

Integrative Analysis Reveals the Transcriptional Collaboration between EZH2 and E2F1 in the Regulation of Cancer-Related Gene Expression

Han Xu^{1,2}, Kexin Xu^{1,2,3}, Housheng H. He^{1,2,3}, Chongzhi Zang^{1,2}, Chen-Hao Chen^{1,2}, Yiwen Chen^{1,2}, Qian Qin², Su Wang², Chenfei Wang², Shengen Hu², Fugen Li², Henry Long², Myles Brown^{2,3}, and X. Shirley Liu^{1,2,4}

Abstract

Overexpression of EZH2 is frequently linked to the advanced and metastatic stage of cancers. The mechanisms of its oncogenic function can be context specific, and may vary depending on the protein complexes that EZH2 interacts with. To identify novel transcriptional collaborators of EZH2 in cancers, a computational approach was developed that integrates protein–DNA binding data, cell perturbation gene expression data, and compendiums of tumor expression profiles. This holistic approach identified E2F1, a known mediator of the Rb tumor suppressor, as a transcriptional collaborator of EZH2 in castration-resistant prostate cancer. Subsequent analysis and experimental validation found EZH2 and E2F1 cobind to a subset of

chromatin sites lacking H3K27 trimethylation, and activate genes that are critical for prostate cancer progression. The collaboration of EZH2 and E2F1 in transcriptional regulation is also observed in diffuse large B-cell lymphoma cell lines, where activation of the transcriptional network is concordant with the cellular response to the EZH2 inhibitor.

Implications: The direct collaboration between EZH2 and Rb/E2F1 pathway provides an innovative mechanism underlying the cascade of tumor progression, and lays the foundation for the development of new anticancer targets/strategies. *Mol Cancer Res*; 14(2): 163–72. ©2015 AACR.

Introduction

Deregulation of proteins involved in the epigenetic machineries is commonly observed in cancer, and the resultant aberrant gene expression leads to cancer initiation, progression, and metastasis. One of the epigenetic regulators, the histone methyltransferase EZH2 (enhancer of zeste homolog 2), is frequently misregulated in a broad spectrum of tumors (1). As numerous studies have demonstrated that EZH2 plays a crucial role in multiple steps during tumor progression (2), targeting EZH2 represents a promising therapeutic option for advanced cancers. Several highly selective inhibitors of EZH2 methyltransferase activity have been generated (3–5). However, the biologic processes affected by the EZH2 inhibitors are

incompletely understood, and the molecular determinants of cancer cell responses to the drugs are still unclear.

EZH2 was originally identified as the catalytic subunit of the Polycomb repressive complex 2 (PRC2), which methylates K27 on histone H3 and leads to transcriptional silencing (6, 7). The oncogenic functions of EZH2 have been predominantly contributed to repression of tumor suppressor genes. However, increasing evidences, including our previous work, suggest that EZH2 may exert its oncogenic activities by transcriptional activation of genes critical for cancer progression (8–10). Under these scenarios, EZH2 physically interacts with signaling components other than PRC2 subunits, such as RelA/B, ER, and AR, to integrate the signaling circuitries for cancer-specific transcriptional programs. Therefore, it becomes intriguing and important to address the factors or pathways that cooperate with the oncogenic activity of EZH2.

To explore the transcriptional network in which EZH2 is involved to promote cancer progression, we developed a computational approach, named Gene Signature Association Analysis (GS2A), to infer the novel transcriptional collaborators of EZH2. The term "transcriptional collaborator" here refers to the factors that are colocalized with EZH2 at specific chromatin loci and cooperate to regulate the expression of a significant number of downstream target genes. The candidate collaborator predicted from GS2A was further validated using both experimental approaches and analysis of published datasets in other cancer types.

Materials and Methods

Data collection

The candidate collaborator list of 2196 genes were curated by taking the union of (i) 1,836 sequence-specific DNA-binding

¹Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts. ²Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, Massachusetts. ³Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts. ⁴School of Life Science and Technology, Tongji University, Shanghai, China.

Note: Supplementary data for this article are available at Molecular Cancer Research Online (<http://mcr.aacrjournals.org/>).

H. Xu and K. Xu contributed equally to this article.

Corresponding Authors: Xiaole Shirley Liu, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215. Phone: 617-632-2472; Fax: 617-632-2444; E-mail: xslu@jimmy.harvard.edu; and Myles A. Brown, myles_brown@dfci.harvard.edu

doi: 10.1158/1541-7786.MCR-15-0313

©2015 American Association for Cancer Research.

transcription factors (TF) published by Vaquerizas and colleagues (11), and (ii) 1,469 TFs, 296 transcription cofactors, and 117 chromatin remodeling factors available at AnimalTFDB database (12). CistromeMap database (<http://cistrome.org/pc/>) was used for the query of publicly available ChIP-seq datasets (13). E2F1 and BRCA1 ChIP-seq datasets from ENCODE cell lines were downloaded from UCSC genome browser databases.

Identification of EZH2 solo peaks from ChIP-seq

EZH2 and H3K27me3 ChIP-seq data in abl cell line were processed using MACS v2.0 based on $FDR < 0.01$ (14). To measure the EZH2 and H3K27me3 ChIP-seq intensity, we count the number of ChIP-seq reads within a 400-bp window centered by the summit of each EZH2 peak. If a read partially overlap with the window, the fraction of the read within the window is added to the count. To define the threshold for determining EZH2 solo peaks, we selected the top 1,000 EZH2 peaks with highest EZH2-binding intensity, and plotted the cumulative distribution of H3K27me3 intensity (Supplementary Fig. S1). A plateau is observed in the curve, confirming the existence of bi-model. The threshold was then chosen at the point where the slope of the curve is minimal.

Determining signature genes

The EZH2 (E2F1)-activated signature genes were selected on the basis of: (i) downregulated upon EZH2 (E2F1) silencing with $FDR < 0.01$ and fold change > 1.5 ; (ii) bound with EZH2 solo (E2F1) peaks at the promoter, that is, ± 2 kb from transcription start sites (TSS), in ChIP-seq data.

Preprocessing patient sample gene expression profile

For microarray datasets, raw measurements were normalized using Robust Multiarray Average. For RNA-seq data from The Cancer Genome Atlas (TCGA), measurements in reads per kilobases per million reads (RPKM) were transformed into log-scale, where a pseudocount of 1 was added to the RPKM measure. The normalized data underwent inverse normal transformation (INT) over the expression vector of each gene. INT transforms the distribution of variables into Gaussian, and represses the influence of the outliers on the correlation computation.

Association analysis between candidate gene and a set of signature genes

Suppose g_c is a candidate gene, G is the set of all genes in a data matrix, and $S = \{g_1, g_2, \dots, g_k\}$ is a set of k signature genes. We compute the Pearson correlation $\text{Cor}(g_c, x)$ between g_c and any other gene x in G . Let m_c and sd_c be the mean and SD of $\text{Cor}(g_c, x)$ for any $x \in G - \{g_c\}$, the significance of coexpression association between g_c and the signature gene set S is computed as:

$$E_c = \frac{1}{k} \sum_{i=1}^k \frac{\text{Cor}(g_c, g_i) - m_c}{sd_c}$$

This definition of association score is analogous to the modified Gene Set Enrichment Analysis (GSEA) score proposed by Irazzary and colleagues, which takes the form of a z-score (15).

Robust rank aggregation

To integrate multiple rank lists from multiple datasets, we employed the Robust Rank Aggregation (RRA) algorithm proposed by Kolde and colleagues (16). Suppose $R = \{r_1, r_2, \dots, r_n\}$ is the vector of ranks of a gene in n lists, where the number of genes in the lists are represented as $M = \{m_1, m_2, \dots, m_n\}$. RRA first normalizes the ranks into percentiles $U = \{u_1, u_2, \dots, u_n\}$ where $u_i = r_i/m_i$ ($i = 1, 2, \dots$). Under null hypotheses where the percentiles follow uniform distribution between 0 and 1, the k th smallest value among u_1, u_2, \dots, u_n is an order-statistic which follows beta distribution $B(k, n + 1 - k)$. RRA computes a P value p_k for the k th smallest value based on beta distribution. The significance score of the gene, that is, ρ value, is $\rho = \min(p_1, p_2, \dots, p_n)$.

In GS2A, each candidate gene is assigned a " ρ value" computed from the order-statistic, and a consensus rank list is generated in ascending order of the ρ values. GS2A randomly permutes the order of candidates in each list to determine the distribution of ρ value under the null hypothesis that the rank of a candidate is uniformly distributed. Empirical false discovery rates (FDR) are then computed by comparing the observed ρ values against the random distribution.

Motif enrichment analysis on ChIP-seq peaks

The sequences of 300 bp in length centered on summit were extracted for each ChIP-seq peak, and scanned with Position Weight Matrices. To determine the threshold for a motif "hit," background sequences of same length were randomly extracted from promoter regions (± 2 kb of TSS) of all RefSeq genes. Given a pokeweed mitogen, the threshold was chosen such that 10% of the background sequences have at least one hit.

Prostate cancer cell lines and culture conditions

The prostate cancer cell line LNCaP was obtained from the ATCC and cultured in RPMI1640 medium supplemented with 10% FBS, 2 mmol/L glutamine, 100 U/mL penicillin, and 100 mg/mL streptomycin. LNCaP-abl (abl) cell line was kindly provided by Zoran Culig (Innsbruck Medical University, Austria). It was maintained in phenol-red-free RPMI1640 medium (Gibco) with the same supplements except for 10% charcoal-dextran-stripped FBS. Both cell lines have recently been authenticated at Bio-Synthesis Inc. and confirmed to be mycoplasma-free using MycoAlert Mycoplasma Detection Kit (Lonza).

Transfection and Western blot analysis

Two different expression plasmids for HA-tagged E2F1 were purchased from Addgene. A total of 2 μ g of each E2F1 construct were transfected into LNCaP cells using SuperFect (Qiagen) according to the manufacturer's instructions. Seventy-two hours after transfection, either total RNA or proteins were collected. For immunoblotting, cells were lysed in RIPA lysis buffer [25 mmol/L Tris-HCl (pH 7.6), 150 mmol/L NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS], and the following antibodies were used for Western blot analysis: E2F1 (05-379, Millipore); HA-Tag (2367, Cell Signaling Technology); EZH2 (612666, BD Biosciences); and GAPDH (sc-32233, Santa Cruz Biotechnology).

RNA isolation and RT-PCR

RNA was extracted and purified using the TRIzol Reagent and RNeasy Mini Kit (Qiagen) according to the manufacturer's

protocols. qRT-PCR was performed upon cDNA synthesis using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Gene expression was calculated relative to the level of GAPDH using the $2^{-\Delta\Delta C_t}$ method as described previously (9). Sequences of primers were listed in Supplementary Table S5.

Standard ChIP and ChIP-seq assays

Chromatin immunoprecipitation (ChIP) experiments were performed as previously described (9). ChIP DNA was purified using PCR Purification Kit (Qiagen) and quantified by Quant-iT dsDNA HS Assay Kit (Invitrogen). Five to 10 ng of ChIP-enriched DNA were prepared for the ChIP-seq libraries using the ThruPLEX-FD Prep Kit (Rubicon Genomics) according to the manufacturer's protocol, and libraries were sequenced on the NextSeq 500. For targeted ChIP, extracted DNA was used for qRT-PCR with the specific primers as listed in Supplementary Table S6. The antibodies used for ChIP and ChIP-seq assays include: normal rabbit IgG (sc-2027, Santa Cruz Biotechnology); E2F1 (554213, BD Pharmingen and sc-193, Santa Cruz Biotechnology); and EZH2 (39933, Active Motif).

Coimmunoprecipitation assays

The coimmunoprecipitation (co-IP) experiments of endogenous EZH2 and E2F1 proteins were performed in abl cells as describe previously (9). Antibodies used for the co-IP assays include 6 μ g of E2F1 (sc-193, Santa Cruz Biotechnology); 8 μ L of EZH2 (39933, Active Motif); and equal amount of normal rabbit IgG (sc-2027, Santa Cruz Biotechnology).

RNA-seq protocols and data analysis

RNA-seq library was prepared using Illumina True-seq RNA sample preparation kit and sequenced to 50 bp using Illumina Hi-seq platform. The RNA-seq data was mapped to human genome build 37 (hg19) using TopHat version 2.0.6 (17). Gene expression level was calculated as gene-based fragment count per kilobases per million reads (FPKM) using Cufflinks version 2.0.2 (18), and the differential gene expression was measured as the fold change of the FPKM between knockdown and control conditions.

Genome resequencing of LNCaP and abl cell lines

Prostate cancer cell lines LNCaP and abl genomic DNA were extracted using Qiagen QIAamp DNA Mini Kit. 100 bp paired sequencing was performed using Illumina Hi-Seq platform. Paired reads of DNA sequences were aligned with BWA software (19), followed by realignment, recalibration, quality control, and genomic variant calling using GATK pipeline version 1.1 with default settings (20). Profile of relative copy number variations between LNCaP and abl were computed using VarScan version 2.2 (21).

Software availability

Source codes of GS2A modules are publicly accessible at <https://github.com/xh1974/GS2A>.

Data access

RNA-seq and ChIP-seq datasets generated along with the work have been submitted to Gene Expression Omnibus (GEO) with accession numbers GSE56331 and GSE67809, respectively. The relative copy number changes between LNCaP and abl, as pro-

cessed by VarScan on whole genome resequencing data, are available in Supplementary Table S7.

Results

Bi-model of EZH2 chromatin binding in castration-resistant prostate cancer cells

To explore the oncogenic function of EZH2 in cancer, we mapped EZH2 chromatin-binding sites and H3K27 trimethylation (H3K27me3) signals in LNCaP-abl (abl; ref. 9), a prostate cancer cell model that strongly resembles the clinical cases of metastatic, hormone-refractory prostate tumors (22). The ChIP-seq profiling revealed two types of EZH2-binding sites: "ensemble" peaks with both EZH2 and H3K27me3 enrichment (Fig. 1A), and "solo" peaks with only EZH2-binding signals (Fig. 1B). We analyzed the ChIP-seq data by quantitatively measuring the H3K27me3 signal across 26,645 EZH2 peaks in the genome, and confirmed the bi-model EZH2 binding in abl (Fig. 1C), with 22,042 (82.7%) ensemble peaks and 4,603 (17.3%) solo peaks (Supplementary Table S1).

To gain insights into the biologic functions of these two distinct patterns of EZH2 chromatin binding, we examined the association of the peaks with EZH2-regulated genes (Fig. 1D). Using gene expression profiling in abl cells, we identified 197 downregulated and 28 upregulated genes upon EZH2 silencing. The overrepresentation of EZH2-induced genes suggests a gene activation function of EZH2 in castration-resistant prostate cancer (CRPC) cells. Interestingly, 56 of EZH2-induced genes (28.4%) are bound by EZH2 solo peaks at their promoter regions, compared with only 15.5% of all RefSeq genes bound by the solo peaks ($P = 1.73E-6$). Gene ontology analysis on the 56 genes showed significant enrichment of cell-cycle regulators ($P = 2.0E-6$) and DNA packaging genes ($P = 1.7E-7$; ref. Fig. 1E). It is interesting to note that no enrichment of ensemble peaks at EZH2-repressed genes was observed. Among these EZH2-activated genes, several have been indicated to play critical roles in cancer progression, such as ATAD2 (23), CDK1 (24), and FOXM1 (25). In prostate cancer patient samples (26), these signature genes are significantly upregulated in the metastatic prostate tumors as compared with the primary cancer samples (Fig. 1F). These lines of evidence imply a novel yet important function of EZH2 in transcriptional regulation, which employs different mechanisms from the canonical PRC2 complex.

Computational inference of transcriptional collaborator of EZH2 in prostate cancer

Because EZH2 binds to solo peaks independent of its Polycomb repressor activity, we hypothesized that a distinct set of transcriptional collaborators mediates the recruitment of EZH2 to solo peaks and its gene activation function in abl cells. Therefore, we performed computational analysis that integrates multiple data sources to systematically screen for the potential transcriptional collaborators of EZH2.

The workflow of our computational approach, named GS2A, is illustrated in Fig. 2A. GS2A first selected the 56 EZH2-activated targets as "signature genes," whose expressions are expected to reflect the activities of EZH2-involved transcriptional network in gene activation (Supplementary Table S2). Next, GS2A screened 2,196 candidates, including TFs, cofactors, and chromatin regulators, to identify those coexpressed

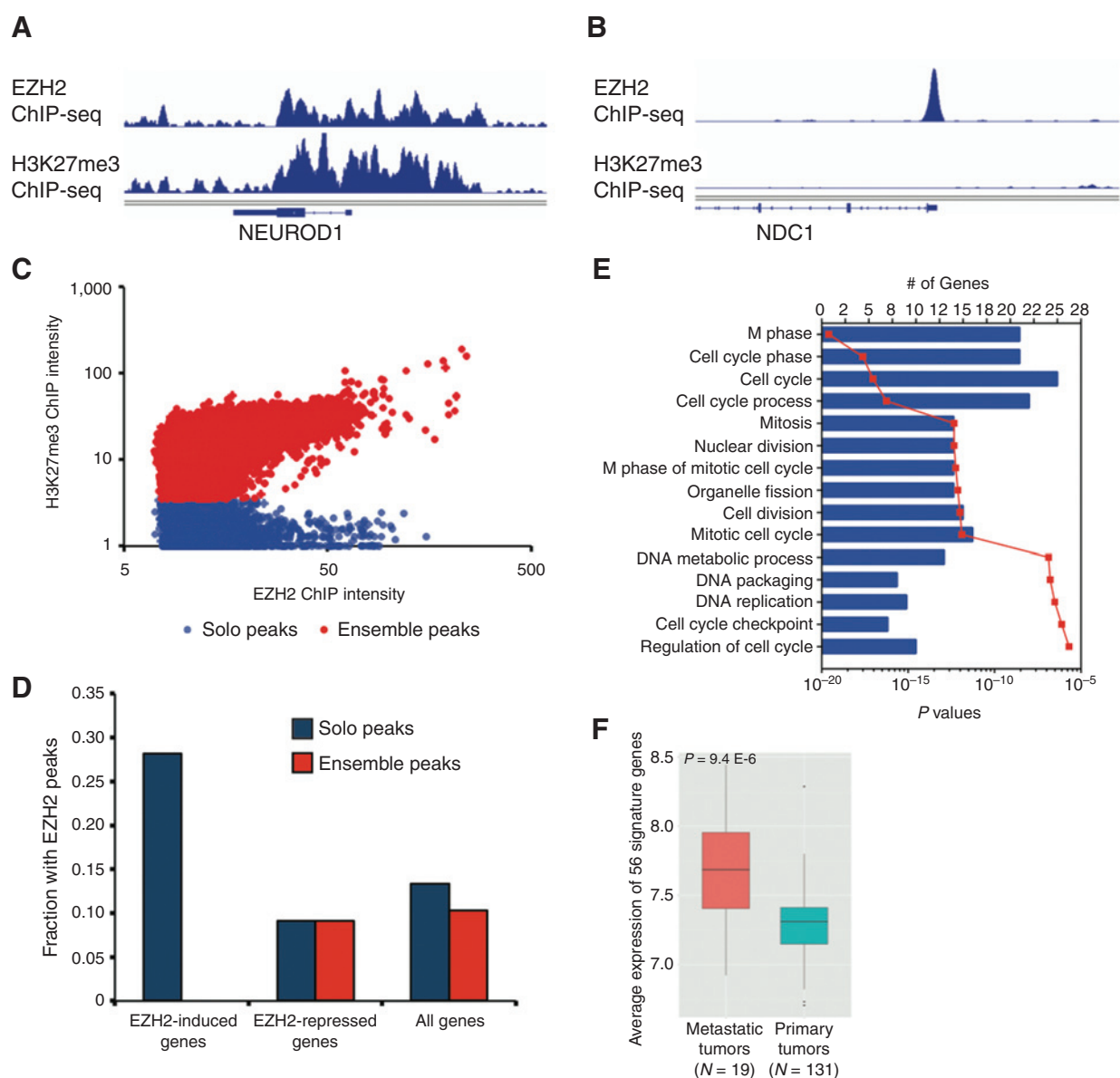
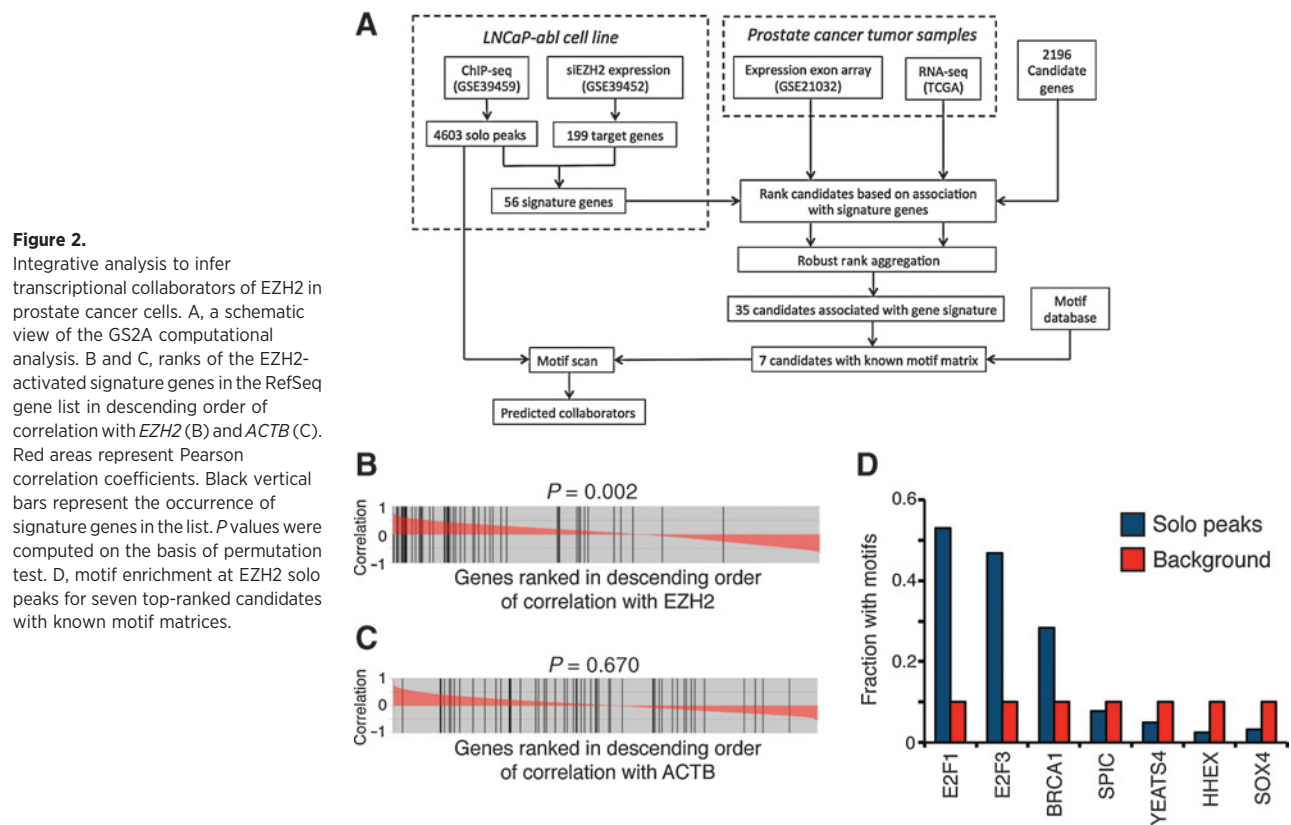


Figure 1. Bi-model of EZH2 chromatin-binding patterns in abl cells. A and B, ChIP-seq profiles demonstrating examples of EZH2 ensemble peaks (A) and solo peaks (B). C, H3K27me3 signals at EZH2-binding sites. Each dot in the scatter plot represents one EZH2-binding peak. D, bar plots showing the association between EZH2 chromatin-binding peaks and EZH2-dependent genes in abl cells. There are none EZH2-induced genes bound by the ensemble peaks. E, gene ontology analysis of the 56 EZH2-activated signature genes using DAVID functional analysis tool at <https://david.ncicrf.gov>. Blue bars, the number of genes in the specific category; red dots, P values of each functional group. F, comparison of the expression levels of the 56 signature genes between metastatic and primary prostate tumors (26). N , numbers of clinical samples in the specific stage of tumors. P value was calculated on the basis of Kolmogorov-Smirnov test.

with the signature genes in the compendiums of gene expression profiles from clinical prostate tumor samples. To measure the significance of coexpression association between a candidate regulator and a signature gene set, all the genes in a profiling dataset are ranked in order of their correlation with the candidate, followed by GSEA that measures the enrichment of signature genes in the top of the ranked list (15). A proof-of-concept hit is EZH2 itself, as most signature genes are at the top of the list sorted in descending order of the correlation with EZH2 (Fig. 2B). In contrast, the signature genes are

almost uniformly distributed with a housekeeping gene *ACTB* (Fig. 2C).

To improve the accuracy and reproducibility of the association analysis, we included two independent gene expression profiling datasets: a compendium of 131 primary prostate tumor samples on Affymetrix exon arrays (26) and TCGA RNA-seq data that includes 140 primary prostate tumor samples (27). As these expression cohorts have different profiling platforms, batch effects, and sample collection bias, GS2A generates one rank-ordered list for each cohort based on GSEA score, and



employs a RRA approach to compute a consensus rank list that integrates evidence from both cohorts. To this end, GS2A identified 35 genes that are reproducibly coexpressed with the signature genes (Supplementary Table S3). As expected, EZH2 itself is top ranked.

Among the 35 candidate genes, 6 have established sequence-specific motif matrixes. BRCA1 does not directly bind to DNA, but a *de novo* BRCA1 motif can be derived from published BRCA1 ChIP-seq data. We scanned the motifs of these seven factors in the pool of EZH2 solo peaks, and found that E2F1-binding sequence to be the most significantly enriched (Fig. 2D). Indeed, the coexpression association result showed a positive correlation of E2F1 with EZH2-activated signature genes (Supplementary Fig. S2A). Interestingly, the expression levels of EZH2 and E2F1 are only marginally correlated (Supplementary Fig. S2B), implying a direct correlation of E2F1 with EZH2-activated signature genes. As E2F1 directly binds to DNA, investigation on its role in EZH2-associated pathway may clarify the molecular mechanisms underlying the recruitment of EZH2 to solo peaks. Therefore, we focused on E2F1 for subsequent analysis.

Colocalization of EZH2 and E2F1 at chromatin regulatory elements lacking of H3K27me3 signals

To validate the genomic colocalization of EZH2 and E2F1, we first performed direct ChIP against either EZH2 or E2F1 at a number of selected locations. Although EZH2 is enriched at both types of sites, E2F1 can only be detected at solo peaks (Fig. 3A). We then generated the E2F1 ChIP-seq profile in *abl* cell, and the global chromatin-binding profiling confirmed the

enrichment of E2F1 binding at the proximal regions of EZH2 solo peaks, but not at the ensemble peaks (Fig. 3B). We also retrieved two public available E2F1 ChIP-seq datasets, one in the cervical cancer cell line Hela and the other in the breast cancer cell line MCF-7 (28). We found that more than 60% of EZH2 solo peaks overlap with E2F1-binding sites in at least one of the cell lines mentioned above, whereas less than 10% of the ensemble peaks overlap with E2F1 localization (Fig. 3C). This suggests that the EZH2 solo peaks are mostly conserved E2F1-binding sites in human cancers. We next performed the co-IP assays, and detected a clear interaction between endogenous EZH2 and E2F1 proteins in *abl* cells (Fig. 3D), suggesting that these two factors bind to each other physically. To this end, we conclude that EZH2 and E2F1 colocalize within a transcriptional complex at a specific subset of chromatin regulatory elements that lack the signals of repressive histone mark H3K27me3.

Mutual activation of downstream targets by the transcriptional complex consisting of EZH2 and E2F1

To validate the cooperation between EZH2 and E2F1 on transcriptional control, we compared the transcriptional programs induced by either protein. We found significant overlap between EZH2- and E2F1-activated genes (Fig. 4A), but observed little overlap between the repressed genes (Supplementary Fig. S3), suggesting that both proteins are involved in the same transcriptional network leading to similar gene activation patterns. Specifically, most EZH2-stimulated signature genes are downregulated upon E2F1 knockdown (Fig. 4B). When we overexpressed E2F1 in prostate cancer cells, the transcriptional

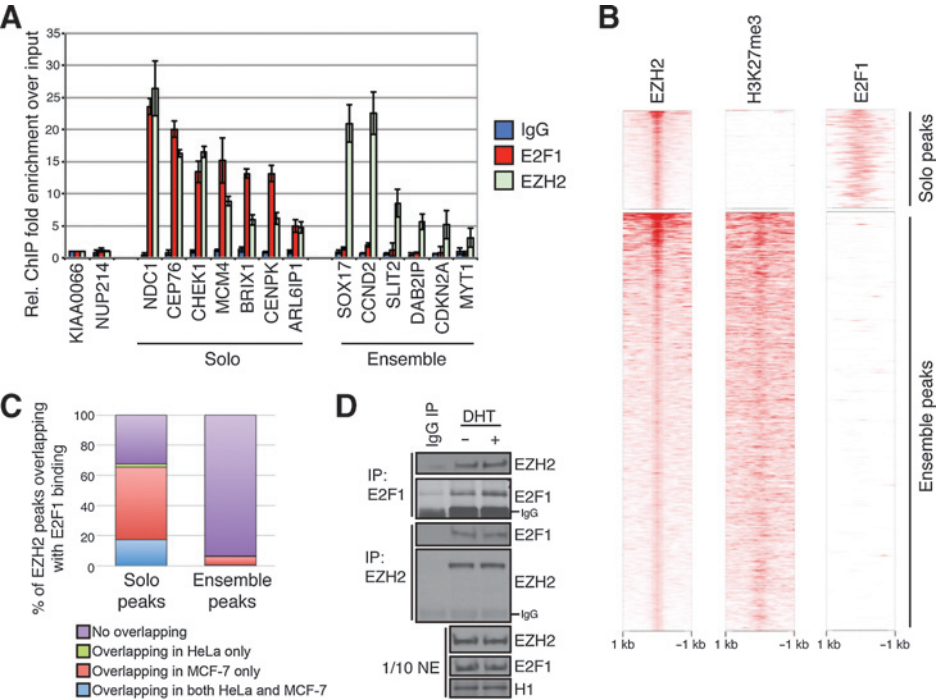


Figure 3. Colocalization of EZH2 and E2F1 at solo peaks. A, ChIP qPCR validation of EZH2 and E2F1 binding at solo or ensemble peaks in abl cells. Error bars represent the SD from three biologic replicates. B, heat maps showing ChIP-seq intensities of EZH2, H3K27me3, and E2F1 in a window size of 2 kb centered at the summits of EZH2 solo or ensemble peaks in abl. C, bar charts showing the percentages of EZH2 solo or ensemble peaks overlapping with E2F1-binding sites in HeLa and MCF-7 cells. D, Western blots of co-IP using indicated antibodies in abl cells. NE, nuclear extracts.

levels of EZH2-induced genes are significantly increased (Supplementary Fig. S4). Similarly, the expression levels of genes directly activated by E2F1 (Supplementary Table S4) are also

decreased when silencing EZH2 (Fig. 4C). The EZH2-mediated regulation of E2F1 targets is not likely through the modulation of E2F1 level, as the E2F1 protein expression did not show

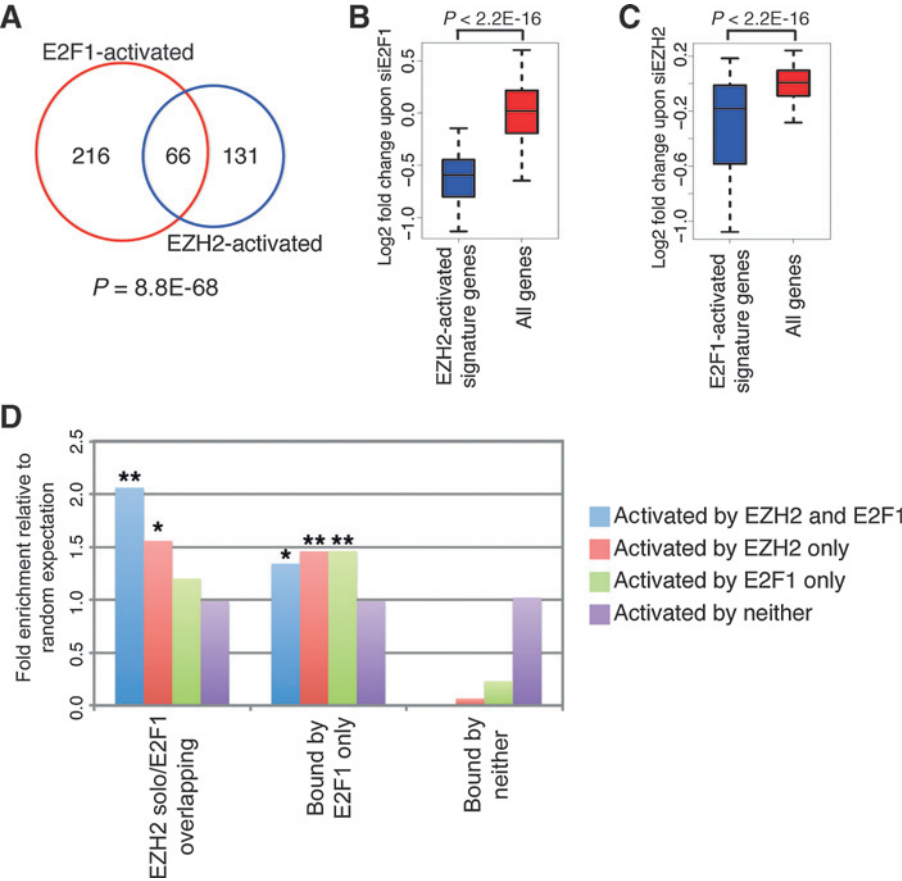


Figure 4. Collaboration between EZH2 and E2F1 in transcriptional activation of downstream targets. A, Venn diagram showing overlap of differential genes downregulated by silencing of EZH2 or E2F1 in abl cells. Numbers, number of genes that are activated by either EZH2 or E2F1 or both. P value was computed on the basis of hypergeometric distribution. B and C, box plots showing downregulation of EZH2-activated signature genes upon E2F1 knockdown (B) or E2F1-activated signature genes upon EZH2 knockdown (C) in abl cells. P values were computed using Kolmogorov-Smirnov test. D, bar charts showing the enrichment of EZH2 solo peaks and/or E2F1 chromatin binding within the promoter regions of genes activated by either/both proteins. The random expectations were computed by assuming independence between binding events and gene activation in 3*4 contingency table. P values were computed on the basis of hypergeometric distribution. n.s., not significant.

dramatic changes upon EZH2 silencing (Supplementary Fig. S5). These results further imply the direct crosstalk between EZH2 and E2F1 in transcriptional control. The collaboration between EZH2 and E2F1 can also be demonstrated in some other biologic systems. Data in mouse fibroblast showed similar upregulation pattern of EZH2-activated signature genes upon E2F1 overexpression (ref. 29; Supplementary Fig. S6A). Likewise, E2F1-activated signature genes were significantly downregulated when EZH2 was depleted in the primary erythroid progenitor cells (ref. 30; Supplementary Fig. S6B).

To determine the functional significance of the collaboration between EZH2 and E2F1, we correlated the chromatin-binding profiles of EZH2 and E2F1 with the gene patterns activated by either protein in *abl* cells (Fig. 4D). As 98.8% of the genes containing EZH2 solo peaks are also bound by E2F1 (Supplementary Fig. S7), we included unique E2F1-binding sites and background sequences as controls. It is noteworthy that the enrichment of binding sites near activated genes became the most prominent when considering the colocalization of EZH2 solo and E2F1 peaks at target genes coactivated by both. These data further confirmed the significance of EZH2 and E2F1 collaboration on direct activation of downstream genes.

Association of EZH2–E2F1 pathway with genomic aberrations in prostate cancer

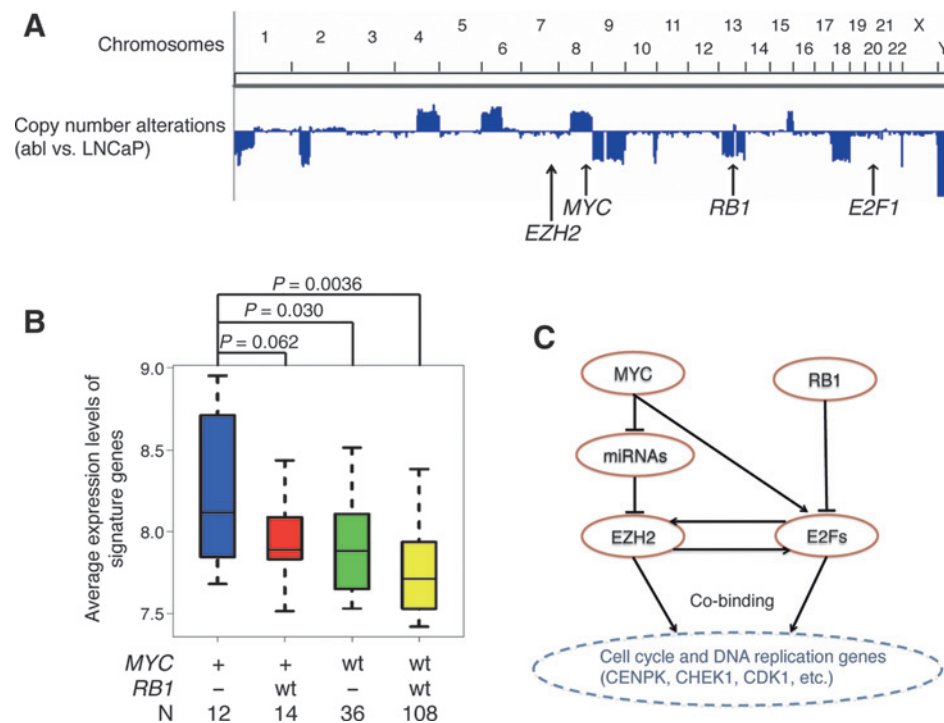
Somatic mutations of EZH2 have been identified in various lymphoid and myeloid neoplasms (31). However, EZH2 is frequently overexpressed but rarely mutated in prostate tumors (32). It is thereby compelling to explore the genetic alterations that activate EZH2-associated signaling. The Rb/E2F pathway has been well established as the upstream reg-

ulator of EZH2 expression (33). MYC coordinately stimulates EZH2 expression (34) and induces the expression as well as the transcriptional activity of E2F1 (35). Therefore, MYC amplification and *RB1* deletion, the genomic abnormalities that are frequently observed in a wide range of cancer types, may drive an EZH2/E2F1-dependent transcriptional program for cancer progression.

We firstly compared the sequence of castration-resistant *abl* cells against the parental androgen-dependent LNCaP using whole genome resequencing. MYC amplification and *RB1* deletion were both observed in *abl* compared with LNCaP (Fig. 5A), whereas no copy number aberrations were detected at EZH2 and E2F1 loci. We next asked whether the activation of EZH2–E2F1 pathway is associated with MYC amplification and *RB1* deletion in prostate tumors. We categorized the TCGA prostate cancer samples into four groups based on the copy number statuses of MYC and *RB1* ($MYC^{+}/RB1^{-}$, $MYC^{+}/RB1^{wt}$, $MYC^{wt}/RB1^{-}$, $MYC^{wt}/RB1^{wt}$; ref. 27). The signature genes are marginally overexpressed in $MYC^{+}/RB1^{wt}$ and $MYC^{wt}/RB1^{-}$ categories when comparing with the $MYC^{wt}/RB1^{wt}$ category, and are most significantly upregulated when MYC is amplified and *RB1* is deleted (Fig. 5B). These observations suggest that the combination of MYC amplification and *RB1* deletion may contribute to the activation of EZH2 and E2F1 signaling during prostate cancer progression.

In summary, we propose a schematic model of EZH2–E2F1 oncogenic pathway in prostate cancer, as shown in Fig. 5C. The MYC amplification and *RB1* deletion drive the activation of the transcriptional network consisting of EZH2 and E2F1. Cobinding of EZH2 and E2F1 at a specific set of chromatin regulatory sites further activates downstream genes that play critical roles in cancer cell proliferation.

Figure 5. Activation of EZH2–E2F1 transcriptional complex is associated with MYC and *RB1* genetic aberrations. A, genome-scale view of relative copy number alterations in *abl* cells compared with LNCaP cells. Arrows, the genetic loci of indicated genes. B, the average expression levels of EZH2-activated signature genes in prostate tumor samples with the indicated copy number alterations of MYC and *RB1*. *P* values were computed from *t* test. +, gene amplification; –, gene deletion; wt, wild type. *N*, number of clinical samples available in each corresponding category. C, the graphic view of the proposed model for oncogenic signaling pathway in prostate cancer.



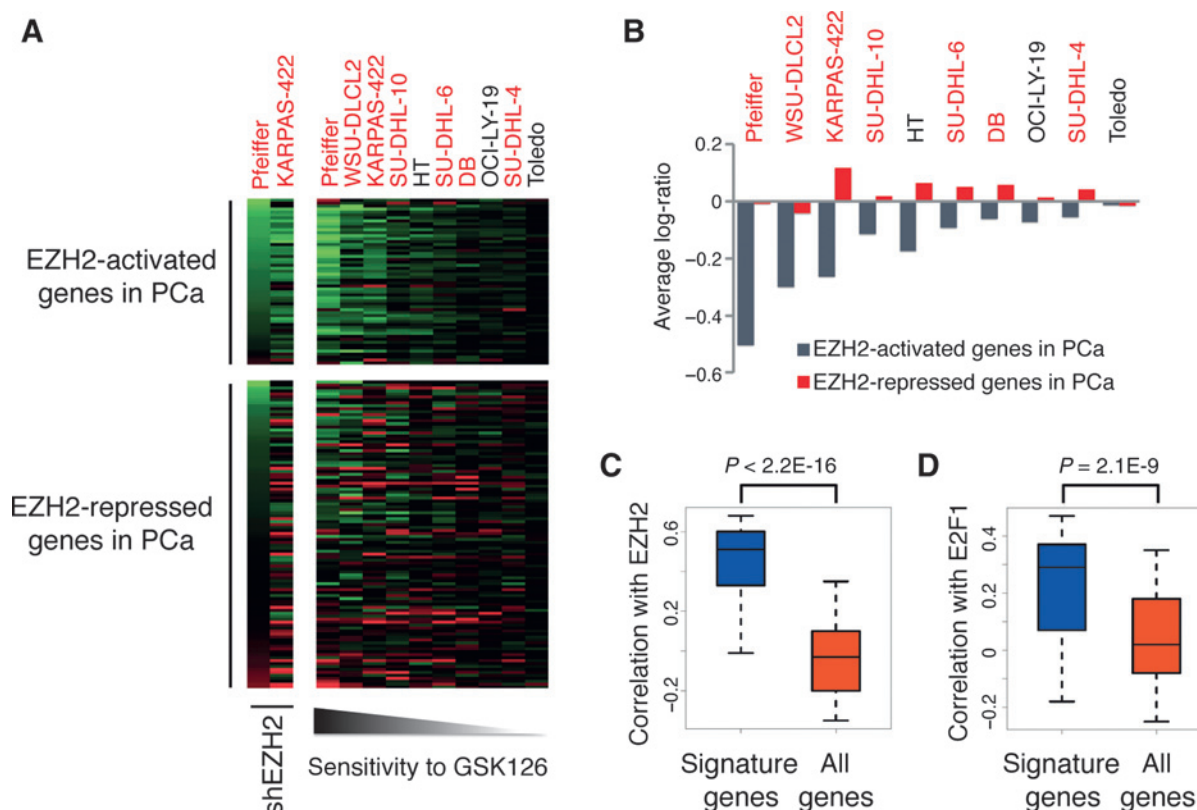


Figure 6. Collaboration between EZH2 and E2F1 in transcriptional control is conserved in DLBCL. **A**, heat maps showing the relative gene expression levels upon EZH2 knockdown (shEZH2) or EZH2 inhibitor (GSK126) treatment in DLBCL cell lines. For EZH2 silencing, cells were infected with either control shRNA or specific shRNA against EZH2 (shEZH2) for 10 days; for inhibitor treatment, cells were incubated with either 0.1% DMSO or 500 nmol/L GSK126 for 3 days. Top, EZH2-activated signature genes; bottom panel, EZH2-repressed genes identified in abl cells. Genes were sorted from top to bottom according to the expression change levels upon EZH2 knockdown in Pfeiffer cells. DLBCL cell lines with red fonts bear EZH2 mutations, and those with black fonts are EZH2 wild-type. Cells were sorted from left to right according to the sensitivity to GSK126. PCa, prostate cancer. **B**, changes of average log₂-ratio expression levels of EZH2-activated signature genes (blue bars) or EZH2-repressed genes (red bars) upon GSK126 treatment in DLBCL cell lines. **C** and **D**, box plots showing the correlation of EZH2-activated signature genes with EZH2 (**C**) and E2F1 (**D**) in DLBCL tumor samples. *P* values were computed on the basis of Kolmogorov-Smirnov test.

Conservation of EZH2–E2F1 pathway in DLBCLs and its clinical implication

To investigate the conservation of EZH2–E2F1 pathway in other cancer types and extend our discoveries to potential clinical applications, we collected the expression profiles in diffuse large B-cell lymphomas (DLBCL; ref. 3), and examined the expression patterns of EZH2-activated signature genes derived from prostate cancer cells under this scenario (Fig. 6A). This dataset includes the transcriptional profiles in 10 DLBCL cell lines treated with the small-molecule EZH2 inhibitor GSK126, as well as the profiles upon EZH2 knockdown in two of the GSK126-sensitive DLBCL cell lines, Pfeiffer and KARPAS-422.

We found that majority of the signature genes derived from abl cells are also downregulated upon EZH2 silencing or inhibitor treatment in DLBCL lines. However, the EZH2-repressed genes do not show a trend of derepression upon EZH2 perturbation. The magnitudes of the average expression changes for the activated signature genes are maximal in GSK126-sensitive cell lines, and are attenuated in GSK126-insensitive cell lines. In contrast, the expression changes of the canonical EZH2-repressed genes do not

agree with the cell responsiveness to GSK126 (Fig. 6B). We specifically examined the levels of three previously reported EZH2-repressed genes, *CDKN1A*, *IRF4*, and *PRDM1* (36, 37). These genes do not show significant or consistent upregulation in both DLBCL and prostate cancer cells upon either GSK126 treatment or EZH2 knockdown (Supplementary Fig. S8). We found similar correlation between the efficacy of GSK126 and gene expression changes in prostate cancer cells. GSK126 displayed relatively minor effects in LNCaP cells, but dramatically retarded the androgen-independent growth of abl (Supplementary Fig. S9). Interestingly, the genes that are coactivated by EZH2 and E2F1 were significantly downregulated by GSK126 in abl cells, and their levels were barely changed in LNCaP (Supplementary Fig. S10). Taken together, these results suggest that the activity of EZH2–E2F1 complex may determine the cancer cell response to the pharmacologic inhibition of EZH2.

To further confirm the conclusion in clinical scenarios, we computed the gene expression correlation in the dataset from DLBCL patient samples (38). Both EZH2 and E2F1 are significantly correlated with the signature genes in clinical samples (Fig. 6C and D). The close correlation of EZH2 and E2F1 with

their coactivated genes in clinical cases as well as with the cell responsiveness to EZH2 inhibitors support the ideas that the regulatory network between EZH2 and E2F1 plays an important role in determining the cancer-specific transcriptional profiles, and that activation of this signaling may contribute to the cellular behavior in response to EZH2-targeting drugs for cancer treatment.

Discussion

In this study, we demonstrated that EZH2 and E2F1 form a transcriptional complex colocalizing at specific chromatin sites to activate the downstream target genes. By integrative analysis using the GS2A pipeline, we found E2F1 as the top candidate that significantly correlates with the expression levels of EZH2-activated signature genes, and E2F1 colocalizes with EZH2 at specific chromatin sites in cancer cells. Most importantly, our discoveries from the cell models fit well with the sceneries in clinical tumors, and the findings provide clinical implications for EZH2-targeting inhibitors as anticancer therapy.

As a strong regulator of cell growth, the Rb/E2F pathway is commonly deregulated in many types of human cancers, such as Rb loss and E2F1 amplification/overexpression. Although being originally recognized to induce apoptosis after DNA damage, E2F1 is frequently upregulated in high-grade tumors and the increased level is closely associated with poor patient outcomes (39). The downstream target genes of E2F1, in particular those involved in cell-cycle progression, are abnormally overexpressed in cancers. Although regulation of E2F1 activity by EZH2 has been indicated in human cancers (40, 41), our findings provided a novel insight into the mechanisms of how perturbations of the Rb/E2F1 signaling lead to lethal tumor phenotypes through interaction with EZH2. The direct crosstalk between EZH2 and Rb/E2F1 pathway reinforce the gene expression patterns critical for cancer cell growth and thus constitutes a deadly alliance that underpins tumor progression.

Our previous work demonstrated the interplay between EZH2 and AR signaling critical for prostate tumor progression (9). Although the androgen response element is not listed among the top TF-binding motifs enriched in EZH2 ChIP-seq dataset, we confirmed its significant enrichment in the proximal region of EZH2 solo peaks. It is interesting to note that the Rb/E2F1 pathway has been shown as a critical regulator of AR expression and transcriptional activity (42–44). Although the activator E2Fs are generally believed to have a great deal of functional redundancy, only E2F1 and E2F3 were shown to enhance AR activity, whereas E2F2 was insufficient for this function (42). Interestingly, besides E2F1, our GS2A analysis implied the potential involvement of E2F3 in the transcriptional network with EZH2 as well, but not E2F2. This result not only confirms that GS2A is a powerful and practical approach to infer the biologically important signaling, but also indicates a decisive role of EZH2–E2F–AR axis in controlling downstream genes important for cancer progression.

Several questions remain unsolved for the mechanisms underpinning this critical transcriptional network for cancer progression. First, it is unclear about the precise molecular events that link the interplay between EZH2 and E2F1 to the transcriptional activation. Nonhistone substrates for EZH2, such as the TFs GATA4 and STAT3, have been identified (45, 46), and the modifications influence the transcriptional

activities by modulating either the protein stability or other posttranslational modifications critical for transactivation. As the enzymatic activity of EZH2 is required for the gene expression programs coregulated by E2F1, it is possible that E2F1 is another nonhistone substrate of EZH2. Second, it is intriguing that derepression of EZH2-repressed genes were not always observed in response to either EZH2 silencing or inhibitor treatment. In contrast, the genes that are directly activated by EZH2 were consistent downstream targets of the EZH2–E2F1 complex. The activities of EZH2 as either a transcriptional repressor or a transcriptional inducer have been both indicated, and the potential mechanisms of its conflicting functions may be the involvement of EZH2 in specific transcriptional complexes or pathways (47). It is therefore worth exploring the signaling/factors that determine the diverse and context-specific functions of EZH2. Third, although we showed strong correlation between MYC amplification and EZH2/E2F1 coregulated genes, the connection of the genetic alterations with the activities of EZH2–E2F1 network is yet incomplete. It is plausible that such genetic aberration induces the transactivation of EZH2 and E2F1 through modulation of critical *cis*-acting elements, and therefore amplifies the cascades of the transcriptional network (48).

In summary, our work uncovered the collaboration between EZH2 and E2F1 in controlling downstream genes that are highly expressed in aggressive tumors. The mechanisms underlying EZH2–E2F1 complex for cancer progression and their therapeutic potential await future research to be better understood and utilized.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: H. Xu, H.H. He, M. Brown

Development of methodology: H. Xu, C. Zang, Y. Chen, H. Long, X.S. Liu

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): K. Xu, H.H. He, C. Zang

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): H. Xu, K. Xu, C. Zang, C.-H. Chen, S. Wang, C. Wang, F. Li, H. Long, M. Brown, X.S. Liu

Writing, review, and/or revision of the manuscript: H. Xu, K. Xu, H.H. He, C. Zang, Y. Chen, M. Brown, X.S. Liu

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): H. Xu, C. Zang, Q. Qin, S. Hu

Study supervision: X.S. Liu

Acknowledgments

The authors thank Graham McVicker, Cliff Meyer, and Xiaoxing Wang for their helpful discussions.

Grant Support

This work was funded by NIH grants (GM056211 and HG4069 to X. Shirley Liu, 1K99CA178199-01A1 to Kexin Xu, K99CA172948 to Househeng H. He, and K99CA175290 to Yiwen Chen), Prostate Cancer Foundation Young Investigator award (to Kexin Xu), Dana-Farber Cancer Institute supports (to the Center of Functional Cancer Epigenetics), and National Science Foundation of China (NSFC 31329003 to X. Shirley Lu).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 16, 2015; revised November 16, 2015; accepted November 21, 2015; published OnlineFirst December 9, 2015.

References

1. Sparrmann A, van Lohuizen M. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* 2006;6:846–56.
2. Chang CJ, Hung MC. The role of EZH2 in tumour progression. *Br J Cancer* 2012;106:243–7.
3. McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, et al. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* 2012;492:108–12.
4. Qi W, Chan H, Teng L, Li L, Chuai S, Zhang R, et al. Selective inhibition of Ezh2 by a small molecule inhibitor blocks tumor cells proliferation. *Proc Natl Acad Sci U S A* 2012;109:21360–5.
5. Knutson SK, Wigle TJ, Warholc NM, Sneeringer CJ, Allain CJ, Klaus CR, et al. A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nat Chem Biol* 2012;8:890–6.
6. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 2002;298:1039–43.
7. Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, et al. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* 2002;111:197–208.
8. Lee ST, Li Z, Wu Z, Aau M, Guan P, Karuturi RK, et al. Context-specific regulation of NF-kappaB target gene expression by EZH2 in breast cancers. *Mol Cell* 2011;43:798–810.
9. Xu K, Wu ZJ, Groner AC, He HH, Cai C, Lis RT, et al. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science* 2012;338:1465–9.
10. Shi B, Liang J, Yang X, Wang Y, Zhao Y, Wu H, et al. Integration of estrogen and Wnt signaling circuits by the polycomb group protein EZH2 in breast cancer cells. *Mol Cell Biol* 2007;27:5105–19.
11. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 2009;10:252–63.
12. Zhang HM, Chen H, Liu W, Liu H, Gong J, Wang H, et al. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res* 2012;40:D144–9.
13. Qin B, Zhou M, Ge Y, Taing L, Liu T, Wang Q, et al. CistromeMap: a knowledgebase and web server for ChIP-Seq and DNase-Seq studies in mouse and human. *Bioinformatics* 2012;28:1411–2.
14. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R137.
15. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
16. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* 2012;28:573–80.
17. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25:1105–11.
18. 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;7:562–78.
19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
20. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
21. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 2009;25:2283–5.
22. Culig A, Hoffmann J, Erdel M, Eder IE, Hobisch A, Hittmair A, et al. Switch from antagonist to agonist of the androgen receptor bicalutamide is associated with prostate tumour progression in a new model system. *Br J Cancer* 1999;81:242–51.
23. Caron C, Lestrat C, Marsal S, Escoffier E, Curtet S, Virolle V, et al. Functional characterization of ATAD2 as a new cancer/testis factor and a predictor of poor prognosis in breast and lung cancers. *Oncogene* 2010;29:5171–81.
24. Willder JM, Heng SJ, McCall P, Adams CE, Tannahill C, Fyffe G, et al. Androgen receptor phosphorylation at serine 515 by Cdk1 predicts biochemical relapse in prostate cancer patients. *Br J Cancer* 2013;108:139–48.
25. Aytes A, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, et al. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* 2014;25:638–51.
26. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* 2010;18:11–22.
27. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;455:1061–8.
28. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
29. Ma Y, Croxton R, Moorer RL Jr, Cress WD. Identification of novel E2F1-regulated genes by microarray. *Arch Biochem Biophys* 2002;399:212–24.
30. Xu J, Shao Z, Li D, Xie H, Kim W, Huang J, et al. Developmental control of polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. *Mol Cell* 2015;57:304–16.
31. Hock H. A complex Polycomb issue: the two faces of EZH2 in cancer. *Genes Dev* 2012;26:751–5.
32. Bachmann N, Hoegel J, Haeusler J, Kuefer R, Herkommer K, Paiss T, et al. Mutation screen and association study of EZH2 as a susceptibility gene for aggressive prostate cancer. *Prostate* 2005;65:252–9.
33. Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* 2003;22:5323–35.
34. Sander S, Bullinger L, Klapproth K, Fiedler K, Kestler HA, Barth TF, et al. MYC stimulates EZH2 expression by repression of its negative regulator miR-26a. *Blood* 2008;112:4202–12.
35. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 2005;435:839–43.
36. Beguelin W, Popovic R, Teater M, Jiang Y, Bunting KL, Rosen M, et al. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* 2013;23:677–92.
37. Caganova M, Carrisi C, Varano G, Mainoldi F, Zanardi F, Germain PL, et al. Germinal center dysregulation by histone methyltransferase EZH2 promotes lymphomagenesis. *J Clin Invest* 2013;123:5009–22.
38. Monti S, Chapuy B, Takeyama K, Rodig SJ, Hao Y, Yeda KT, et al. Integrative analysis reveals an outcome-associated and targetable pattern of p53 and cell cycle deregulation in diffuse large B cell lymphoma. *Cancer Cell* 2012;22:359–72.
39. Engelmann D, Putzer BM. The dark side of E2F1: in transit beyond apoptosis. *Cancer Res* 2012;72:571–5.
40. Gao SB, Zheng QF, Xu B, Pan CB, Li KL, Zhao Y, et al. EZH2 represses target genes through H3K27-dependent and H3K27-independent mechanisms in hepatocellular carcinoma. *Mol Cancer Res* 2014;12:1388–97.
41. Wu ZL, Zheng SS, Li ZM, Qiao YY, Aau MY, Yu Q. Polycomb protein EZH2 regulates E2F1-dependent apoptosis through epigenetically modulating Bim expression. *Cell Death Differ* 2010;17:801–10.
42. Sharma A, Yeow WS, Ertel A, Coleman I, Clegg N, Thangavel C, et al. The retinoblastoma tumor suppressor controls androgen signaling and human prostate cancer progression. *J Clin Invest* 2010;120:4478–92.
43. Davis JN, Wojno KJ, Daignault S, Hofer MD, Kuefer R, Rubin MA, et al. Elevated E2F1 inhibits transcription of the androgen receptor in metastatic hormone-resistant prostate cancer. *Cancer Res* 2006;66:11897–906.
44. Mallik I, Davila M, Tapia T, Schanen B, Chakrabarti R. Androgen regulates Cdc6 transcription through interactions between androgen receptor and E2F transcription factor in prostate cancer cells. *Biochim Biophys Acta* 2008;1783:1737–44.
45. Kim E, Kim M, Woo DH, Shin Y, Shin J, Chang N, et al. Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. *Cancer Cell* 2013;23:839–52.
46. He A, Shen X, Ma Q, Cao J, von Gise A, Zhou P, et al. PRC2 directly methylates GATA4 and represses its transcriptional activity. *Genes Dev* 2012;26:37–42.
47. Deb G, Singh AK, Gupta S. EZH2: not EZHY (easy) to deal. *Mol Cancer Res* 2014;12:639–53.
48. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 2013;153:307–19.