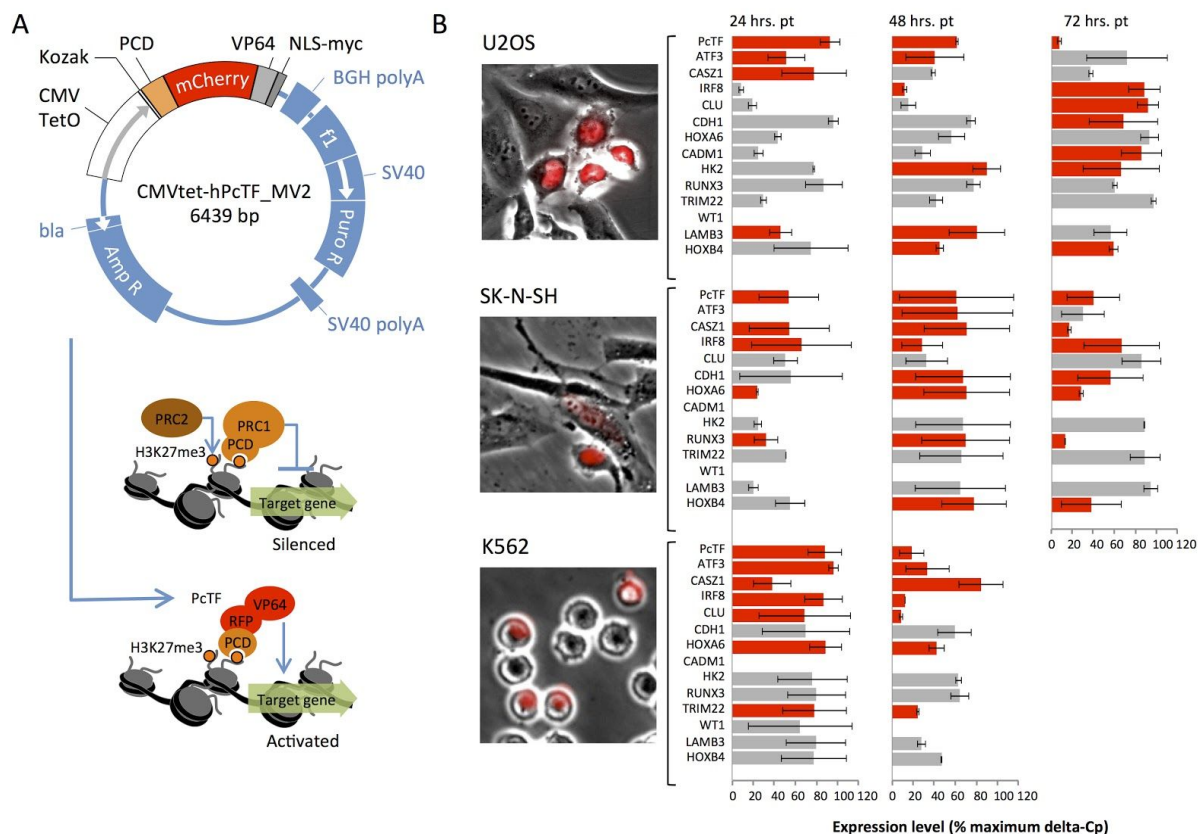


Figure 1. PcTF expression stimulates upregulation of known targets of Polycomb in transiently transfected cells. (A) Map of the PcTF-expressing plasmid (top). The natural PRC1 complex mediates gene silencing (middle). PcTF expression leads to accumulation of PcTF at H3K27me3 and gene activation (bottom)¹⁴. (B) Transiently-transfected, PcTF-positive U-2 OS, SK-N-SH and K562 cells are visualized via the red fluorescent protein tag mCherry (RFP). qRT-PCR was used to determine mRNA levels of PcTF and a panel of 13 target genes. Red bars represent 2-fold or greater GAPDH-normalized expression in PcTF-plasmid transfected cells compared to cells mock-transfected with the vehicle (Lipofectamine LTX) only. Hrs. pt = hours post-transfection. Error bars = standard deviation for two independent transfections.

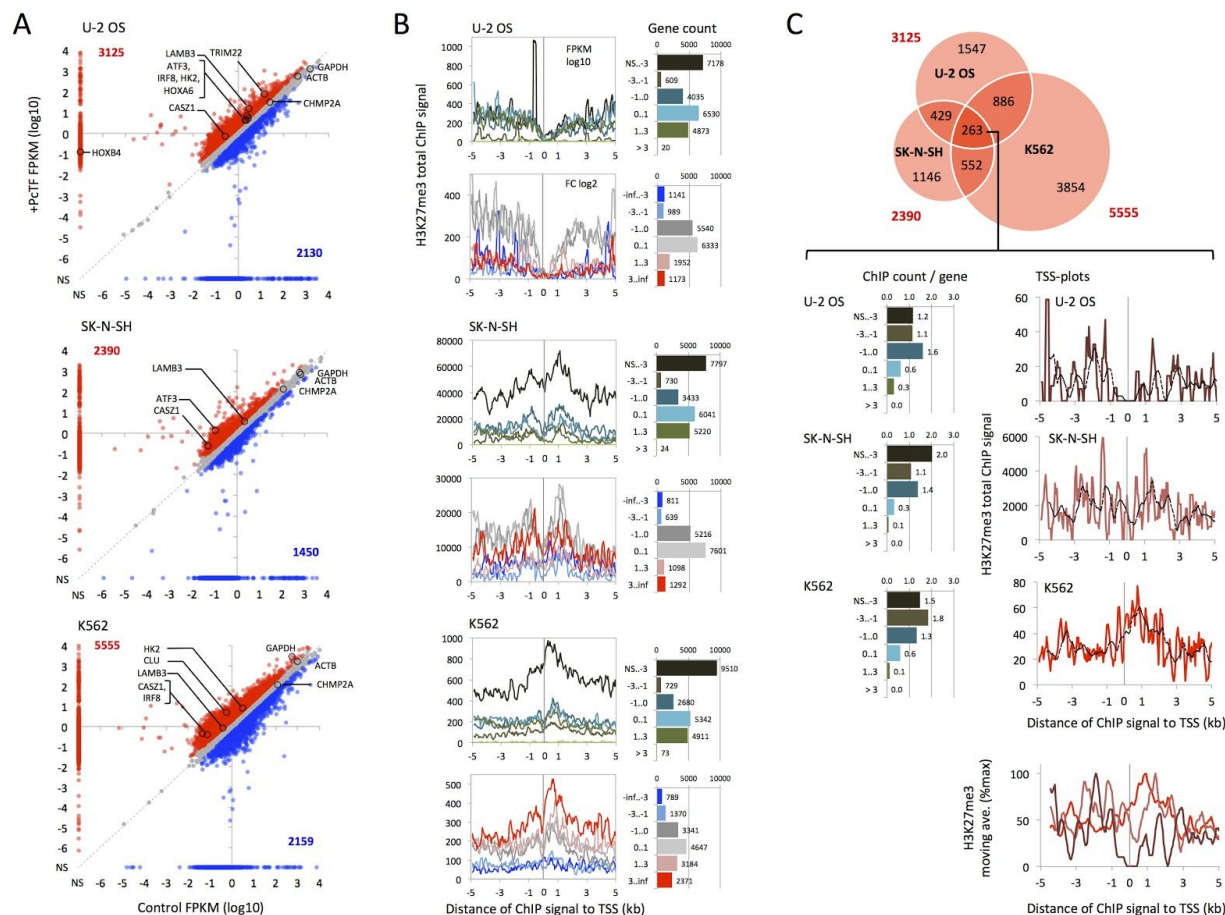
PcTF stimulates expression of a large subset of H3K27me3-marked genes in U-2 OS, SK-N-SH, and K562 cells.

Whole-transcriptome analysis confirmed up-regulation of the known Polycomb targets and

revealed a larger subset of commonly up-regulated genes. We performed next generation deep sequencing of total RNA (RNA-seq) from mock transfected (control) and PcTF-expressing cells (+PcTF) up to 96 hours post-transfection. Whole-transcriptome analyses of U-2 OS, SK-N-SH, and K562 corroborated the qRT-PCR results for 9 of the 13 known Polycomb targets: *ATF3*, *CASZ1*, *CLU*, *HK2*, *HOXA6*, *HOXB4*, *IRF8*, *LAMB3*, and *TRIM22*. Comparison of fragments per kilobase of gene model per million mapped reads (FPKM) for control versus +PcTF samples showed that 10.3 - 23.9% of 23,245 annotated Refseq genes became up-regulated 2-fold or higher (Fig. 2a). This result is consistent with a mechanism where PcTF accumulates at silenced, H3K27me3-enriched loci and activates transcription through its C-terminal VP64 domain. We also observed down-regulation (fold-change ≤ 2.0) at 6.2 - 9.3% of the genes in the three cell lines we tested. Repression is inconsistent with the transcriptional activation function of VP64. In some cases, PcTF might disrupt proper gene expression by placing VP64 in a suboptimal position relative to the promoter^{31, 32}.

We looked for an epigenetic signature of PcTF-sensitive genes by mapping enrichment of H3K27me3 near transcription start sites (TSS's). Overall, H3K27me3 enrichment was depleted near transcription start sites at silenced and active genes, indicated by a sharp valley at position zero in stratified TSS-plots (Fig. 2b). This profile has been observed in other reports; similar plots show a valley in the TSS profile at the nucleosome-free region (NFR) of promoters in mammalian cells^{33, 34}. Genes with lower basal expression levels, prior to PcTF expression, show the highest levels of H3K27me3 enrichment which is consistent with the role of H3K27me3 in gene silencing. Genes that became up-regulated in response to PcTF expression showed the highest H3K27me3 levels compared to non-responsive and down-regulated genes in K562 cells. This result provides evidence of a mechanism where PcTF accumulates at methyl-histone marks through the N-terminal PCD and activates silenced genes through the C-terminal VP64 domain. In SK-N-SH and U-2 OS however, up-regulated genes showed the same or lower H3K27me3 levels as non-responsive genes. Therefore, PcTF-sensitivity may not be strictly determined by methylation signal levels that were determined by ChIP analysis.

We hypothesized that a specific H3K27me3 distribution pattern around the TSS determines which genes are activated by PcTF. We identified a subset of 263 commonly upregulated genes (Fig. 2c) in order to exclude genes that behave in a cell-type-specific manner. In U-2 OS and SK-N-SH, we observed depletion of H3K27me3 at the TSS, surrounded by signal peaks at intervals of roughly 1 kb. ChIP hotspot densities around the TSS were highest for genes with low basal expression levels (0.001 - 1.0 FPKM) (Fig. 2c). The highest peaks for K562 were observed immediately downstream of the TSS and 3 kb upstream (Fig. 2c). Altogether, these data suggest that PcTF-mediated activation of silenced genes may rely on H3K27me3 enrichment at positions distal to the TSS.



ChIP-seq analyses show PCD-dependent enrichment of PcTF near transcription start sites

In order to further investigate the mechanism of PcTF-mediated gene regulation, we used ChIP-seq to compare the distribution of PcTF and H3K27me3 throughout the genome and at TSS regions. We performed ChIP-seq on chromatin from an isogenic PcTF-expressing U-2 OS cell line. The cell line “U2OS-PcTF” carries a chromosomally-integrated, TetR-repressed gene

that expresses myc-tagged PcTF when the cells are treated with doxycycline (Fig. 3a), as described previously¹⁴. We used this system for ChIP-seq instead of transiently transfected cells in order to eliminate negative background from PcTF-minus cells. ChIP was performed as previously described using either an antibody against the PcTF N-terminal myc tag for U2OS-PcTF cells or anti-H3K27me3 for the Flp-in T-REx parental line¹⁴. ChIP-seq analysis of a truncated version of PcTF that lacks the PCD domain (Δ TF) showed little overlap with PcTF and H3K27me3 (Fig. S2), which indicated that PcTF and H3K27me3 distributions were not due to random crosslinking. We analyzed the distributions of PcTF and H3K27me3 ChIP at different genomic features and observed the greatest enrichment over background for intergenic regions (Fig. 3b). PcTF, but not H3K27me3, showed signal enrichment at promoters. These results indicate that PcTF and H3K27me3 co-occupy non-coding regions, and that PcTF is frequently enriched at promoter sites.

We inspected promoter regions more closely by identifying PcTF and H3K27me3 enriched regions surrounding TSS's. We identified 5535 regions that were marked with PcTF or H3K27me3 or both. A TSS plot showed a peak of PcTF enrichment at the TSS over a 2 kb interval (Fig. 3c). Randomly crosslinked Δ TF did not show the same enrichment profile, therefore mCherry-tagged VP64 alone does not account for the enrichment pattern of PcTF. In contrast to PcTF, H3K27me3 was depleted at the TSS. The subset of 316 regions in which we detected both PcTF and H3K27me3 showed a similar but shorter PcTF enrichment peak. H3K27me3 signal was depleted near the TSS, with dispersed peaks of H3K27me3 further from the TSS. The profiling data suggest that at co-occupied regions, PcTF enrichment appears at and adjacent to the TSS, while H3K27me3 is farther away either upstream or downstream. Overall, these results suggest that PcTF accumulates near the TSS, and that this accumulation depends upon interaction with distal H3K27me3 through the N-terminal PCD peptide.

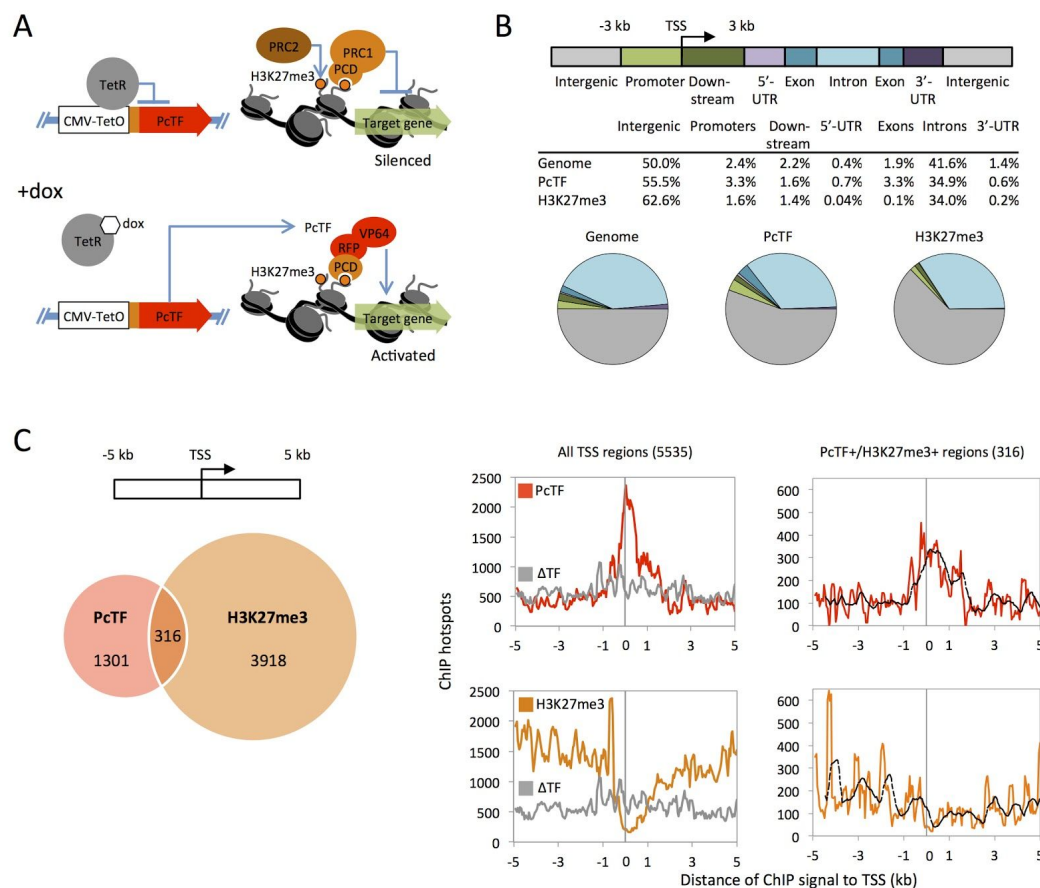


Figure 3. ChIP-seq analysis of PcTF and H3K27me3 distribution in a transgenic U2-OS cell line. (A) Illustrated mechanism of PcTF function. Doxycycline (dox) induced PcTF expression in U2OS-PcTF cells leads to accumulation of PcTF at H3K27me3-positive promoters and gene activation¹⁴. (B) Distribution of PcTF and H3K27me3 ChIP signals (hotspots) in genes and intergenic regions compared to background (Genome) distributions. (C) The venn diagram compares genes that have PcTF- and H3K27me3-marked TSS regions. TSS-centered plots (left) show the total ChIP hotspot signal value within a sliding window of 200 bp (step size = 50 bp). ΔTF is a negative control protein that lacks the PCD (H3K27me3-binding) motif. The right-most plots show the subset of 316 genes where the TSS region is co-occupied by PcTF and H3K27me3. Dashed lines show the moving average (window = 10).

Up-regulated genes are marked by H3K27me3 upstream and PcTF near the TSS in U2OS-PcTF cells

Up-regulation of PcTF-sensitive genes corresponds with H3K27me3 enrichment at roughly 1 - 2 kb upstream of the transcription start site in regions where we also detected PcTF enrichment near the TSS. To identify up-regulated genes, we compared RNA-seq FPKM values of 23,244 genes for U2OS-PcTF cells that were untreated or treated with doxycycline (dox) to induce PcTF expression. Known Polycomb targets *LAMB3*, *IRF8*, *CASZ1*, *RUNX3*, *HOXB4*, and *CDH1* were up-regulated 2-fold or higher. For transcript levels that changed at least two-fold in either direction, the predominant response was an increase in expression of genes that had a low

initial expression state in control cells. FPKM values of silenced genes are typically lower than 1.0³⁵. The frequency of upregulated genes within the non-expressing FPKM range of 10⁻⁴ - 10⁰ was 23%, whereas only 3% of active genes were upregulated (Fig. 4a).

Stratified TSS plots showed that PcTF was frequently enriched immediately downstream of the TSS at up-regulated genes (Fig. 4b). H3K27me3 was depleted near the TSS and showed peaks of enrichment 1 kb, 2 kb and 4 kb upstream. Median distances of the nearest H3K27me3 signal were -2545.5 bp for 2- to 8-fold up-regulated genes and -1365 bp for genes that were up-regulated 8-fold or higher (Fig. 4c). We observed a significant ($p < 0.05$) increase in expression across two independent experimental replicates for five of the upregulated genes: *PTRF*, *CDKRAP3*, *PPIA*, *ELFN2*, *POR*, and *CARD10* (Fig. 4d). For genes that showed no change or were down-regulated, median distances ranged from -588.5 to 93.5 bp. Positioning of H3K27me3 very close to the TSS at non-responsive and down-regulated genes might interfere with VP64-driven gene activation, or cause VP64 to recruit transcription factors to a site that is not optimal for proper transcription. These data indicate that a distal H3K27me3 mark positions PcTF at an optimal position to allow the VP64 activation domain to stimulate gene transcription at target promoters.

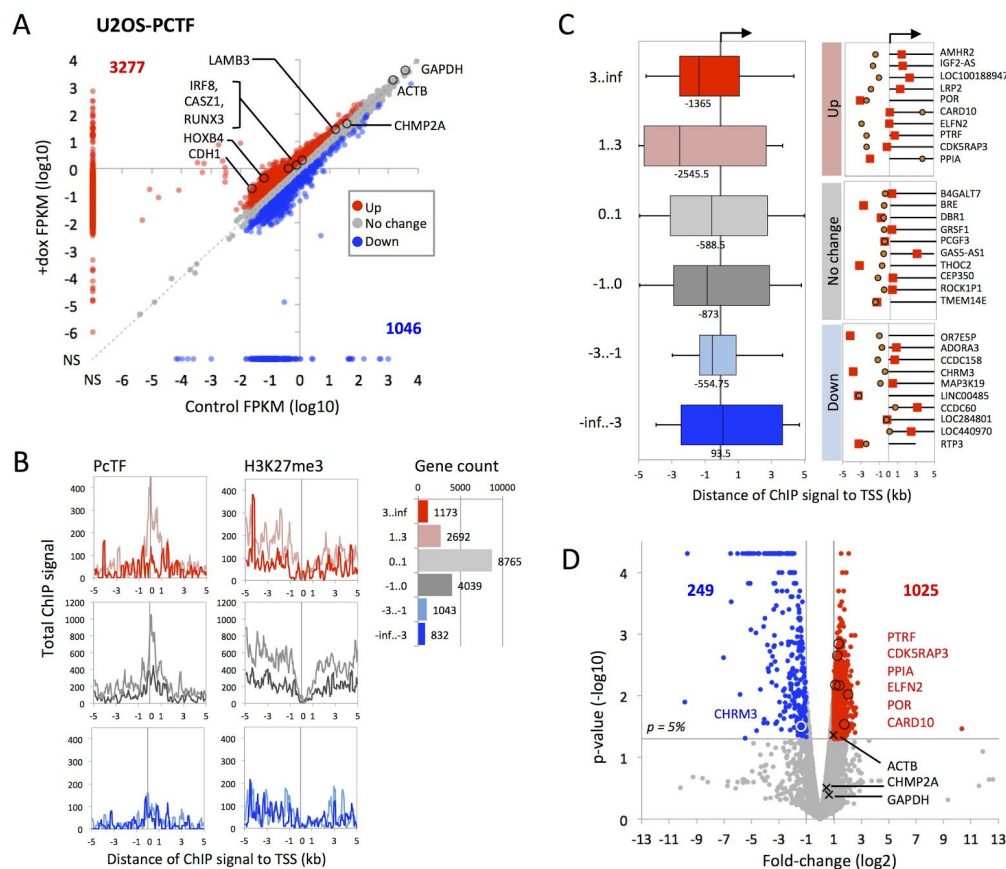


Figure 4. Analysis of PcTF and H3K27me3 localization and gene regulation. (A) The x/y scatter plot compares RNA-seq signals (FPKM log10) of U2OS-PcTF cells without PcTF (Control) and after PcTF induction (+dox). Known Polycomb-regulated genes and control genes from Figures 1 and 2 are highlighted

for comparison with previous experiments. (B) TSS-centered plots show the total ChIP hotspot signal value (window = 200 bp, step size = 50 bp) stratified by log2 fold-change expression. (C) H3K27me3 ChIP signal distances from the TSS were compared for genes of lengths ≥ 2 kb in the subset of 316 genes where the 10 kb region is co-occupied by PcTF and H3K27me3. Genes were grouped by fold-change expression (as in B). Box plots show median values (solid vertical line), 25th (left box) and 75th (right box) percentiles, and minimum (left whisker) and maximum values (right whisker). TSS maps drawn to scale show the midpoints of ChIP signals for H3K27me3 (orange circle) and PcTF (red square) relative to the TSS. Genes are sorted from highest to lowest log2 fold-change value. (D) The volcano plot shows statistical significance versus fold change (same color scheme as in Fig. 1a) for duplicate RNA-seq experiments. Up- and down-regulated genes from panel C are marked with circles. Control genes are marked with an “x.”

H3K27me3 enrichment occurs near enhancers at PcTF-sensitive loci

Sites that are distal from the TSS are known as enhancers, which are often bound by regulatory proteins that interact with transcription factors at the TSS and stimulate or repress gene expression. We hypothesized that PcTF stimulates expression at genes where a nearby enhancer is enriched for H3K27me3. Investigation of the chromosomal positions of PcTF-sensitive genes, enhancers, and H3K27me3 revealed that histone methylation at upstream enhancers corresponds with PcTF-mediated gene activation. Enhancer regions were identified from the 15 classes of human chromatin states from the ENCODE project for 9 different human cell types³⁶, including K562. Enhancer coordinates were based on class 4 (strong enhancer) and class 5 (medium enhancer) hg19 annotations from the public K562 cell data. We identified the closest enhancer element near 23,245 TSS's at both the upstream and downstream position and calculated median distances to the nearest H3K27me3 signal (Fig. 5a). Genes with FPKM values of zero in both control and treated cells were excluded from the calculations. We observed that H3K27me3 ChIP signals occur close to upstream enhancer midpoints at higher frequencies near up-regulated genes. U-2 OS cells showed the strongest correlation (Pearson $R^2 = 0.79$) between H3K27me3-proximal, upstream, class 4 enhancers and fold-change expression in response to PcTF.

Close inspection of gene regions further supported the significance of H3K27me3 positioning at enhancers near PcTF-sensitive genes. At *CASZ1*, one of the known Polycomb targets, the upstream enhancer regions were positioned near H3K27me3 enrichment signals in all three cell types. *ELFN2*, a gene that became up-regulated after PcTF expression in all three cell types, also showed H3K27me3 proximity to upstream enhancers. At *AMHR2*, H3K27me3 ChIP hotspots appear about 1 kb from the TSS, between the nearest enhancer and the nearest PcTF ChIP signal in U-2 OS and K562 (Fig. 5b). *AMHR2* was up-regulated by PcTF in U-2 OS and K562, but not in SK-N-SH where we found no H3K27me3 signal within the same region. These results suggest that H3K27me3 at enhancer elements upstream of the TSS contributes to PcTF-mediated gene activation.

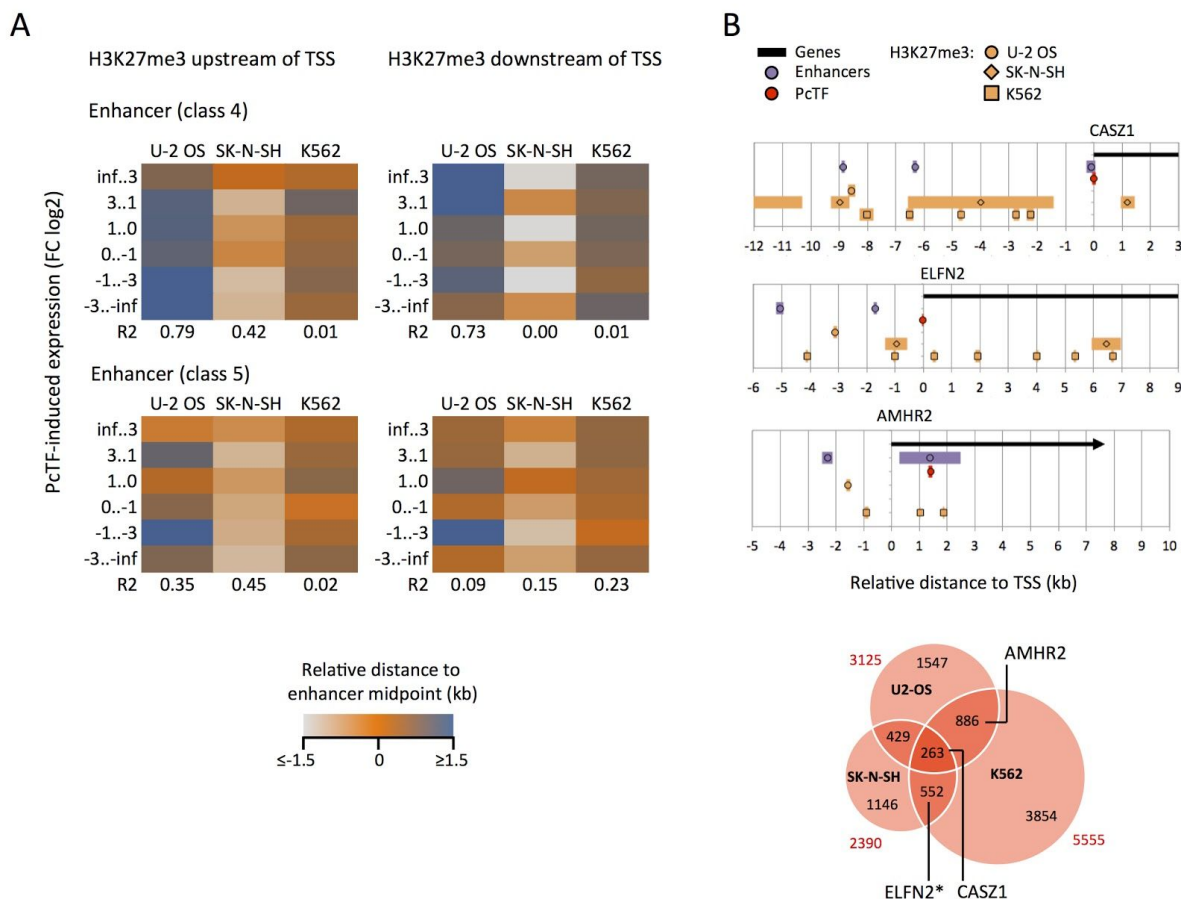


Figure 5. H3K27me3 enrichment at enhancer regions. (A) Color scales show median distances between midpoints of nearest H3K27me3 ChIP hotspots and enhancer elements closest to the TSS of all genes. Data are stratified by fold-change in expression level after PcTF expression. Enhancer coordinates were determined using chromatin states classes from the ENCODE project (UCSC Browser HMM track for K562 cells)³⁶. R2 = Pearson correlation coefficient of median distance versus fold-change bin. (B) High-resolution maps of features at select genes that are upregulated in one or more cell lines. The Venn diagram from Fig. 2c is shown here with labels for genes that are shown in the maps. *ELFN was up-regulated in U2OS-PcTF and not in transiently transfected U-2 OS cells.

DISCUSSION

The work presented here provides new insights into the mechanism of PcTF, a synthetic chromatin-based transcription factor. PcTF couples recognition of a silencing-associated mark with gene activation. PcTF enrichment near TSS sites is similar to the patterns that have been observed for subunits of the transcription initiation complex^{33, 37}, suggesting a strong interaction with endogenous transcription factors. The C-terminal domain of PcTF includes VP64 which is composed of four copies of a core acidic transcription activation domain (TAD) from VP16³⁸. The TAD is derived from the H1 region of VP16, which has been shown to bind with high affinity (high nanomolar range) to the Mediator complex subunit MED25³⁹. The VP16 TAD also

interacts with MED17, several members of the Transcription Factor II (TFII) family, and the TATA promoter motif-binding protein TBP (reviewed in^{40, 41}). In contrast, the C-terminal PCD has relatively low affinity (high micromolar range) for H3K27me3⁴². PCD-H3K27me3 binding may be weak *in vivo*, but it is still necessary for PcTF function¹⁴ and for TSS-proximal enrichment (Fig. 3C). Enrichment of PcTF in regions that are near to but do not overlap with H3K27me3 was observed in our previous ChIP-PCR studies¹⁴. This offset of PcTF ChIP signals from H3K27me3 sites suggests that PcTF engages with H3K27me3 and then becomes trapped near the TSS through interactions with the transcription initiation proteins in crosslinked chromatin. The data provide strong evidence for a mechanism where PcTF bridges distal histone methylation marks with PolII-associated transcription factors at the nucleosome-free TSS (Fig. 6).

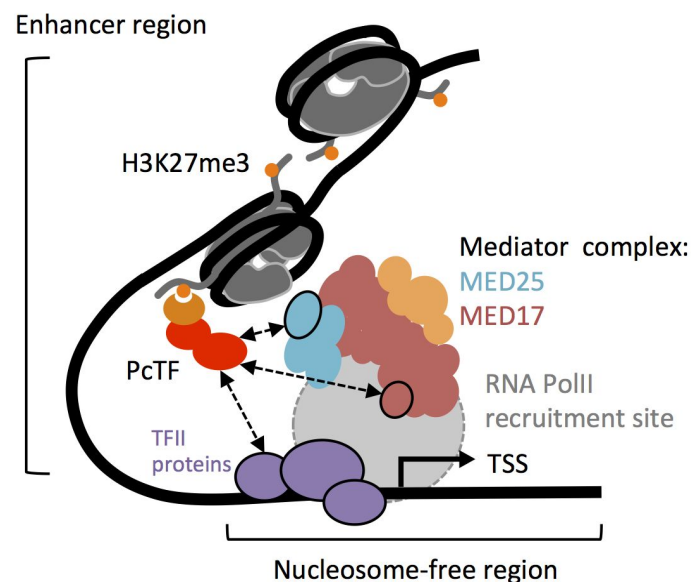


Figure 6. Model for PcTF-mediated gene activation. PcTF binds to H3K27me3 (orange dots) through the N-terminal PCD domain (orange crescent) at an enhancer region that is distal to the transcriptional start site (TSS). The C-terminal VP64 domain (red) interacts with members of the transcription initiation complex, including MED25, MED17, and TFII proteins. DNA is shown as a thick black line. The core histone complexes of nucleosomes and histone H3 tails are depicted in dark grey.

RNA-seq results for genes that do not appear to be direct targets of PcTF-mediated activation provide additional insights into gene regulation. Many of the genes we analyzed in the RNA-seq experiments do not change more than two-fold or become down-regulated after PcTF expression. In these cases, suboptimal positioning of the VP64 domain of PcTF relative to the promoter might neutralize its activity or lead to reduced expression below basal levels^{31, 32}. We observed several genes that became up- or down-regulated, but lacked a PcTF signal within 10 kb of the TSS. Regulation of these genes could be mediated by transcription regulators that are direct targets of PcTF. For example up-regulated, PcTF-marked genes including *ATF3* and *CASZ1* encode broadly-acting transcriptional regulators that control multiple targets^{43, 44}.

Our work has significant implications for medicine. Currently, epigenetic engineering relies

largely on small-molecule inhibitors that are intended to erase cancer-associated epigenetic silencing and indirectly activate therapeutic genes. These drugs have major shortcomings such as incomplete erasure of cancer-associated methyl-histone marks and DNA damage. Here, we used a synthetic chromatin protein, PcTF, to convert the methyl-histone signal into direct activation of therapeutic genes instead of inhibitors to erase cancer-associated marks. Genes that were consistently upregulated by PcTF in all three cell lines we tested have therapeutic potential. Work by others has shown that induced upregulation of the transcription factor *ATF3* antagonizes cancer cell growth and invasion in certain tissues, such as ovarian (OVAC)⁴⁵ and esophageal (ESCC)⁴⁶. Overexpression of the *CASZ1* transcription regulator inhibits proliferation of neuroblastoma (SH-SY5Y)⁴⁷. Additional work to determine the specificity of PcTF in cancer versus healthy cells and to enhance efficient synthetic protein delivery will bring this promising new technology closer to translation.

ONLINE METHODS

DNA constructs

The doxycycline-inducible PcTF-expressing plasmid has been described previously¹⁴. The plasmid is constitutively expressed in the absence of a TetR protein (in cell lines U-2 OS, SK-N-SH, and K562). The full annotated sequence of “hPCD-TF” is available online at Benchling - Hayneslab: Synthetic Chromatin Actuators (<https://benchling.com/hayneslab/f/S0I0WLoRFK-synthetic-chromatin-actuators/seq-fSG2GUbr-hpcd-tf/>).

Chromatin Immunoprecipitation followed by deep sequencing (ChIP-seq)

Chromatin immunoprecipitation was performed on U-2 OS Flp-In T-Rex cells, which carry a chromosomal insert of a TetR-repressible PcTF gene as previously described¹⁴. A ChIP-Seq DNA Sample Prep Kit (Illumina IP-102-1001) was used to prepare deep sequencing libraries from DNA that was purified from immunoprecipitated (IP) and non-IP (input) chromatin. End-repair was carried out at 20°C for 30 min. in the following 50 µl reaction: 30.0 µl ChIP DNA, 1x T4 DNA ligase buffer, 0.4 mM dNTP, 1.0 µl T4 DNA Polymerase, 1.0 µl Klenow DNA polymerase, 1.0 µl T4 polynucleotide kinase. End-repair products were concentrated into a final volume of 34 µl Qiagen Elution Buffer (QIAquick PCR Purification kit 28104). 3'-end adenine base extension was carried out at 37°C for 30 min. in the following 50 µl reaction: 34.0 µl end-repair reaction product, 1x Klenow buffer, 0.2 mM dATP, 1.0 µl Klenow exonuclease. Base extension products were concentrated into a final volume of 10 µl H₂O (Zymo Clean and Concentrator D4003). Adapter ligation was carried out at room temperature for 15 min. in the following 30 µl reaction: 10 µl 3'-base extension reaction product, 1x ligase buffer, 1.0 µl adapter oligo mix, 4.0 µl DNA ligase. Ligation products were concentrated into a final volume of 10 µl H₂O (Zymo Clean and Concentrator D4003). 150 - 200 bp fragments from each ligation reaction were resolved and purified via gel electrophoresis and extraction (Zymoclean Gel DNA recovery kit D4001). DNA was back-eluted from the column twice with 10 µl H₂O and brought to a final volume of 36 µl. Adapter-modified 150-200 bp fragments were enriched by the polymerase

chain reaction (PCR) in the following 50 μ l reaction: 36.0 μ l gel-purified DNA, 1x Phusion buffer, 0.3 mM dNTP, 1.0 μ l PCR primer 1.1, 1.0 μ l PCR primer 2.1, 0.5 μ l Phusion polymerase. The cycling program was 98°C/30 sec, 18x[98°C/10 sec, 65°C/30 sec, 72°C/30 sec], 72°C/5 min, 4°C/ ∞ . PCR products were concentrated into a final volume of 15 μ l H₂O (Zymo Clean and Concentrator D4003). Libraries were size-confirmed on an Agilent 2100 Bioanalyzer and subjected to deep sequencing with single-end 100 bp reads on an Illumina Hi-Seq SR flow cell platform.

Cell culture and transfection

Complete growth media contained 10% tetracycline-free fetal bovine serum and 1% penicillin and streptomycin (pen/strep). U-2 OS and U2OS-PcTF cells were cultured in McCoy's 5A. K562, and SK-N-SH cells were cultured in IMDM, or EMEM, respectively. Cells were grown at 37 °C in a humidified CO₂ incubator. U2OS-PcTF stable cell lines were generated by previously published work¹⁴. PcTF-expressing U-2 OS, K562, and SK-N-SH cells were generated by transfecting 5x10⁵ cells in 6-well plates with DNA/Lipofectamine complexes: 2 μ g of plasmid DNA, 7.5 μ l of Lipofectamine LTX (Invitrogen), 2.5 PLUS reagent, 570 μ l OptiMEM. Transfected cells were grown in pen/strep-free growth medium for 18 hrs. The transfection medium was replaced with fresh, pen/strep-supplemented medium and cells were grown for up to 48 hrs. (K562) or up to 96 hrs. (U-2 OS and SK-N-SH).

Quantitative Reverse Transcription PCR (qRT-PCR)

Total messenger RNA was extracted from ~90% confluent cells (~1-2x10⁶). Adherent cells (U2OS and SK-N-SH) were lysed directly in culture plates with 500 μ l TRIzol. Suspended cells (K562) were collected into tubes, pelleted by centrifugation at 1000 xg for 3 min. at room temperature, separated from the supernatant, and lysed in 500 μ l TRIzol. TRIzol cell lysates were extracted with 100 μ l chloroform and centrifuged at 12,000 xg for 15 min. at 4 °C. RNA was column-purified from the aqueous phase (Qiagen RNeasy Mini kit 74104). SuperScript III (Invitrogen) was used to generate cDNA from 2.0 μ g of RNA. Real-time quantitative PCR reactions (15 μ l each) contained 1x LightCycler 480 Probes Master Mix (Roche), 2.25 pmol of primers (see Supplemental Table 1 for sequences), and 2 μ l of a 1:10 cDNA dilution (1:1000 dilution for GAPDH and mCh). Crossing point (C_p) values, the first peak of the second derivative of fluorescence over cycle number, were calculated by the Roche LightCycler 480 software. For each experiment (one cell type, one target gene, all time points) C_p values were normalized to the highest value. "Expression level" was calculated as ΔC_p , $2^{[C_p \text{ GAPDH} - C_p \text{ experimental gene}]}$, for PcTF-expressing cells. Fold-change was determined as Expression level divided by ΔC_p for non-PcTF control cells.

RNA-seq

RNA-seq was performed using two replicates per experimental condition. Total RNA was prepared as described for qRT-PCR. 50 ng of total RNA was used to prepare cDNA via single primer isothermal amplification using the Ovation RNA-Seq System (Nugen 7102-A01) and automated on the Apollo 324 liquid handler (Wafergen). cDNA was sheared to approximately 300 bp fragments using the Covaris M220 ultrasonicator. Libraries were generated using Kapa

Biosystem's library preparation kit (KK8201). In separate reactions, fragments from each replicate sample were end-repaired, A-tailed, and ligated to index and adapter fragments (Bioo, 520999). The adapter-ligated molecules were cleaned using AMPure beads (Agencourt Bioscience/Beckman Coulter, A63883), and amplified with Kapa's HIFI enzyme. The library was analyzed on an Agilent Bioanalyzer, and quantified by qPCR (KAPA Library Quantification Kit, KK4835) before multiplex pooling and sequencing on a Hiseq 2000 platform (Illumina) at the ASU CLAS Genomics Core facility.

Bioinformatics analysis

ChIP-seq alignments were carried out using the Bowtie2 algorithm⁴⁷ with the hg19 reference genome (Feb. 2009 GRCh37). Enrichments normalized to input were calculated using the Hotspot algorithm (distribution version 4)⁴⁸. ChIP-seq data for SK-N-SH was provided by B. Bernstein and K562 data was retrieved from the UCSC Genome Browser website⁴⁹. Galaxy (<http://www.usegalaxy.org>)⁵⁰ was used to identify overlaps with a minimum of 1 bp between gene intervals and ChIP-seq enrichment intervals. ChIP-seq profiling across genes and intergenic regions was performed using the Cistrome platform⁵¹. RNA-seq alignments were carried out with de-multiplexed 50-bp single-end reads and the hg19 transcriptome using Bowtie 2. Transcript abundance and differential expression were calculated using Tophat and Cuffdiff⁵² within the Galaxy online platform using the *Homo sapiens* UCSC hg19 gene transfer format (gtf) file from Illumina iGenomes. Outdated genes were identified by cross-referencing gene symbols with the NCBI database. Out-dated genes and duplicate transcription start sites were removed from the Cuffdiff output in Excel, resulting in 23,245 genes. Analysis of distances between features (TSS's, ChIP signals, and enhancers) was performed and all graphs and charts were generated in Excel.

ACKNOWLEDGEMENTS

D. Nyer was supported by the Arizona Department of Health Services. D. Vargas was supported by the Western Alliance to Expand Student Opportunities (WAESO). C. Hom was supported by the Fulton Undergraduate Research Initiative (FURI). K.A. Haynes was supported by the NIH NCI (K01 CA188164) and the NSF Synthetic Biology Research Engineering Center (SynBERC). DNA oligos were purchased with support from Integrated DNA Technologies, a SynBERC IAB member. The authors thank J. Kramer, J. Steel, J. Park, and N. Briones from the Biodesign Institute at ASU for assistance with ChIP-seq and RNA-seq, and B. Bernstein for SK-N-SH ChIP-seq data.

REFERENCES

1. Jenuwein T, Allis CD. (2001) Translating the histone code. *Science*. 293:1074-80.
2. Bannister AJ, Kouzarides T. (2011) Regulation of chromatin by histone modifications. *Cell Res*. 21:381-95.

3. Sharma S, Kelly TK, Jones PA. (2010) Epigenetics in cancer. *Carcinogenesis*. 31:27-36.
4. Venugopal B, Evans TR. (2011) Developing histone deacetylase inhibitors as anti-cancer therapeutics. *Curr Med Chem*. 18:1658-71.
5. Schotta G, Ebert A, Dorn R, Reuter G. (2003) Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. *Semin Cell Dev Biol*. 14:67-75.
6. Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev*. 20:1123-36.
7. Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. (2015) Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol*. 33:661-7.
8. Keung AJ, Joung JK, Khalil AS, Collins JJ. (2015) Chromatin regulation at the frontier of synthetic biology. *Nat Rev Genet*. 16:159-71.
9. Thakore PI, Black JB, Hilton IB, Gersbach CA. (2016) Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat Methods*. 13:127-37.
10. Yamaguchi H, Hung MC. (2014) Regulation and Role of EZH2 in Cancer. *Cancer Res Treat*. 46:209-22.
11. Wang W, Qin JJ, Voruganti S, Nag S, Zhou J, Zhang R. (2015) Polycomb Group (PcG) Proteins and Human Cancers: Multifaceted Functions and Therapeutic Implications. *Med Res Rev*. 35:1220-67.
12. Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., & Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & Development*, 17(15), 1870–1881.
13. Min, J., Zhang, Y., & Xu, R.-M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes & Development*, 17(15), 1823–1828.
14. Haynes, K.A. and Silver, P.A. (2011) Synthetic reversal of epigenetic silencing. *J Biol Chem*, 286:27176–27182.
15. Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev*. 20:1123-36.
16. Tan J, Yang X, Zhuang L, Jiang X, Chen W, Lee PL, Karuturi RK, Tan PB, Liu ET, Yu Q. (2007) Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes Dev*. 21:1050-63.
17. Cao Q, Yu J, Dhanasekaran SM, Kim JH, Mani RS, Tomlins SA, Mehra R, Laxman B, Cao X, Yu J, Kleer CG, Varambally S, Chinnaiyan AM. (2008) Repression of E-cadherin by the polycomb group protein EZH2 in cancer. *Oncogene*. 27:7274-84.
18. Nuytten M, Beke L, Van Eynde A, Ceulemans H, Beullens M, Van Hummelen P, Fuks F, Bollen M. (2008) The transcriptional repressor NIPP1 is an essential player in EZH2-mediated gene silencing. *Oncogene*. 27:1449-60.
19. Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultschi AI, Wutz A, Bickmore WA. (2010) Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell*. 38:452-64.

20. Kalushkova A, Fryknäs M, Lemaire M, Fristedt C, Agarwal P, Eriksson M, Deleu S, Atadja P, Osterborg A, Nilsson K, Vanderkerken K, Oberg F, Jernberg-Wiklund H. Polycomb target genes are silenced in multiple myeloma. (2010) PLoS One. 5:e11483.
21. Wang C, Liu Z, Woo CW, Li Z, Wang L, Wei JS, Marquez VE, Bates SE, Jin Q, Khan J, Ge K, Thiele CJ. (2012) EZH2 Mediates epigenetic silencing of neuroblastoma suppressor genes CASZ1, CLU, RUNX3, and NGFR. Cancer Res. 72:315-24.
22. 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. 23:660-76.
23. Jene-Sanz A, Váraljai R, Vilkova AV, Khramtsova GF, Khramtsov AI, Olopade OI, Lopez-Bigas N, Benevolenskaya EV. (2013) Expression of polycomb targets predicts breast cancer prognosis. Mol Cell Biol. 33:3951-61.
24. Tiffen JC, Gunatilake D, Gallagher SJ, Gowrishankar K, Heinemann A, Cullinane C, Dutton-Regester K, Pupo GM, Strbenac D, Yang JY, Madore J, Mann GJ, Hayward NK, McArthur GA, Filipp FV, Hersey P. (2015) Targeting activating mutations of EZH2 leads to potent cell growth inhibition in human melanoma by derepression of tumor suppressor genes. Oncotarget. 6:27023-36.
25. Xu F, Liu L, Chang CK, He Q, Wu LY, Zhang Z, Shi WH, Guo J, Zhu Y, Zhao YS, Gu SC, Fei CM, Li X. (2016) Genomic loss of EZH2 leads to epigenetic modifications and overexpression of the HOX gene clusters in myelodysplastic syndrome. Oncotarget. 7:8119-30.
26. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 125:301-13.
27. Zhao XD, Han X, Chew JL, Liu J, Chiu KP, Choo A, Orlov YL, Sung WK, Shahab A, Kuznetsov VA, Bourque G, Oh S, Ruan Y, Ng HH, Wei CL. (2007) Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell. 1:286-98.
28. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE. (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. 4:e1000242
29. Chapman-Rothe N, Curry E, Zeller C, Liber D, Stronach E, Gabra H, Ghaem-Maghami S, Brown R. (2013) Chromatin H3K27me3/H3K4me3 histone marks define gene sets in high-grade serous ovarian cancer that distinguish malignant, tumour-sustaining and chemo-resistant ovarian tumour cells. Oncogene. 32:4586-92.
30. John S, Sabo PJ, Thurman RE, Sung M, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA. (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nature Genetics. 43:264-268.

31. Farzadfard F, Perli SD, Lu TK. (2013) Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth Biol.* 2:604-13.
32. Hu J, Lei Y, Wong WK, Liu S, Lee KC, He X, You W, Zhou R, Guo JT, Chen X, Peng X, Sun H, Huang H, Zhao H, Feng B. (2014) Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Res.* 42:4375-90
33. Young MD, Willson TA, Wakefield MJ, Trounson E, Hilton DJ, Blewitt ME, Oshlack A, Majewski IJ. (2011) ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res.* 39:7415-27
34. Hussein SM, Puri MC, Tonge PD, Benevento M, Corso AJ, Clancy JL, Mosbergen R, Li M, Lee DS, Cloonan N, Wood DL, Munoz J, Middleton R, Korn O, Patel HR, White CA, Shin JY, Gauthier ME, Lê Cao KA, Kim JI, Mar JC, Shakiba N, Ritchie W, Rasko JE, Grimmond SM, Zandstra PW, Wells CA, Preiss T, Seo JS, Heck AJ, Rogers IM, Nagy A. (2014) Genome-wide characterization of the routes to pluripotency. *Nature.* 516:198-206.
35. Hart T, Komori HK, LaMere S, Podshivalova K, Salomon DR. (2013) Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics.* 14:778.
36. Koch F, Fenouil R., Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, et al. (2011) Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural & Molecular Biology,* 18:956–963.
37. Beerli RR, Segal DJ, Dreier B, Barbas CF. (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *PNAS* 95:14628–14633.
38. Vojnic E, Mourão A, Seizl M, Simon B, Wenzek L, Larivière L, Baumli S, Baumgart K, Meisterernst M, Sattler M, Cramer P. Structure and VP16 binding of the Mediator Med25 activator interaction domain. (2011) *Nat Struct Mol Biol.* 18:404-9.
39. Hirai H, Tani T, Kikyo N. (2010) Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *Int J Dev Biol.* 54:1589-96.
40. Malik S, Roeder RG. (2010) The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet.* 11:761-72.
41. 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. *The Journal of Biological Chemistry,* 286:521–529.
42. Thompson MR, Xu D, Williams BR. (2009) ATF3 transcription factor and its emerging roles in immunity and cancer. *J Mol Med,* 87:1053-60.
43. Liu Z, Lam N, Thiele CJ. Zinc finger transcription factor CASZ1 interacts with histones, DNA repair proteins and recruits NuRD complex to regulate gene transcription. (2015) *Oncotarget,* 6:27628-40.
44. Syed V, Mukherjee K, Lyons-Weiler J, Lau KM, Mashima T, Tsuruo T, Ho SM. (2005) Identification of ATF-3, caveolin-1, DLC-1, and NM23-H2 as putative antitumorigenic, progesterone-regulated genes for ovarian cancer cells by gene profiling. *Oncogene.* 24:1774-87.
45. Xie JJ, Xie YM, Chen B, Pan F, Guo JC, Zhao Q, Shen JH, Wu ZY, Wu JY, Xu LY, Li

- EM. (2014) ATF3 functions as a novel tumor suppressor with prognostic significance in esophageal squamous cell carcinoma. *Oncotarget*. 5:8569-82.
46. Liu Z, Rader J, He S, Phung T, Thiele CJ. (2013) CASZ1 inhibits cell cycle progression in neuroblastoma by restoring pRb activity. *Cell Cycle*. 12:2210-8.
47. Langmead B, Salzberg S. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9:357-359.
48. John, S., Sabo, P. J., Thurman, R. E., Sung, M.-H., Biddie, S. C., Johnson, T. A., et al. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genetics*. 43: 264–268.
49. Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, 32.
50. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. (2016) *Nucleic Acids Res*. pii: gkw343.
51. Liu T, Ortiz J, Taing L, Meyer C, Lee B, Zhang Y, Shin H, Wong S, Ma J, Lei Y, et al. (2011) Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biology*. 12: R83.
52. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 7:562-78.