Activation of tumor suppressor genes in breast cancer cells by a synthetic chromatin effector

Kimberly C. Olney[2], David B. Nyer[1], Melissa A. Wilson Sayres[2,3], and Karmella A. Haynes[1]

1. School of Biological and Health Systems Engineering, Arizona State University
2. School of Life Sciences, Arizona State University
3. The Biodesign Institute Center for Evolutionary Medicine, Arizona State University

Running title: Synthetic regulator of breast cancer tumor suppressors

Key words: chromatin, breast cancer, Polycomb, tumor suppressor genes

Financial support: M.A.W.S. and K.C.O. - startup to M.A.W.S from the School of Life Sciences and the Biodesign Institute at Arizona State University; D.B.N. - ADHS 14-082976; K.A.H. - NIH NCI K01 CA188164.

Corresponding author:
Karmella A. Haynes; 501 E Tyler Mall, Box 9709, Tempe, AZ 85287; Phone - 480-965-4636; Fax - 480-727-7624; karmella.haynes@asu.edu

ABSTRACT

Disease states, such as breast cancer, arise from the disruption of chromatin, the central DNA-protein structures that package human genetic material. Mounting evidence from genome-wide studies across cancers show that Polycomb-mediated repression of sets of genes, called Polycomb modules, is strongly linked to a poor prognosis. We developed a synthetic transcriptional activator to release silenced genes from the repressed state. The Polycomb-based Transcription Factor (PcTF) is a synthetic effector that accumulates at methyl-histone marks and regulates hundreds of gene targets, including tumor suppressors. We recently reported the activity of PcTF in bone, blood, and brain sarcoma-derived model cell lines. Here, we expand our investigation of PcTF to three breast cancer-derived cell lines. We expressed PcTF in drug-responsive (MCF-7, BT-474) and nonresponsive triple negative (BT-549) breast cancer cell lines. RNA-seq showed that hundreds of genes were up- or down-regulated by PcTF as early as 24 hours after transfection. BT-549, the triple-negative cancer cell line, showed the highest number of PcTF-activated genes. We demonstrate the anti-cancer potential of PcTF by identifying 15 tumor suppressor genes that are upregulated across the three cell types. The data also provide new mechanistic insights into the relationship between chromatin organization and PcTF-mediated regulation of genes. Our results have exciting implications for cancer treatment with engineered biologics.

INTRODUCTION

Eukaryotic chromosomes are organized as chromatin, a dynamic network of interacting proteins, DNA, and RNA in eukaryotic nuclei. These interactions regulate gene transcription and coordinate distinct, genome-wide expression profiles in different cell types. Chromatin mediates epigenetic inheritance [1,2] by regulating expression states that persist through cellular mitosis and in generations of sexually reproducing organisms [3,4]. Trimethylation of histone H3 - a component of the nucleosome protein octamer core, the fundamental subunit of chromatin - at lysine 27 (H3K27me3) plays a central role in the epigenetic regulation of genes that control cell differentiation [5,6]. Several landmark studies have revealed that hyperactivity of the histone-methyltransferase enhancer of zeste 1 and 2 (EZH1, EZH2), which generates H3K27me3, is a feature shared by many types of cancer (recently reviewed in [7]). In breast cancer, elevated EZH2 has been linked to cell proliferation and metastasis [8,9] and a poor prognosis for breast cancer patients [10–13]. In stem cells and cancer cells, EZH2 generates the repressive H3K27me3 mark at nucleosomes near the promoters of developmental genes to prevent differentiation and maintain the proliferative state in stem cells or to generate neoplasia in cancer (reviewed in [5]). Polycomb Repressive Complex 1

(PRC1) binds to the H3K27me3 mark through the polycomb chromodomain (PCD) motif of the CBX protein to stabilize the repressed state. Silencing is reinforced by other chromatin regulators including histone deacetylase (HDAC) and DNA methyltransferase (DMT) [14].

The PRC module, a group of genes that is silenced by H3K27me3 and Polycomb transcriptional regulators [15,16], is a high priority for cancer research and the development of epigenetic drugs. Epigenetic therapy targets aberrant chromatin within cancer cells. This approach overcomes the problem of resistance to hormone therapy, which requires the presence of specific transmembrane receptors on the surfaces of cancer cells. Relatively high expression or upregulation of PRC module genes is associated with a non-proliferative state, cell adhesion, organ development, and normal anatomical structure morphogenesis [15]. Knockdown (depletion) of chromatin proteins (reviewed in [16,17]) and inhibition of Polycomb proteins with low molecular weight compounds and peptides [18–20] stimulates expression of developmental genes and perturbs cancer-associated cell behavior. The success of epigenetic interventions in clinical trials [21,22] demonstrates that mis-regulated chromatin is a druggable target in cancer. Basic research has revealed certain limitations for epigenetic inhibitor compounds. Inhibitors indirectly activate silenced genes by blocking repressors, generate incomplete conversion of silenced chromatin into active chromatin [23,24], interact with off-target proteins outside of the nucleus [25], and do not affect resistant Polycomb protein mutants [26–28]. These limitations can be addressed with alternative molecular technologies.

Biologics, therapies composed of macromolecules such as proteins, are richer in biochemical information compared to low molecular weight compounds. Until recently, none have been designed to decode epigenetic information in cancer cells. To this end, we developed the Polycomb-based Transcription Factor (PcTF), which binds H3K27me3 [29] and recruits endogenous transcription factors to PRC-silenced genes. In bone, brain, and blood-cancer derived cell lines, PcTF expression stimulates transcriptional activation of several anti-oncogenesis genes [30]. PcTF-mediated activation leads to the eventual loss of the silencing mark H3K27me3 and elevation of the active mark H3K4me3 at the tumor suppressor locus *CASZ1*. To advance PcTF towards medical translation, we sought to investigate the behavior of this protein in breast cancer cells lines that have been established as models for tumorigenesis [31–33].

Here, we extend our investigation of PcTF target genes to three breast cancer-relevant cell lines. First, we investigated the transcription profiles of predicted PRC module genes in drug-responsive (MCF-7, BT-474) and nonresponsive triple negative (BT-549) breast cancer cell lines. Receptor-negative BT-549 cells have a transcription profile and histology similar to aggressive tumor

cells from patient samples [34,35]. We show that the transcription profiles of the untreated breast cancer cells are distinct from MCF10A, a breast tissue-derived control cell line. We also show that predicted Polycomb-regulated genes are repressed in the breast cancer cells compared to MCF10A. Over expression of PcTF in transfected breast cancer cells led to the upregulation of dozens of genes, including many predicted PRC module genes and 15 well-characterized tumor suppressor genes, as early as 24 hours after transfection. The transcriptome of BT-549 (triple-negative) showed the highest degree of PcTF-sensitivity. Our results also provide new mechanistic insights into the relationship between chromatin structure and activation of genes by an artificial regulator. We observed that PcTF-sensitive genes are enriched for silencing marks and low levels of activation-associated chromatin marks, suggesting that PcTF regulates genes that are poised for activation.

RESULTS

Differential regulation of genes in breast cancer cell lines

To determine expression levels of predicted PRC module genes, we profiled the transcriptomes of three breast cancer cell lines and the non-invasive, basal B cell line MCF10A [36,37] using next-generation deep sequencing of total RNA (RNA-seq). MCF7, BT-474, and BT-549 represent luminal A, luminal B, and basal B subtypes of breast cancer, respectively (Table 1) [31]. Previous studies have shown that gene expression profiles distinguish two major categories of cancer cell lines, luminal and basal, in patient-derived samples [38,39]. The basal class exhibits a stem-cell like expression profile [40], which is consistent with high levels of Polycomb-mediated repression at genes involved in development and differentiation [41,42]. Levels of the repressor protein EZH2 and the histone modification that it generates (H3K27me3) are elevated in MCF7, BT-474, and BT-549 compared to non-metastatic cells such as MCF10A (Table 1). A mechanistic link between Polycomb-mediated repression and tumor aggressiveness has been supported by a study where stimulation of the phosphoinositide 3-kinase (PI3K) signaling pathway, which induces a metastatic phenotype in MCF10A, is accompanied by increased H3K27me3 at several target genes [43,44]. We hypothesized that known Polycomb-repressed genes (the PRC module) would be down-regulated in the cancerous cell lines compared to MCF10A.

Cell line	ATCC	Sub-type	Markers [31]	EZH2	H3K27me3
MCF7	HTB-22	Luminal A	ER+, PR+	Elevated[a,b,c] [45–47]	Elevated[a] [43,46]
BT-474	HTB-20	Luminal B	ER+, PR+, HER2+	Elevated[c] [48]	Elevated[d] [43]
BT-549	HTB-122	Basal B, claudin-low	ER-, PR-, *TP53*^{M}	Elevated[c] [8]	Elevated[d] [43]
MCF10A	CRL-10317	Non-invasive/ Basal B	ER-, PR-	n/a	n/a

Table 1. Descriptions of the breast tissue-derived cell lines used in this study. ATCC = American Tissue Culture Center ID. Molecular subtype and marker expression status are from Neve et. al 2006 [31]: Estrogen receptor presence or absence (ER+/-), Progesterone receptor presence or absence (PR+/-), HER2 overexpression (HER2+), and TP53 mutation (*TP53*^{M}). EZH2 and H3K27me3 were shown to be elevated compared to non-metastatic fibroblasts (a) [46], LNCaP (b) [45], MCF10A (c) [8,47,48], and HMEC (d).

Comparison of the expression profiles in untreated cells showed that the three breast cancer model cell lines were transcriptionally dissimilar to the control cell line MCF10A and that BT-549 and MCF7 were more similar to each other than either were to BT-474. Expression levels (FPKM values) across 63,286 gene protein coding transcripts (GRCh38 reference genome) were used to calculate Jensen-Shannon Divergence (JSD) (Methods and Fig. 1A). JSD values correspond to the similarity of the probability distributions of transcript levels for two RNA-seq experiments. Expression values for biological replicates showed the highest similarities (smallest distances) within cell types (Fig 1A, upper grid). The largest distances were observed between MCF10A and the three cancer cell types: 0.461 for BT-549, 0.476 for MCF7, and 0.511 for BT-474 (Fig 1A, lower grid). A similarly high JS distance was observed for BT-549 versus BT-474 (JSD = 0.464), suggesting that these cancer cell lines are transcriptionally distinct. BT-549 and MCF7 showed the highest similarity, with a cumulative JSD of 0.357. This observation contrasts with other reports where BT-549 and MCF7 are described as transcriptionally and phenotypically different [36,49]. Differences in transcription profiling methods, RNA-seq used here and the DNA oligomer microarray chip used by others, may underlie the different outcomes.

Differential expression between cell lines for individual genes (Fig. S1) followed similar trends as those observed for the global JSD analysis. We used an expression comparison algorithm (Cuffdiff [50]) to identify genes that were differentially expressed (2-fold or greater difference in expression, q value ≤ 0.05) or similarly expressed (less than 2-fold difference, q value ≤ 0.05) between cell types. Comparisons that included MCF10A showed the highest numbers of differentially-expressed genes,

as well as the lowest numbers of similarly expressed genes. This result further supports transcriptional differences between the cancerous cell lines and MCF10A (Fig. S1).

Next, we determined expression levels within groups of predicted PRC-regulated genes and observed that expression within these subsets is lower in the three cancer cell types than in MCF10A. We used data from other breast cancer cell line studies of MCF7 and MDA-MB-231 to classify a subset of PRC target genes based on H3K27me3 enrichment or binding of EZH2, an enzyme that generates the H3K27me3 mark (see Methods). Only 245 gene IDs were shared between the H3K27me3 and EZH2 subsets. Although these two groups are mostly distinct, both showed low median expression values (FPKM < 2), which suggests epigenetic repression (Fig. 1B). This result is consistent with the roles of H3K27me3 and EZH2 in gene silencing. Median expression levels of predicted PRC module genes were reduced in the cancer cell lines compared to the non-cancer cell line. The H3K27me3-marked subset showed median log10(FPKM) values for BT-474 (-1.66), MCF7 (-1.16), and BT-549 (-1.15) that were slightly lower than MCF10A (-1.10) (Fig. 1B, middle plot). The median FPKM values for EZH2 targets were dramatically lower (zero signal) in the cancer cell lines, while the median value was higher (-1.65) for MCF10A (Fig. 1B, right). This result suggests that EZH2 enrichment in MDA-MB-231 is a better predictor for strong epigenetic repression in breast cancer cell lines.

To determine whether individual predicted PRC target genes were similarly regulated across cell lines, we compared two groups of genes that were categorized by expression level: silenced (FPKM < 2) [51,52] or expressed (FPKM ≥ 2) (Fig. 1C). In each cell type, genes with silenced expression levels included 70.2% - 79.3% of the H3K27me3-marked loci and 78.4% - 82.2% of the EZH2-enriched loci. Roughly a quarter of the genes (17.8% - 29.8%) showed some expression (FPKM > 2) and only 16.7% - 8.2% were expressed at FPKM ≥ 10. Although the majority of the genes were similarly expressed or silenced in the cancer lines and in MCF10A, several genes were uniquely regulated (Fig. 1C non-overlapping regions). We hypothesized that the silenced genes would become activated in the presence of the synthetic regulator PcTF which recognizes the H3K27me3 mark.

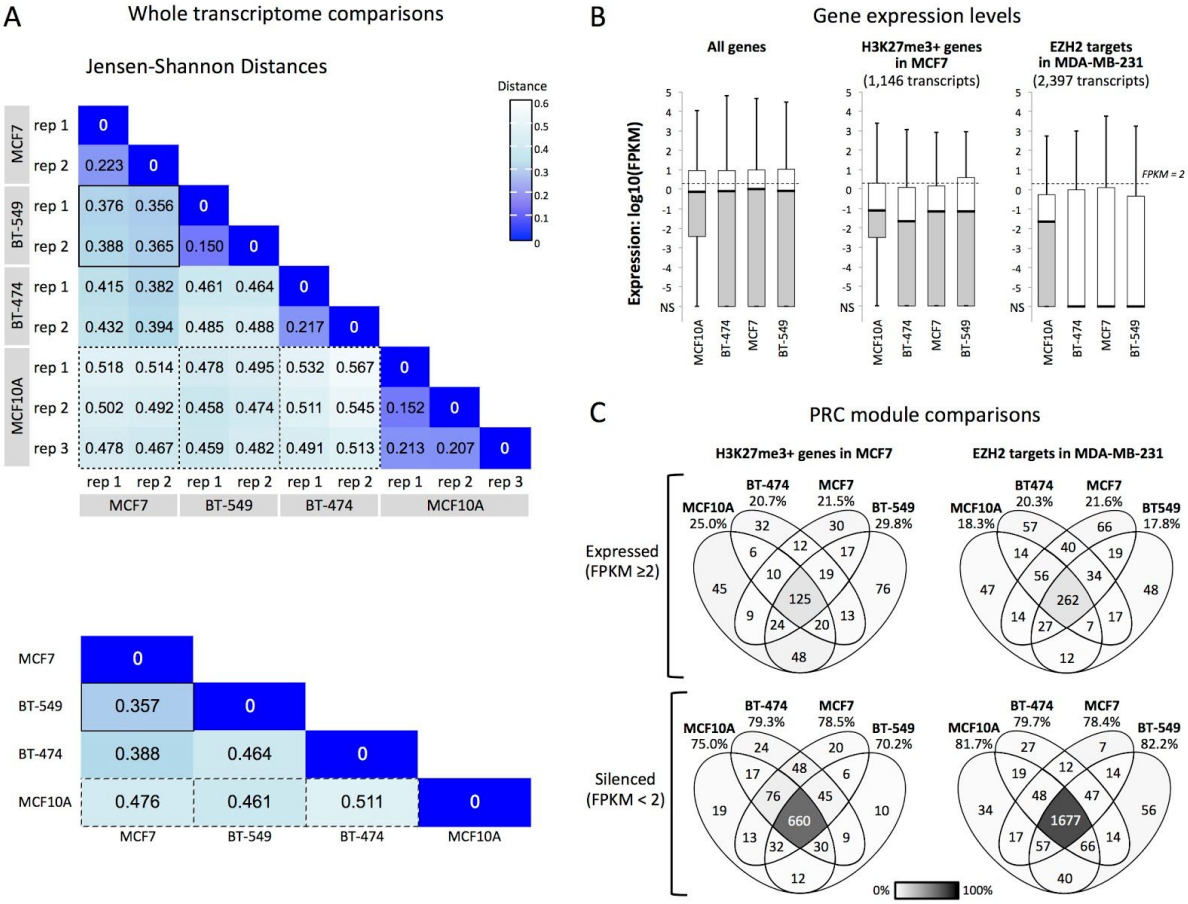

Figure 1. Comparisons of transcription profiles of three model breast cancer lines (MCF7, BT-549, BT-474) and a control non-cancer line (MCF10A). (A) Jensen-Shannon Divergence (JSD) values were calculated as the similarity of the probability distributions of expression levels (FPKM values) for 63,286 total transcripts, which include 22,268 protein-coding transcripts. Each square is a pairwise comparison of RNA-seq biological replicates. In the lower grid, we used cummeRbund [53] to consolidate replicates and to calculate overall JSD between cell types. Solid border = BT-549 vs. MCF7, smallest JSD; dashed border = JSD's for MCF10A vs. cancer cell lines. (B) The boxplots show log10(FPKM) values for all protein-coding transcripts (22,268), H3K27me3-positive (1,146) or EZH2-positive (2,397) protein-coding loci. NS = no signal. (C) The Venn diagrams include transcripts from panel B. Genes are categorized by expression level: expressed (FPKM ≥ 2) or silenced (FPKM < 2). The percentages of genes in each category (expressed or silenced) is shown under the label for each cell line.

PcTF-responsive genes include PRC module genes and other loci

We investigated PcTF-mediated gene regulation in the three breast cancer cell lines by profiling the transcriptomes of PcTF-expressing cells (Fig. 2). We transfected cells with PcTF-encoding plasmid DNA (previously described [30]) via Lipofectamine LTX and and allowed transfected cells to grow for 24, 48, and 72 hours before extracting total RNA for sequencing. RNA-seq reads were aligned to a

human reference genome GRCh38 that included the coding region for PcTF (see Methods). No reads aligned to the PcTF coding sequence in control, untransfected cells. In the transfected cells, PcTF expression levels were highest at 24 hours and decreased 1.6 to 5.5-fold every 24 hours (Fig. 2A). We observed a similar trend with other cancer cell lines in a previous study [30]. One outlier sample, a replicate for BT-474 cells expressing PcTF for 48 hours, had a markedly different PcTF expression level (Fig. 2A) and genome-wide transcription profile (Fig. S2) and was therefore omitted from further analyses. Different subsets of genes were up- or down-regulated at least two fold (q value ≤ 0.05) early, late, or across all time points during PcTF expression (Fig. 2B). Of the genes that showed at least a two-fold change in either direction, the vast majority were up-regulated (Fig. 2B). Genes that were significantly up- or down-regulated at least 2-fold at all three time points (Fig. 2C) or at one or two time points (Fig. S3) included several PRC module genes. Genes that were not identified as PRC-repressed might be downstream targets of direct PcTF targets. It is also possible that the ChIP-seq data did not identify all Polycomb-regulated genes. Overall, our results are consistent with PcTF's function as an activator that targets H3K27me3-enriched genes.

Nineteen genes were upregulated at least 2-fold (q value ≤ 0.05) at all time points in all three cell lines (Fig. 2C). These genes are part of the interferon and innate immunity pathway: *C19orf66*, *DDX58, DTX3L, HERC6, IFI27, IFI44L, IFI6, IFIH1, ISG15, LGALS3BP, MX1, OAS1, OAS3, PARP9, PARP14, PLSCR1, SP100, UBE2L6*, and *XAF1*. The most significantly enriched GO terms for this set include defense response to virus, negative regulation of viral life cycle, response to stimulus, and immune system process (Fig. S4). Four of these genes (*IFI6, MX1, OAS3*, and *UBE2L6*) may have become activated as a nonspecific response to foreign plasmid DNA and RNA, as previously reported by others [54–58]. These genes may also be direct targets of PcTF. *MX1*, *HERC6*, and *UBE26L* belong to the PRC module groups shown in Figure 1 (panels B and C). Other studies have linked high levels of expression from interferon pathway genes with a non-cancerous phenotype. In breast cancer, an immune response gene-expressing subgroup, which includes *ISG15*, *MX1*, and other interferon genes, has been associated with improved prognosis in triple negative breast cancers [59,60]. Therefore, our results suggest that PcTF shifts the transcription profiles of breast cancer cells towards an anti-cancer state. Our discovery of 19 commonly upregulated genes indicates that diverse cancer subtypes can be similarly affected by a single synthetic transcriptional regulator.

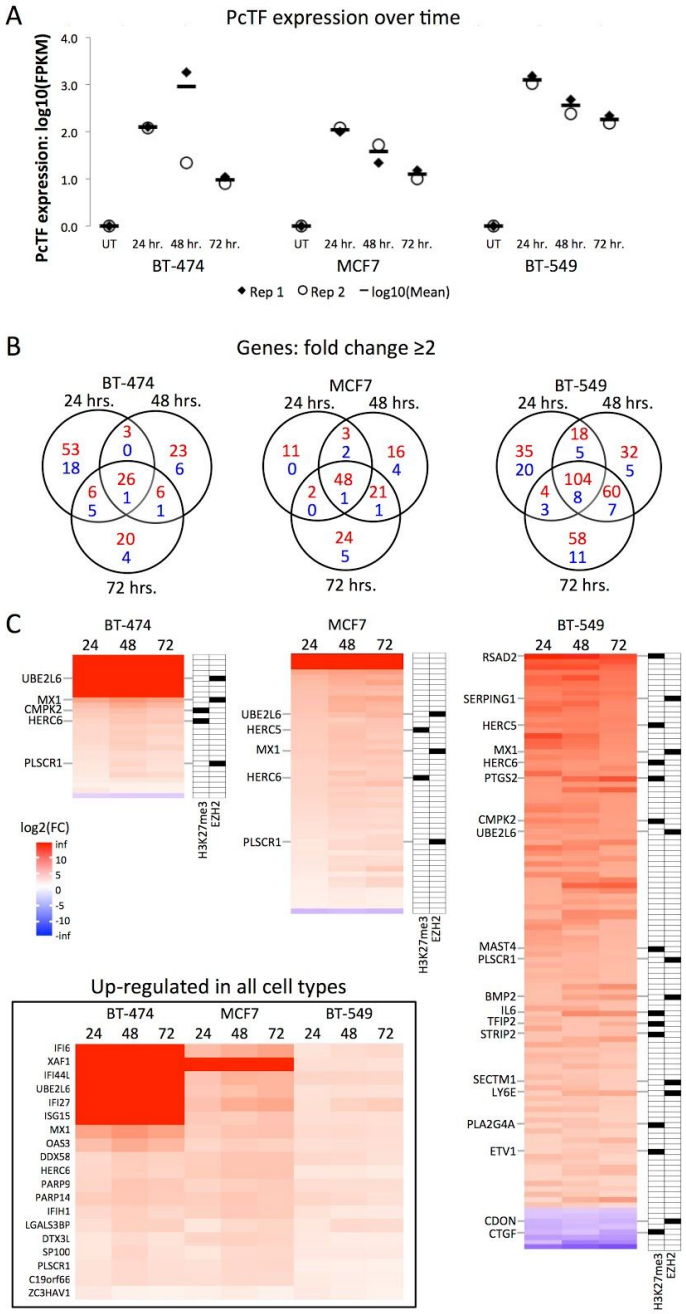

Figure 2. PcTF-expressing breast tissue-derived cell lines show upregulation of PRC-repressed genes and other genes. (A) The chart shows log10(FPKM) of PcTF for untransfected cells (UT) and at 24, 48, and 72 hours following transfection of BT-474, MCF7, and BT-549. The BT-474 outlier (48 hrs. rep1) was omitted from subsequent analyses. (B) The Venn diagrams show genes with expression levels that changed at least 2-fold in either direction (q value ≤ 0.05) in PcTF-expressing cells versus untransfected cells. Red = up-regulated, blue = down-regulated. (C) The heat maps show fold-change (log2(FC)) values for subsets of genes that significantly changed (q ≤ 0.05) at all three time points (center regions of the Venn diagrams in B). PRC module genes from Figure 1B, C are labeled. Filled boxes to the right of each heat map indicate H3K27me3 enrichment in MCF7 and EZH2 targets in MDA-MB-231. The inset shows 19 genes that were upregulated across all time points in all

three cell lines. Data for the other genes, where log2(FC) >= 2 and $q \leq 0.05$ at one or two time points, are shown in Figure S3.

Genes become upregulated over time in the presence of stable PcTF levels

We sought to determine the dynamics of transcriptional regulation at PcTF-regulated genes. Using transient transfections, we established that PcTF-mediated activation of genes could be detected over background at multiple time points. However, in this experiment PcTF levels decreased over time (Fig. 2A), which prevents us from distinguishing time- versus dose-dependent effects on gene regulation. Therefore, we constructed stable transgenic MCF7 cell lines to enable constant expression of the fusion protein over time. Expression of PcTF or a control fusion protein that lacks the histone-binding domain ($Pc_{\Delta}TF$) was placed under the control of the rtTA activator, which binds to the *pTet* promoter in the presence of doxycycline (dox) (Fig. 3A). Expression of rtTA was indicated by constitutive GFP expression, and inducible nuclear localization sequence-tagged PcTF was detected as a nuclear RFP signal in the presence of doxycycline (Fig. 3B). We used quantitative reverse transcription PCR (qRT-PCR) to measure the expression levels of PcTF and a subset of PcTF-sensitive genes that were identified in the RNA-seq experiment (Fig. 2).

QRT-PCR using a universal mCherry-specific primer set confirmed that PcTF expression levels decreased over time in transiently transfected cells (Fig. 3C) as observed in the RNA-seq experiment (Fig. 2A). The stable transgenic cells showed low levels of fusion protein mRNA in the initial uninduced (dox-minus) state compared to untransfected MCF7 cells. Exposure to 1 µg/mL dox increased PcTF and $Pc_{\Delta}TF$ levels by an order of magnitude. These levels were slightly higher than the PcTF expression levels observed in transiently transfected cells at the 72-hour time point, and remained relatively constant over time. Fold-change (compared to untransfected cells) remained within values of 67 - 192 at 24, 48, and 72 hours.

For qRT-PCR analysis of PcTF target genes, we were able to design and validate specific assays for a subset of the PcTF-induced genes that were identified in the RNA-seq experiments. We selected transcripts that represented a range of basal expression levels, from low to high FPKM values, in the untreated state: *XAF1* (2.9E-03), *SAMD9L* (0.03), *GBP1* (0.31), *CEACAM1* (0.96), *CASP14* (1.14), and *SP100* (14.57). *XAF1* was the most strongly upregulated across all three time points (18 to 36-fold). This result agreed with the high fold-change values we observed in the RNA-seq experiment for transiently transfected cells (Fig. 2C). The other five genes showed slight upregulation (1.4 to 1.8-fold) in response to dox-induced PcTF expression. The relatively weak response of these genes compared to *XAF1* could be explained by a smaller dynamic range, where

there is little difference between the basal versus activated expression level. Furthermore, these genes may have been slightly upregulated prior to dox treatment since PcTF was detected at low levels before induction (Fig. 3C).

In all cases except for *SP100*, the truncated fusion protein $Pc_{\Delta}TF$ did not upregulate the target genes over the three time points, showing that the H3K27me3-binding domain (PCD) was required for PcTF-mediated regulation. At the 24 hour time point, *XAF1*, *GBP1*, and *CEACAM1* became up-regulated in truncation-expressing cells, suggesting an initial nonspecific response to transgene activation. At 48 and 72 hours, gene expression decreased in the presence of $Pc_{\Delta}TF$, but remained upregulated in the presence of PcTF. Overall, these results demonstrate that target genes are upregulated at steady levels of PcTF expression. Maintenance of this activated state requires the H3K27me3-binding PCD.

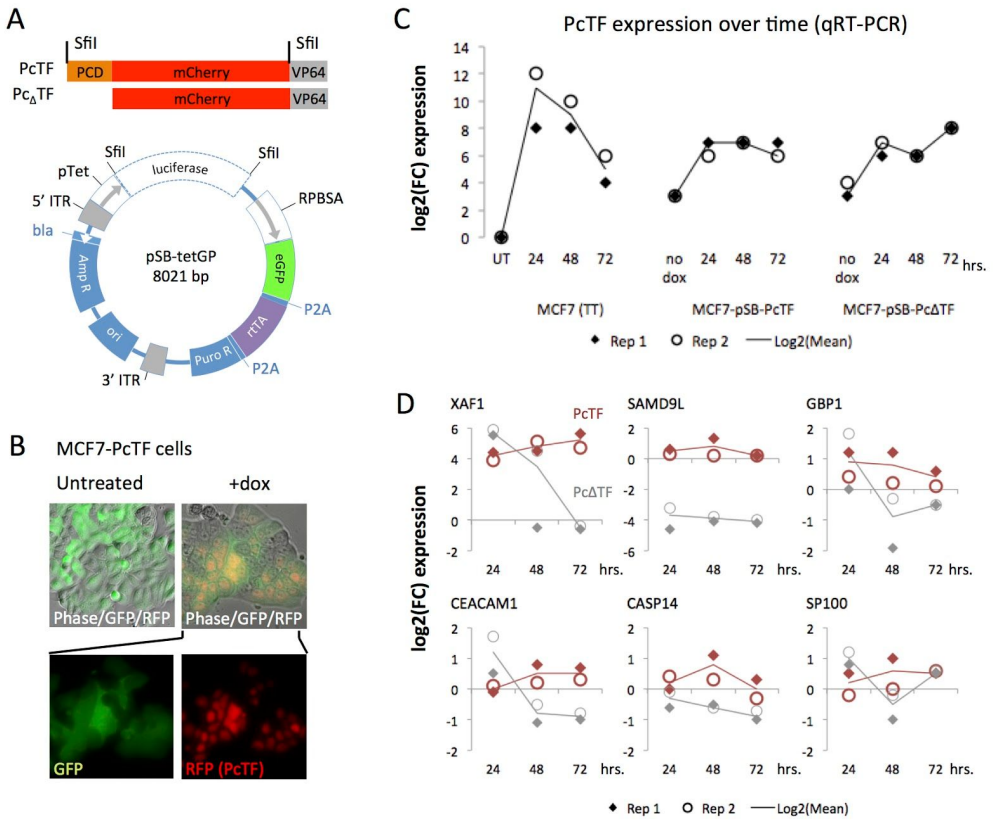

Figure 3. QRT-PCR analysis of gene expression in stable, transgenic PcTF-expressing cells. (A) *SfiI*-flanked PcTF or $Pc_{\Delta}TF$ constructs (top) were cloned into the pSB-tetGP expression vector (bottom), resulting in the replacement of the *luciferase* reporter with fusion protein ORFs. (B) Fluorescence microscopy of the MCF7-PcTF transgenic cell line. (C) Time course qRT-PCR for PcTF. (D) Time course qRT-PCR for select genes. For all qRT-PCR experiments n = two cDNA libraries from independent transfections or dox treatments. FC = fold change relative to "no dox" controls, calculated as double delta C_p (see Methods).

PcTF-sensitive genes are located in chromatin regions that contain both silencing and activation-associated marks

To investigate the contribution of local chromatin states to PcTF-mediated gene regulation, we compared chromatin modifications at several loci with the expression profiles of the corresponding genes. Previously, we showed that *CASZ1*, a direct target of PcTF, was enriched for H3K27me3 up to 10 kb upstream of the promoter and showed an immediate (10 - 24 hours) response to PcTF in a U-2 OS model cell line [30]. Here, we utilized the extensive public ChIP-seq data that is available for MCF7 to investigate chromatin modifications at PcTF-responsive genes. In untreated MCF7 cells, *CASZ1* was expressed at a moderate level (FPKM = 9.75) and became upregulated up to 1.8-fold after 72 hours of PcTF expression (Fig. 4A). *CASZ1* was upregulated over 2-fold across all time points in BT-549, but not in BT-474 where *CASZ1* had the highest basal expression level. In MCF7, enrichment of H3K27me3 and H3K9me3 within 10 kb of the transcription start site of *CASZ1* was lower than at the repressed locus *FCAR* (FPKM = 0, signal below detection limits) (Fig. 4A). Higher levels of H3K27ac and H3K4me3 at CASZ1 suggest an active transcriptional state, but RNA PolII was depleted compared the active *HSPA8* gene (FPKM = 298.88). In MCF7, *CASZ1* may represent a down-regulated, but not strongly repressed, gene that responds slightly to PcTF regulation.

Next, we investigated the chromatin states of the 48 genes that became significantly upregulated (FC > 2, $q \leq 0.05$) by PcTF in MCF7 across all time points (RNA-seq analysis, Fig. 2B, C). Sixteen of these genes showed low basal expression (FPKM < 2) and 32 showed higher basal expression (FPKM range = 2.06 to 63.88, mean = 9.21, median = 14.08) in untreated cells. We labeled these sets as Class 1 and Class 2 genes, respectively. We compared chromatin modifications of the 48 PcTF-responsive genes to chromatin profiles of fifty repressed (Class 3) and fifty active (Class 4) genes. The active gene classes (2 and 4) were relatively depleted for H3K27me3, which is consistent with the role of H3K27me3 in silencing. Enrichment for H3K27me3, the natural target of the N-terminal PCD motif in PcTF, was highest for the Class 1 PcTF-responsive genes and the Class 3 silenced, non-responsive genes (Fig. 3B). These data provide further evidence that PcTF acts upon a select subset of H3K27me3-enriched loci, as we previously observed in other cell lines [30]. Class 2 PcTF-responsive genes generally lack the H3K27 methylation mark and may represent downstream targets of the products expressed from direct PcTF targets.

To investigate potential off-target binding with a similar histone modification, we analyzed enrichment profiles for the mark H3K9me3. PCD peptides have shown some affinity for both H3K9me3 and H3K27me3 peptide tails *in vitro* [61,62]. The gene subsets showed no significant differences in mean H3K9me3 enrichment (Fig. 4B), a modification that is frequently found at

constitutive pericentric heterochromatin and non-coding DNA [63–65]. We observed that PcTF-responsive genes tended to be distributed along chromosome arms rather than concentrated near centromeres (Fig. S4). PcTF target sites coincide more closely with the distribution of facultative chromatin and epigenetically-regulated cell development genes [41,66], which are characterized by H3K27me3 enrichment in specific cell types.

Enrichment for two active gene-associated marks, H3K27ac and H3K4me3, at PcTF-responsive genes (Classes 1 and 2) were at intermediate levels between the silenced and active PcTF non-responsive genes (Classes 3 and 4). RNA PolII was relatively depleted at PcTF-responsive genes compared to the active, non-responsive genes. From these findings, we conclude that strongly repressed genes are resistant to PcTF-mediated activation and that an intermediate regulatory state, where silent and active marks are present, supports PcTF activity.

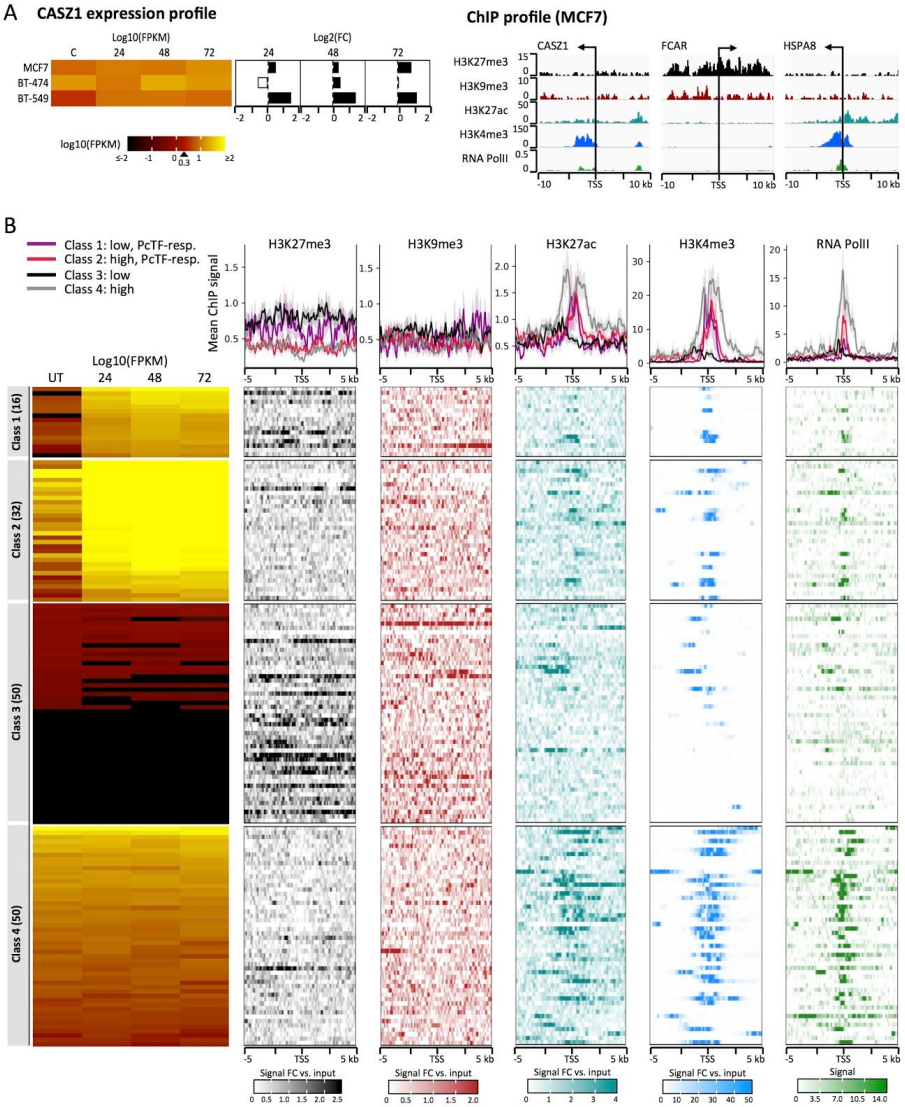

Figure 4. Expression profiles of PcTF-responsive and nonresponsive genes in MCF7. (A) Expression profile of *CASZ1* and ChIP-seq signals within +/- 10 kb of the transcription start sites of *CASZ1*, a silenced gene (*FCAR*), and an active gene (*HSPA8*). (B) Expression profiles of genes where fold-change (FC, PcTF-treated vs. UT, untreated control) is greater than two at all time points. Within each class, genes are sorted from highest FC to lowest. Numbers of genes are indicated in parentheses.

Tumor suppressor and BRCA pathway genes become upregulated in PcTF-expressing cells

Next, we explored the clinical potential of PcTF-mediated transcriptional regulation by determining the representation of known tumor suppressor genes amongst PcTF-responsive loci. For this analysis we used a tumor suppressor gene set that includes 983 candidate anti-cancer targets that are down-regulated in tumor samples (Methods). Of these, 589 include BRCA human tumor suppressor genes (TSGs) that are repressed in invasive carcinoma samples compared to normal tissue samples [67,68]. The genes were classified as tumor suppressors based on text-mining of cancer research literature, and manual assessment of relevant cancer types and molecular pathways (TSGene 2.0) [67,68].

We identified a subset of genes that showed a minimum FPKM of 2 in treated cells and at least 2-fold upregulation ($q \leq 0.05$) in at least one of the three breast cancer cell lines (BT-474, MCF7, BT-549) across all three time points. To identify TSGs that are upregulated in response to PcTF, we compared the upregulated subset to the 983 candidate anti-cancer genes identified by TSGene 2.0. Fifteen of the 983 TSGs were upregulated across all three time points in at least one of the cell lines (Fig. 5B). Information genecards.org [69] further validated the association of these 15 genes with tumor suppressor activity. Of the fifteen upregulated TSGs, seven belong to the breast cancer susceptibility (BRCA) pathway: *CDKN1A*, *PML*, *ANGPTL4*, *CEACAM1*, *BMP2*, *SP100*, *TFPI2*.

Cell line comparisons of RNA-seq FPKM values for the fifteen tumor suppressor genes showed that median expression was lower in untreated BT-474 and MCF7 than in the non-cancerous MCF10A cell line (Fig. 5A). This result is consistent with the idea that epigenetic repression of TSGs supports a cancerous cell phenotype. In PcTF-expressing cells, the median expression of the fifteen tumor suppressor genes was increased at all time points compared to the untreated samples for each cancer cell line (Fig. 5A). Interestingly, the median FPKM value for the 15 TSGs was higher in BT-549 than in MCF10A. We examined the expression levels of the individual genes and found that *BMP2*, *CEACAM1*, *CDKN1A*, *DSP* are lower in BT-549, as well as BT-474 and MCF7, than in MCF10A (Fig. 5B). These genes become upregulated in PcTF-expressing cells.

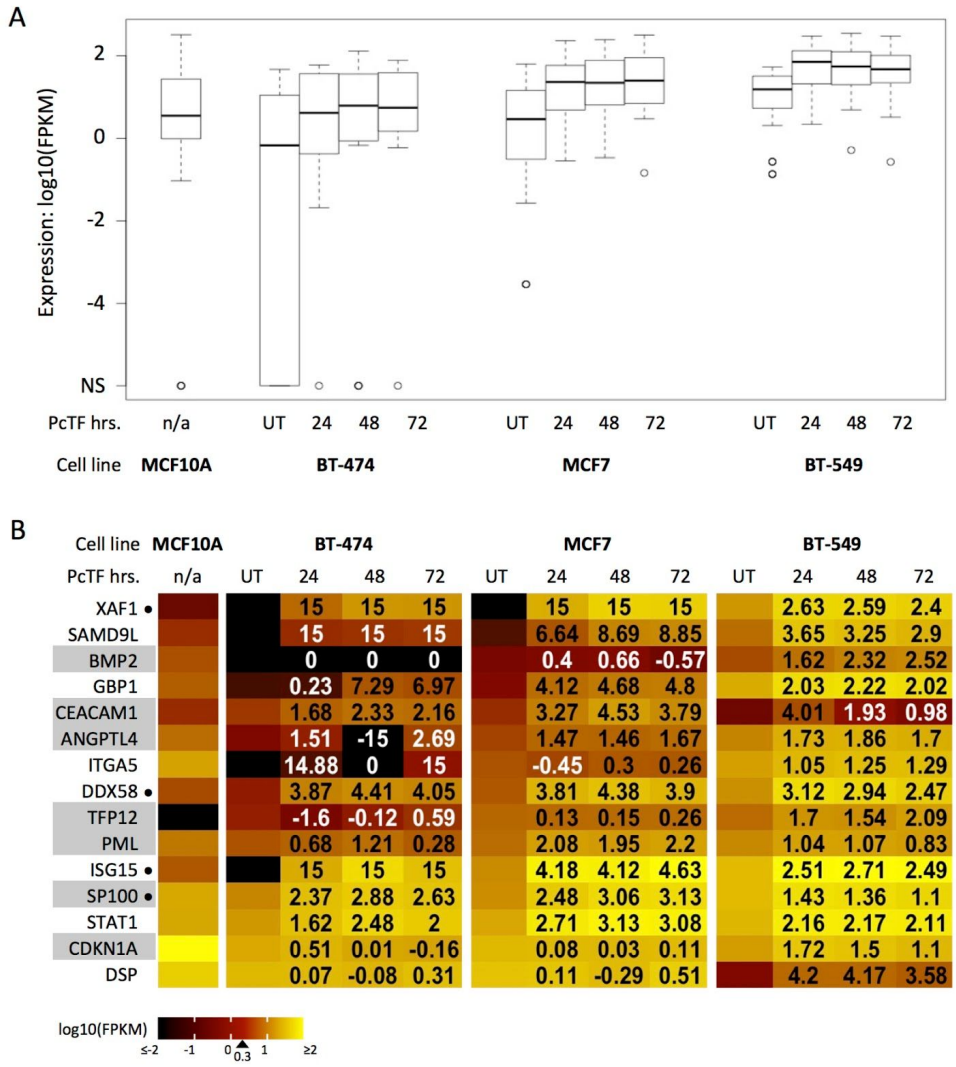

Figure 5. Tumor suppressor genes show increased expression in PcTF-expressing cancer cell lines. (A) Mean upregulation across three time points (24, 48, and 72 hours) is shown for fifteen tumor suppressor genes that show at least two-fold upregulation ($q \leq 0.05$) relative to the untreated control (UT) in at least one of the cell lines. (B) Individual log10(FPKM) (color scale) for each of the tumor suppressor genes in A. BRCA pathway genes are shaded in grey. ● = FC ≥ 2, $q \leq 0.05$ at all time points in all cell lines. Data for all cell types and conditions are shown. Genes are sorted from lowest to highest expression in untreated MCF7 cells. Numbers in the PcTF-treatment columns show log2 fold change values compared to UT. 15 = infinite positive fold change, where no expression was detected in untreated cells, -15 = infinite negative fold-change, where no expression was detected in treated cells.

CONCLUSIONS & DISCUSSION

We have demonstrated that PcTF, a synthetic transcription factor that is designed to recognize H3K27me3, leads to broad changes in the transcription profiles of cell lines that represent different breast cancer subtypes. We hypothesized that genes enriched for Polycomb-associated chromatin marks would become up-regulated by PcTF. While many H3K27me3-enriched genes were upregulated in MCF7, many were non-responsive. Several genes that were not identified as Polycomb-enriched became up-regulated by PcTF. Enrichment of H3K9me3 did not distinguish PcTF-responsive genes, therefore we can rule out off-target binding which has been observed for PCD peptides in biochemical studies [61,62]. At PcTF-responsive genes, levels of H3K4me3 and H3K27ac were higher than at silenced non-responsive genes, but RNA PolII was depleted compared to active non-responsive genes. Therefore, the chromatin at PcTF-responsive genes may support a low or intermediate expression state. Berrozpe et al. recently reported that Polycomb complexes preferentially accumulate at weakly expressed genes rather than strongly silenced or highly expressed genes [70]. In our experiments, specific PRC-regulated genes may have been expressed at low to intermediate levels and then further upregulated upon exposure to PcTF. Our analysis of PcTF-regulated genes and chromatin states paves the way for future studies to further resolve chromatin features that distinguish regulatable PRC-repressed genes in cancer cells.

Deregulation of histone-modifying enzymes contributes to epigenomic diversity in breast cancer cells [13]. Other factors can also contribute to transcription profile variations, such as differences in the abundance of wild type or mutated transcription factors, or mutations that impact the stability and turnover of RNA transcripts. Past work has begun to illuminate the relationship between phenotypic subclasses and transcription profiles [15,36,49,71]. Such investigations help to elucidate cancer mechanisms and drug targets for more effective treatments. However, the link between transcriptome and phenotype is not entirely straight-forward. We observed that the transcription profile of BT-549 (invasive basal B) is more similar to MCF7 (luminal) than either were to BT-474 (luminal). In contrast, other reports have shown clear distinctions between the transcription profiles and phenotypes of BT-549 and MCF7 [36,49]. Differences in transcript profiling methods, our RNA-seq and JSD analysis versus the DNA oligomer arrays used by others, may account for this conflicting result. Further, we acknowledge that the JSD may be driven by a few genes with high expression and high variance, which could account for some of the patterns. Cancer cell transcriptome diversity poses a formidable challenge for the development of drugs that pinpoint specific genes and pathways. The results reported here demonstrate that PcTF co-regulates cohorts of genes, many of which are associated with anticancer functions, in diverse model breast cancer cell lines with different basal

gene expression levels. BT-549 showed higher basal expression levels for many genes, including predicted PRC module genes, that became upregulated by PcTF. In contrast, the same genes were strongly repressed in BT-474. In spite of this difference, several of these genes became activated in the presence of PcTF.

PcTF represents a new class of biologic, a medicinal protein-based macromolecule. So far, low molecular weight compounds, rather than biologics, are the predominant method for epigenetic intervention. Their ease of delivery, orally or intravenously, make these compounds a very attractive approach for cancer treatment. However, small compounds have a very limited range of biological activity, e.g. as ligands for specific proteins, compared to macromolecules. Transgenic and synthetic transcription factors expand the repertoire of epigenetic drug activity by allowing selective control of therapeutic genes in cancer cells [72–75]. Protein expression often relies on inefficient and possibly mutagenic nucleic acid delivery, which poses a significant barrier for many potential biologics. Recent advances in large molecule carriers such as cell penetrating peptides [76–78] provide a positive outlook for cellular delivery of purified proteins. We have recently shown that PcTF has affinity for histone H3K27me3 peptides *in vitro* [29]. Here we demonstrate the potential utility of PcTF in additional cancer types. It will be eventually important to determine if cell-penetrating PcTF proteins meet or exceed the efficacy of small molecule epigenetic drugs in tumor models. At present, PcTF shows promise for activating tumor suppressor genes and provides a potential alternative to traditional tumor therapies.

MATERIALS AND METHODS

DNA constructs

Plasmids were constructed to express fusion proteins either constitutively or in the presence of doxycycline. The plasmid for constitutive expression of PcTF, hPCD-TF_MV2 (KAH126), was constructed as previously described [79]. The doxycycline-inducible transgene PcTF_pSB-GPtet was constructed by ligating 50 ng of PCR amplified, SfiI-digested PcTF fragment with a SfiI-linearized pSB-GPtet vector [80] (Addgene #60495) at a ratio of 5 insert to 1 vector in a 10 uL reaction (1 uL 10x buffer, 1 uL T4 ligase). The same procedure was used to build constructs for dox-inducible $Pc_{\Delta}TF$ expression. Primers used for the PCR amplification step are as follows: Forward 5'-tgaaGGCCTCTGAGGCCaattcgcggccgcatctaga, Reverse 5'-gcttGGCCTGACAGGCtgcagcggccgctactagt. Template-binding sequences are underscored. Adjacent nucleotides were designed to add *SfiI* restriction sites (uppercase) to each end. The full annotated sequences of all plasmids reported here are available online at Benchling - Hayneslab:

Synthetic Chromatin Actuators
(https://benchling.com/hayneslab/f/S0I0WLoRFK-synthetic-chromatin-actuators/).

Cell culture, transfection, and stable cell lines

MCF7 (ATCC HTB-22) cells were cultured in Eagle's Minimal Essential Medium supplemented with 0.01 mg/mL human recombinant insulin, 10% fetal bovine serum, and 1% penicillin and streptomyicn. BT-474 cells (ATCC HTB-20) were cultured in ATCC Hybri-Care Medium supplemented with 1.5 g/L sodium bicarbonate, 10% fetal bovine serum, and 1% penicillin and streptomycin. BT-549 cells (ATCC HTB-122) were cultured in RPMI-1640 Medium supplemented with 0.0008 mg/mL human recombinant insulin, 10% fetal bovine serum, and 1% penicillin and streptomycin. MCF-10A cells (ATCC CRL-10317) were cultured in Mammary Epithelial Cell Growth Medium (Mammary Epithelial Cell Basal Medium and BulletKit supplements, excepting gentamycin-amphotericin B mix), supplemented with 100 ng/mL cholera toxin. Cells were grown at 37 °C in a humidified CO_2 incubator. PcTF-expressing MCF7, BT-474, and BT-549 cells were generated by transfecting $5x10^5$ cells in 6-well plates with DNA/Lipofectamine complexes: 2 µg of hPCD-TF_MV2 plasmid DNA, 7.5 µl of Lipofectamine LTX (Invitrogen), 2.5 PLUS reagent, 570 µl OptiMEM. Control cells were mock-transfected with DNA-free water. 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 72 hrs. To generate PcTF-inducible cell lines, MCF7, BT-474, BT-549, and MCF-10A cells were transfected with either hPCD-TF_pSB-tetGP or TF_pSB-tetGP, following the same treatment conditions as above. 24 hrs after transfection, the transfection medium was replaced with fresh medium supplemented with puromycin at 0.5 µg/mL (1 µg/mL for MCF10A cell lines). Cells were then grown until either individual GFP-positive colonies could be isolated (low transfection efficiency lines BT-549 and MCF-10A), or cell cultures were >90% GFP-positive as measured by flow cytometry (high transfection efficiency lines BT-474 and MCF-7). Total culture time was 2-3 weeks per cell line.

Preparation of total mRNA

Total messenger RNA was extracted from ~90% confluent cells (~$1\text{-}2x10^6$). Adherent cells were lysed directly in culture plates with 500 µl TRIzol. TRIzol cell lysates were extracted with 100 µl 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).

Quantitative reverse transcription PCR (qRT-PCR)

SuperScript III (Invitrogen) was used to generate cDNA from 2.0 µg of RNA. Real-time quantitative PCR reactions (15 µl each) contained 1x LightCycler 480 Probes Master Mix (Roche), 2.25 pmol of

primers (see Supplemental Table 1 for sequences), and 2 µl of a 1:10 cDNA dilution (or 1:1000 dilution for GAPDH and mCh). The real time PCR program was run as follows: Pre-incubation, ramp at 4.4°C*sec^{-1} to 95°C, hold 10 min.; Amplification, 45 cycles (ramp at 4.4°C*sec^{-1} to 95°C, hold 10 sec., ramp at 2.2°C*sec^{-1} to 60°C, hold 30 sec., single acquisition); Cooling, ramp at 2.2°C*sec^{-1} to 40°C, hold 30 sec. 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. Expression level was calculated as *delta* C_p = 2^[C_p *GAPDH* - C_p *experimental gene*]. Fold change was determined as *double delta* C_p = *delta* C_p *treated cells* / *delta* C_p *mock* for PcTF expression levels (Fig. 3C), or as *double delta* C_p = C_p *dox treated cells* / *delta* C_p *no dox* for gene expression levels in the stable cell lines (Fig. 3D).

Transcriptome profiling with RNA-seq

RNA-seq was performed using two biological replicates per cell type, treatment, and time point for transiently transfected cells and three replicates for untransfected MCF10A. 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. Samples were sequenced at 8 per lane to generate an average of 2.5E+07 reads per sample. Read values ranged from 5.7E+06 (minimum) to 1.11E+08 (maximum) per sample.

Transcriptome analysis

RNA-seq reads were quality-checked before and after trimming and filtering using FastQC [81]. TrimmomaticSE was used to clip bases that were below the PHRED-scaled threshold quality of 10 at the 5' end and 25 at the trailing 3' end of each read for all samples [82]. A sliding window of 4 bases was used to clip reads when the average quality per base dropped below 30. Reads of less than 50 bp were removed. A combined reference genome index and dictionary for GRCH38.p7 (1-22, X, MT, and non-chromosomal sequences) [83] that included the full coding region of the synthetic PcTF protein were created using Spliced Transcripts Alignment to Reference (STARv2.5.2b) [84] and the

picard tools (version 1.1.19) [85]. Trimmed RNA-seq reads were mapped, and splice junctions extracted, using STARv2.5.2b read aligner [84]. Bamtools2.4.0 [86] was used to check alignment quality using the 'stats' command. Mapped reads in BAM format were sorted, duplicates were marked, read groups were added, and the files were indexed using the Bamtools2.4.0 package. CuffDiff, a program in the Cufflinks package [53], was used to identify genes and transcripts that expressed significant changes in pairwise comparisons between conditions. Fastq and differential expression analysis files are available at the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (Accession GSE103520, release date September 8, 2017). CummeRbund [53] was used to calculate distances between features and to generate graphs and charts (JSD plots). R ggplot2 [83,87] and VennDiagrams [88] were used to generate heat maps and Venn diagrams respectively. The entire workflow is provided as a readme file at: https://github.com/WilsonSayresLab/PcTF_differential_expression

Bioinformatics analyses and sources of public shared data

For the results shown in Figure 1B, genome-wide H3K27me3 enrichment in MCF7 cells, determined by chromatin immunoprecipitation followed by deep sequencing (ChIP-seq), was downloaded from the ENCODE project (accession UCSC-ENCODE-hg19:wgEncodeEH002922) [89]. We classified genes with a ChIP-seq peak within 5000 bp up or downstream of the transcription start site as H3K27me3-positive (1,146 protein-coding transcripts). EZH2-enriched genes (2,397 protein-coding transcripts) for MDA-MB-231 [15] were provided as a list from E. Benevolenskaya (unpublished). For the results shown in Figure 4, MCF7 ChIP-seq data (from the P. Farnham, J. Stamatoyannopoulos, and V. Iyer labs) was downloaded from the ENCODE project [89]: H3K27me3 (ENCFF081UQC.bigWig), H3K9me3 (ENCFF754TEC.bigWig), H3K27ac (ENCFF986ZEW.bigWig), H3K4me3 (ENCFF530LJW.bigWig), and RNA PolII (ENCFF690CUE.bam) and used to generate plots using DeepTools [90] (computeMatrix, plotProfile, plotHeatmap) in the Galaxy online platform at usegalaxy.org [91]. Prior to plotting, the RNA PolII data was converted to bigWig format using bamCoverage. Figure S4 was generated using REViGO [92] Unique differentially expressed genes were researched using GeneCards [69] and GOrilla analysis [93]. For the data shown in Figure S5, REViGO [92] was used to compare the biological functions for the differentially expressed genes. The results in Figure 5 are based on human tumor suppressor genes (983 total) that are reported to show lower expressed in cancer samples of the Cancer Genome Atlas (TCGA) compared to the TCGA normal tissue samples was downloaded from https://bioinfo.uth.edu/TSGene/download.cgi. Of these 983 genes, 589 are breast cancer specific [67,68].

ACKNOWLEDGEMENTS

The project was supported by a grant from the Arizona Department of Health Services (ADHS) Arizona Biomedical Research Commission (ABRC) (14-082976 to K.A.H.). M.A.W.S. and K.C.O. were supported by startup to M.A.W.S from the School of Life Sciences and the Biodesign Institute at Arizona State University. K.A.H. was supported by K01 CA188164 from the NIH NCI. D.B.N. was supported by 14-082976 from the ADHS. The authors thank Dr. E.V. Benevolenskaya for the generous gift of EZH2 gene module data for MDA-MB-231.

AUTHOR INFORMATION

Contributions

K.C.O. performed differential expression transcriptome analysis, identification of targeted upregulated genes in response to PcTF, and submission of NGS data to Gene Expression Omnibus (GEO). D.B.N performed cell culturing and transfection, preparation of samples for RNA-seq, and qRT-PCR. M.A.W.S. was responsible for the oversight of the bioinformatics analyses and interpretation of the data. K.A.H. was responsible for the conception of the project, and oversight of molecular cloning, cell culturing, and RNA-seq. All authors contributed to writing of the article.

Competing Interests

The authors declare no conflicts of interest.

REFERENCES

1. Margueron R, Reinberg D. Chromatin structure and the inheritance of epigenetic information. Nat Rev Genet. 2010;11: 285–296.
2. Richards EJ, Elgin SCR. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell. 2002;108: 489–500.
3. Roemer I, Reik W, Dean W, Klose J. Epigenetic inheritance in the mouse. Curr Biol. 1997;7: 277–280.
4. Cavalli G, Paro R. The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell. 1998;93: 505–518.
5. Kim J, Orkin SH. Embryonic stem cell-specific signatures in cancer: insights into genomic regulatory networks and implications for medicine. Genome Med. 2011;3: 75.
6. Sparmann A, van Lohuizen M. Polycomb silencers control cell fate, development and cancer. Nat

Rev Cancer. 2006;6: 846–856.

7. Wang W, Qin J-J, Voruganti S, Nag S, Zhou J, Zhang R. Polycomb Group (PcG) Proteins and Human Cancers: Multifaceted Functions and Therapeutic Implications. Med Res Rev. 2015;35: 1220–1267.

8. Chang C-J, Yang J-Y, Xia W, Chen C-T, Xie X, Chao C-H, et al. EZH2 promotes expansion of breast tumor initiating cells through activation of RAF1-β-catenin signaling. Cancer Cell. 2011;19: 86–100.

9. Alford SH, Toy K, Merajver SD, Kleer CG. Increased risk for distant metastasis in patients with familial early-stage breast cancer and high EZH2 expression. Breast Cancer Res Treat. 2012;132: 429–437.

10. Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc Natl Acad Sci U S A. 2003;100: 11606–11611.

11. Niida A, Smith AD, Imoto S, Aburatani H, Zhang MQ, Akiyama T. Gene set-based module discovery in the breast cancer transcriptome. BMC Bioinformatics. 2009;10: 71.

12. Collett K, Eide GE, Arnes J, Stefansson IM, Eide J, Braaten A, et al. Expression of enhancer of zeste homologue 2 is significantly associated with increased tumor cell proliferation and is a marker of aggressive breast cancer. Clin Cancer Res. 2006;12: 1168–1174.

13. Peña-Llopis S, Wan Y, Martinez ED. Unique epigenetic gene profiles define human breast cancers with poor prognosis. Oncotarget. 2016;7: 85819–85831.

14. Easwaran H, Johnstone SE, Van Neste L, Ohm J, Mosbruger T, Wang Q, et al. A DNA hypermethylation module for the stem/progenitor cell signature of cancer. Genome Res. 2012;22: 837–849.

15. Jene-Sanz A, Varaljai R, Vilkova AV, Khramtsova GF, Khramtsov AI, Olopade OI, et al. Expression of Polycomb Targets Predicts Breast Cancer Prognosis. Mol Cell Biol. 2013;33: 3951–3961.

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

17. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150: 12–27.

18. Tabet S, Douglas SF, Daze KD, Garnett GAE, Allen KJH, Abrioux EMM, et al. Synthetic trimethyllysine receptors that bind histone 3, trimethyllysine 27 (H3K27me3) and disrupt its interaction with the epigenetic reader protein CBX7. Bioorg Med Chem. 2013;21: 7004–7010.

19. Simhadri C, Daze KD, Douglas SF, Quon TTH, Dev A, Gignac MC, et al. Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7). J Med Chem. 2014;57: 2874–2883.

20. Stuckey JI, Dickson BM, Cheng N, Liu Y, Norris JL, Cholensky SH, et al. A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1. Nat Chem Biol. 2016;12:

180–187.

21. Biancotto C, Frigè G, Minucci S. Histone Modification Therapy of Cancer. Advances in Genetics. 2010. pp. 341–386.

22. Mani S, Herceg Z. DNA Demethylating Agents and Epigenetic Therapy of Cancer. Advances in Genetics. 2010. pp. 327–340.

23. McGarvey KM, Fahrner JA, Greene E, Martens J, Jenuwein T, Baylin SB. Silenced Tumor Suppressor Genes Reactivated by DNA Demethylation Do Not Return to a Fully Euchromatic Chromatin State. Cancer Res. 2006;66: 3541–3549.

24. McGarvey KM, Greene E, Fahrner JA, Jenuwein T, Baylin SB. DNA methylation and complete transcriptional silencing of cancer genes persist after depletion of EZH2. Cancer Res. 2007;67: 5097–5102.

25. Su I-H, Dobenecker M-W, Dickinson E, Oser M, Basavaraj A, Marqueron R, et al. Polycomb group protein ezh2 controls actin polymerization and cell signaling. Cell. 2005;121: 425–436.

26. Ueda K, Yoshimi A, Kagoya Y, Nishikawa S, Marquez VE, Nakagawa M, et al. Inhibition of histone methyltransferase EZH2 depletes leukemia stem cell of mixed lineage leukemia fusion leukemia through upregulation of p16. Cancer Sci. 2014;105: 512–519.

27. Xu B, On DM, Ma A, Parton T, Konze KD, Pattenden SG, et al. Selective inhibition of EZH2 and EZH1 enzymatic activity by a small molecule suppresses MLL-rearranged leukemia. Blood. 2015;125: 346–357.

28. Fujiwara T, Saitoh H, Inoue A, Kobayashi M, Okitsu Y, Katsuoka Y, et al. 3-Deazaneplanocin A (DZNep), an inhibitor of S-adenosylmethionine-dependent methyltransferase, promotes erythroid differentiation. J Biol Chem. 2014;289: 8121–8134.

29. Tekel SJ, Vargas DA, Song L, LaBaer J, Haynes KA. Tandem histone-binding domains enhance the activity of a synthetic chromatin effector [Internet]. 2017. doi:10.1101/145730

30. Nyer DB, Daer RM, Vargas D, Hom C, Haynes KA. Regulation of cancer epigenomes with a histone-binding synthetic transcription factor. npj Genomic Medicine. 2017;2. doi:10.1038/s41525-016-0002-3

31. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell. 2006;10: 515–527.

32. Lacroix M, Leclercq G. Relevance of breast cancer cell lines as models for breast tumours: an update. Breast Cancer Res Treat. 2004;83: 249–289.

33. Goodspeed A, Heiser LM, Gray JW, Costello JC. Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. Mol Cancer Res. 2016;14: 3–13.

34. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011;121: 2750–2767.

35. Tseng L-M, Chiu J-H, Liu C-Y, Tsai Y-F, Wang Y-L, Yang C-W, et al. A comparison of the molecular subtypes of triple-negative breast cancer among non-Asian and Taiwanese women.

Breast Cancer Res Treat. 2017;163: 241–254.

36. Kenny PA, Lee GY, Myers CA, Neve RM, Semeiks JR, Spellman PT, et al. The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. Mol Oncol. 2007;1: 84–96.

37. Nagaraja GM, Othman M, Fox BP, Alsaber R, Pellegrino CM, Zeng Y, et al. Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics. Oncogene. 2006;25: 2328–2338.

38. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003;100: 8418–8423.

39. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences. 2001;98: 10869–10874.

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

41. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature. 2006;441: 349–353.

42. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 2006;125: 301–313.

43. Zuo T, Liu T-M, Lan X, Weng Y-I, Shen R, Gu F, et al. Epigenetic silencing mediated through activated PI3K/AKT signaling in breast cancer. Cancer Res. 2011;71: 1752–1762.

44. Lin H-JL, Zuo T, Lin C-H, Kuo CT, Liyanarachchi S, Sun S, et al. Breast cancer-associated fibroblasts confer AKT1-mediated epigenetic silencing of Cystatin M in epithelial cells. Cancer Res. 2008;68: 10257–10266.

45. Ren G, Baritaki S, Marathe H, Feng J, Park S, Beach S, et al. Polycomb protein EZH2 regulates tumor invasion via the transcriptional repression of the metastasis suppressor RKIP in breast and prostate cancer. Cancer Res. 2012;72: 3091–3104.

46. Leroy G, Dimaggio PA, Chan EY, Zee BM, Blanco MA, Bryant B, et al. A quantitative atlas of histone modification signatures from human cancer cells. Epigenetics Chromatin. 2013;6: 20.

47. Derfoul A, Juan AH, Difilippantonio MJ, Palanisamy N, Ried T, Sartorelli V. Decreased microRNA-214 levels in breast cancer cells coincides with increased cell proliferation, invasion and accumulation of the Polycomb Ezh2 methyltransferase. Carcinogenesis. 2011;32: 1607–1614.

48. Dong M, Fan X-J, Chen Z-H, Wang T-T, Li X, Chen J, et al. Aberrant expression of enhancer of zeste homologue 2, correlated with HIF-1α, refines relapse risk and predicts poor outcome for breast cancer. Oncol Rep. 2014;32: 1101–1107.

49. Seals DF, Azucena EF Jr, Pass I, Tesfay L, Gordon R, Woodrow M, et al. The adaptor protein

Tks5/Fish is required for podosome formation and function, and for the protease-driven invasion of cancer cells. Cancer Cell. 2005;7: 155–165.

50. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013;31: 46–53.

51. Gonzàlez-Porta M, Frankish A, Rung J, Harrow J, Brazma A. Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. Genome Biol. 2013;14: R70.

52. Rupp SM, Webster TH, Olney KC, Hutchins ED, Kusumi K, Wilson Sayres MA. Evolution of Dosage Compensation in Anolis carolinensis, a Reptile with XX/XY Chromosomal Sex Determination. Genome Biol Evol. 2017;9: 231–240.

53. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7: 562–578.

54. Sledz CA, Holko M, de Veer MJ, Silverman RH, Williams BRG. Activation of the interferon system by short-interfering RNAs. Nat Cell Biol. 2003;5: 834–839.

55. Olejniczak M, Galka P, Krzyzosiak WJ. Sequence-non-specific effects of RNA interference triggers and microRNA regulators. Nucleic Acids Res. 2010;38: 1–16.

56. Fiszer-Kierzkowska A, Vydra N, Wysocka-Wycisk A, Kronekova Z, Jarząb M, Lisowska KM, et al. Liposome-based DNA carriers may induce cellular stress response and change gene expression pattern in transfected cells. BMC Mol Biol. 2011;12: 27.

57. Jacobsen L, Calvin S, Lobenhofer E. Transcriptional effects of transfection: the potential for misinterpretation of gene expression data generated from transiently transfected cells. Biotechniques. 2009;47: 617–624.

58. Huerfano S, Ryabchenko B, Forstová J. Nucleofection of expression vectors induces a robust interferon response and inhibition of cell proliferation. DNA Cell Biol. 2013;32: 467–479.

59. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol. 2007;8: R157.

60. Xu H, Xian J, Vire E, McKinney S, Wei V, Wong J, et al. Up-regulation of the interferon-related genes in BRCA2 knockout epithelial cells. J Pathol. 2014;234: 386–397.

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

62. Fischle W. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. 2003;17: 1870–1881.

63. Lachner M, Sengupta R, Schotta G, Jenuwein T. Trilogies of histone lysine methylation as epigenetic landmarks of the eukaryotic genome. Cold Spring Harb Symp Quant Biol. 2004;69: 209–218.

64. Haynes KA, Leibovitch BA, Rangwala SH, Craig C, Elgin SCR. Analyzing heterochromatin

formation using chromosome 4 of Drosophila melanogaster. Cold Spring Harb Symp Quant Biol. 2004;69: 267–272.

65. Nishibuchi G, Déjardin J. The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. Chromosome Res. 2017;25: 77–87.

66. Wiles ET, Selker EU. H3K27 methylation: a promiscuous repressive chromatin mark. Curr Opin Genet Dev. 2017;43: 31–37.

67. Min Zhao Jingchun Sun. TSGene: a web resource for tumor suppressor genes. Nucleic Acids Res. 2013; D970–6.

68. Zhao M, Kim P, Mitra R, Zhao J, Zhao Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res. 2015;44: D1023–D1031.

69. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: integrating information about genes, proteins and diseases. Trends Genet. 1997;13: 163.

70. Berrozpe G, Bryant GO, Warpinski K, Spagna D, Narayan S, Shah S, et al. Polycomb Responds to Low Levels of Transcription. Cell Rep. 2017;20: 785–793.

71. Guo X, Xiao H, Guo S, Dong L, Chen J. Identification of breast cancer mechanism based on weighted gene coexpression network analysis. Cancer Gene Ther. 2017; doi:10.1038/cgt.2017.23

72. Kwilas AR, Ardiani A, Dirmeier U, Wottawah C, Schlom J, Hodge JW. A poxviral-based cancer vaccine the transcription factor twist inhibits primary tumor growth and metastases in a model of metastatic breast cancer and improves survival in a spontaneous prostate cancer model. Oncotarget. 2015;6: 28194–28210.

73. Beltran A, Parikh S, Liu Y, Cuevas BD, Johnson GL, Futscher BW, et al. Re-activation of a dormant tumor suppressor gene maspin by designed transcription factors. Oncogene. 2007;26: 2791–2798.

74. Falke D, Fisher M, Ye D, Juliano RL. Design of artificial transcription factors to selectively regulate the pro-apoptotic bax gene. Nucleic Acids Res. 2003;31: e10.

75. Lara H, Wang Y, Beltran AS, Juárez-Moreno K, Yuan X, Kato S, et al. Targeting serous epithelial ovarian cancer with designer zinc finger transcription factors. J Biol Chem. 2012;287: 29873–29886.

76. Akishiba M, Takeuchi T, Kawaguchi Y, Sakamoto K, Yu H-H, Nakase I, et al. Cytosolic antibody delivery by lipid-sensitive endosomolytic peptide. Nat Chem. 2017;9: 751–761.

77. Essafi M, Baudot AD, Mouska X, Cassuto J-P, Ticchioni M, Deckert M. Cell-penetrating TAT-FOXO3 fusion proteins induce apoptotic cell death in leukemic cells. Mol Cancer Ther. 2011;10: 37–46.

78. Staahl BT, Benekareddy M, Coulon-Bainier C, Banfal AA, Floor SN, Sabo JK, et al. Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. Nat Biotechnol. 2017;35: 431–434.

79. Haynes KA, Silver PA. Synthetic reversal of epigenetic silencing. J Biol Chem. 2011;286:

27176–27182.

80. Kowarz E, Löscher D, Marschalek R. Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. Biotechnol J. 2015;10: 647–653.

81. Andrews S. FastQC: a quality control tool for high throughput sequence data. In: Babraham Bioinformatics [Internet]. 2010. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

82. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120.

83. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012;22: 1760–1774.

84. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21.

85. Picard Tools. In: Broad Institute [Internet]. 2003. Available: http://broadinstitute.github.io/picard/

86. Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011;27: 1691–1692.

87. Warnes MGR, Bolker B, Bonebakker L, Gentleman R. Package “gplots”. Various R Programming Tools for Plotting Data. 2016.

88. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12: 35.

89. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489: 57–74.

90. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44: W160–5.

91. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 2016;44: W3–W10.

92. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6: e21800.

93. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10: 48.