Molecular and Cellular Biology

Expression of Polycomb Targets Predicts Breast Cancer Prognosis

Alba Jene-Sanz, Renáta Váraljai, Alexandra V. Vilkova, Galina F. Khramtsova, Andrey I. Khramtsov, Olufunmilayo I. Olopade, Nuria Lopez-Bigas and Elizaveta V. Benevolenskaya

Mol. Cell. Biol. **2013, 33(19):3951. DOI: 10.1128/MCB.00426-13.**
Published Ahead of Print 5 August 2013.

MCB
Journals.ASM.org

Expression of Polycomb Targets Predicts Breast Cancer Prognosis

Alba Jene-Sanz,[a,b] Renáta Váraljai,[b] Alexandra V. Vilkova,[b] Galina F. Khramtsova,[c] Andrey I. Khramtsov,[d] Olufunmilayo I. Olopade,[c] Nuria Lopez-Bigas,[a,e] Elizaveta V. Benevolenskaya[b]

Research Unit on Biomedical Informatics, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain[a]; Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, Illinois, USA[b]; University of Chicago, Department of Pathology, Chicago, Illinois, USA[c]; University of Chicago, Department of Medicine, Chicago, Illinois, USA[d]; Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain[e]

Global changes in the epigenome are increasingly being appreciated as key events in cancer progression. The pathogenic role of enhancer of zeste homolog 2 (EZH2) has been connected to its histone 3 lysine 27 (H3K27) methyltransferase activity and gene repression; however, little is known about relationship of changes in expression of EZH2 target genes to cancer characteristics and patient prognosis. Here we show that through expression analysis of genomic regions with H3K27 trimethylation (H3K27me3) and EZH2 binding, breast cancer patients can be stratified into good and poor prognostic groups independent of known cancer gene signatures. The EZH2-bound regions were downregulated in tumors characterized by aggressive behavior, high expression of cell cycle genes, and low expression of developmental and cell adhesion genes. Depletion of EZH2 in breast cancer cells significantly increased expression of the top altered genes, decreased proliferation, and improved cell adhesion, indicating a critical role played by EZH2 in determining the cancer phenotype.

Recognition of the role of the epigenome during the formation of cancer genomes may help to explain aspects of tumor development such as the late onset of most solid tumors, recurrent disease, tumor heterogeneity, risk factors, and environmental effects (1, 2). Deregulation of Polycomb repressive complex 2 (PRC2) proteins EZH2 and SUZ12 has been linked to the initiation of tumorigenesis through a variety of mechanisms, which ultimately prevent the expression of cell fate regulators and promote a stem cell-like phenotype (3, 4). EZH2 is overexpressed in several cancers and is associated with a worse prognosis in prostate, breast, endometrial, and melanoma tumors (5–8). Increased EZH2 has been shown to be implicated in the expansion of breast tumor-initiating cells (9) and required for promotion of metastasis in several tumors (10, 11). An elevated EZH2 protein level indicated a precancerous state in morphologically normal breast epithelium and increased as breast cancer developed and progressed (12).

The oncogenic function of EZH2 is believed to be mainly mediated through its gene-silencing activity (3). By placing the repressive histone modification mark H3K27me3 at the promoters of developmental regulators, the PRC2 complex is responsible for their silencing in stem cells (13). Consistent with the function of EZH2 as a histone methyltransferase, the loss of the H3K27me3 mark at those locations typically promotes cellular differentiation (14–16). We hypothesized that additional genomic regions occupied by PRC2 or enriched in H3K27 methylation may contribute to the processes leading to neoplasia. If this is the case, then with the development of EZH2 selective inhibitors (17), EZH2 is emerging as a very attractive target for anticancer therapies.

We set up to determine the gene expression pattern in PRC2 locations in aggressive and nonaggressive breast cancers and its association with certain cancer phenotypes at the whole-genome scale. We refer to the set of all genes found in such regions, for example, bound by EZH2, as the "EZH2 module," in reference to the concept of the "gene regulatory module," which designates a group of genes that share a regulatory property (e.g., all genes bound by the transcription factor EZH2 in a certain cell type). Modules can be obtained through analysis of any types of genome-wide studies and can also be literature based. We used systems biology approaches to analyze the largest breast cancer gene expression data sets, as well as determined the transcript and protein levels in primary human breast tumors. We established the relationship between expression of different modules and clinicopathological characteristics of breast tumors. This analysis revealed a high contribution of the EZH2 module to breast cancer characteristics. Using RNA inhibition experiments, we further showed that the expression level of genes in the EZH2 module, as well as the cancer cell properties that were associated with EZH2 module expression, was dependent on EZH2. We conclude that through the expression analysis of a PRC2 module, patients with worse and better prognoses can be reliably stratified, and opportunities for prognosis and application of EZH2 inhibitors should be further explored in patients with advanced breast cancer.

MATERIALS AND METHODS

Human breast cancer specimens. Tissue microarrays were constructed from 95 histologically confirmed breast cancer samples. All tissue samples were obtained and handled in accordance with an approved Institutional Review Board protocol. To identify breast cancer subtypes, we evaluated the tissues for the expression of estrogen receptor (ER), progesterone receptor (PR), HER2, keratins 5 and 6 (CK5/6), and epidermal growth factor receptor (EGFR). Staining with vimentin served as a control to monitor the quality of tissue fixation in archival tumors. Breast cancer subtypes were defined as luminal A (ER^+ and/or PR^+ and $HER2^-$), luminal B (ER^+ and/or PR^+ and $HER2^+$), basal-like (ER^-, PR^-, $HER2^-$,

Received 8 April 2013 Returned for modification 7 May 2013
Accepted 28 July 2013
Published ahead of print 5 August 2013
Address correspondence to Elizaveta V. Benevolenskaya, evb@uic.edu, or Nuria Lopez-Bigas, nuria.lopez@upf.edu.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/MCB.00426-13.

doi:10.1128/MCB.00426-13

CK5/6$^+$, and/or EGFR$^+$), HER2$^+$/ER$^-$ (HER2$^+$, ER$^-$ and PR$^-$), and unclassified (negative for all five markers) as described previously (18). EZH2 expression was measured by staining using an anti-EZH2 antibody (catalog no. 187395; Invitrogen). For the transcript analysis from the breast tissue microarray, RNA samples were prepared from 6 tumor tissues and 6 matching normal tissues. An additional 18 cDNAs isolated from breast ductal carcinomas were analyzed from the TissueScan Cancer Survey tissue quantitative PCR (qPCR) panel 384-1 (CSRT102; OriGene).

Cell culture and inducible shRNA expression. MCF7, MDA-MB-231 and 293T cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Mediatech) supplemented with 10% fetal bovine serum (FBS) (HyClone), and MCF10A cells were grown in DMEM F-12 medium (Mediatech) supplemented with 5% horse serum, 10 μg/ml insulin, 100 ng/ml cholera toxin, and 0.5 μg/ml hydrocortisone. All cells were grown in 37°C in a 5% CO_2 incubator. Cells were seeded in a 24-well tissue culture plate at a density of 3×10^4 cells/well in medium supplemented with doxycycline (Dox$^+$) at the concentration 100 ng/ml or in Dox-free (Dox$^-$) medium. Cells were plated in triplicates. Medium was changed every 3 days. Cell viability was assessed by trypan blue exclusion analysis and was 90 to 100%. Cells were imaged for red fluorescent protein (RFP) using fluorescence microscopy (Leica DM IRB), and images were acquired using Q Capture PRO software. Matrigel assays were performed in 8-well glass chambers with 5×10^3 cells/well in growth medium containing 2% Matrigel (BD Biosciences). Cells were fed every 4 days with assay medium containing 2% Matrigel, which was supplemented with 5 ng/ml epidermal growth factor (EGF) for MCF10A cells. The successful miR-30-based short hairpin RNAs (shRNA) to EZH2 were generated from pTRIPZ V3THS-387506 (pTRIPZ-ON-EZH2-1) and V2THS-17506 (Thermo Fisher) by deletion of sequences encoding the rtTA3 transactivator. Constructs were subjected to restriction digestion with BamHI and religated to give pTRIPZ-OFF-EZH2-1 and pTRIPZ-OFF-EZH2-2, respectively. The same modification was introduced in the control nonsilencing construct pTRIPZ RHS4743 (Thermo Fisher). To make the constructs function as Tet-Off, they were transduced into tTA-Advanced expression cells. The tTA-Advanced MCF7 cell line was described before (19). For virus production, 293T cells were grown at 80% confluence and transfected by Lipofectamine 2000 with packaging constructs pMD2.G and psPAX2, the pTRIPZ construct, and 7 μg/ml Polybrene. After 8 h, the transfection medium was replaced with growth medium. Cells were selected for integrated constructs with puromycin dihydrochloride for 2 days.

Analysis of protein and gene expression. Protein lysates were separated by 6.25% SDS-PAGE. For immunoblot analysis, mouse EZH2 BD43 (Millipore) and α-tubulin T9026 (Sigma) antibodies were used and blots were developed using the ECL enhanced chemiluminescence system. The detection of epithelial-mesenchymal transition (EMT) markers was performed with anti-rabbit antibody to β-catenin (sc-7199; Santa Cruz) and anti-mouse antibodies to E-cadherin (CD324; BD Biosciences) and vimentin (v5255; Sigma). Alexa Fluor 647-labeled anti-mouse (Invitrogen) and Cy3-labeled anti-rabbit (Jackson Laboratory) antibodies were used as the secondary antibodies. Nuclei were stained with DAPI (4′,6-diamidino-2-phenylindole). Immunofluorescent staining of acini cultured in Matrigel was performed essentially as described previously (20). Acini were mounted in FluorSave reagent (Calbiochem), and confocal images were taken using Zeiss LSM 700 laser scanning microscope using Zen 2009 (Zeiss Enhanced Navigation) software. For gene expression analysis, total RNAs from cell lines were isolated using TRIzol (Sigma). Total RNAs from formalin-fixed paraffin-embedded (FFPE) tissue samples were isolated using the MagMAX FFPE total nucleic acid isolation kit (Ambion), and then 1.8 μg of the sample RNA was used for first-strand cDNA synthesis using the Superscript Vilo kit (Life Technologies). Real-time PCR was performed using the SYBR green PCR master mix and the CFX96 system (Bio-Rad). The expression level of experimental genes was normalized to the geometrical mean of three control genes, *SDHA*, *UBC*, and *POLR2A*. The primer sequences used in reverse transcription (RT)-qPCR are available in Table S10 in the supplemental material.

Public data sets. (i) PRC2 modules. We created lists of genes regulated by PRC2 from experimental data in available resources (see Table S1 in the supplemental material). The degree of overlap between these gene lists is shown in Fig. S1B in the supplemental material. The gene lists include human genome-wide occupancy data sets from chromatin immunoprecipitation sequencing (ChIP-seq) experiments in embryonic stem (ES) cells (21, 22) that we processed using Bowtie version 0.12.5 (hg19 genome assembly, unique alignments, allowing 2 mismatches) (23) for short-read alignment. For peak detection of EZH2 and SUZ12, we used MACS (version 1.4.1) (24) setting "--nomodel" and two times of "--bw" (bandwidth) parameter to determine the scan window whenever a control IP was not available); the shift size was estimated using Pyicos (25). For H3K27me3, we used SICER (version 1.1) (26) (setting gap size to 600). Regions were assigned to protein-coding genes (Ensembl v60) if they overlapped either to the gene body or up to 5 kb upstream from the transcription start site, using BedTools (27). Raw reads from each data set were checked for quality with FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/), and overall peak-calling performance was evaluated with CEAS (28).

Breast cancer intrinsic subtype signatures were collected from the MsigDB v3.0 (29). Intrinsic subtypes for the data set from Ivshina et al. (30) were annotated using the robust model in R package genefu (http://www.R-project.org). Intrinsic subtypes from Sabatier et al. (31) were obtained from the authors' annotations.

(ii) Other modules. We obtained and classified breast cancer prognostic gene modules (see Table S6 in the supplemental material) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (32) and Gene Ontology (GO) (33) gene sets. KEGG and GO terms were obtained from Biomart (Ensembl v65).

We used eight publicly available expression profiling data sets downloaded from Gene Expression Omnibus (GEO) (34–37) and the Cancer Genome Atlas (TCGA) Data Portal (38) (see Table S2 in the supplemental material). Each transcriptome data set consists of microarray expression data for primary tumors, except for GSE1133, which contains normal tissues and malignant cell lines, and GSE3526, which consists entirely of healthy tissues. The sample number varies from 159 to 533 across all cancer data sets. Data were preprocessed as previously described (39) and filtered for protein-coding genes (according to Ensembl v60 annotations). The input data for enrichment analysis were obtained by median centering of the expression value of each gene across all the tumor samples (row median) and dividing this value by the standard deviation (row standard deviation) using the R Bioconductor package (40). For all cancer data sets, normal samples were removed before median centering. The obtained value is the measure of expression level for the gene in a sample compared to its expression level in all other samples in the data set.

SLEA. Sample-level enrichment analysis (SLEA) was performed using Gitools version 1.6.0 (41; http://www.gitools.org). We used the *z*-score method as described previously (39, 42). This method compares the mean (or median) expression value of genes in each module to a distribution of the mean (or median) of 10,000 random modules of the same size. Such enrichment analysis was run for each sample and the result was a *z*-score, which is a measure of the difference between the observed and expected mean (or median) expression values for genes in a module. The *P* value related to the *z*-score was corrected for multiple testing using Benjamini-Hochberg false discovery rate (FDR) method (43). We define positively enriched modules in a sample as those with a positive *z*-score and a corrected *P* of <0.01, while negatively enriched modules have negative *z*-scores with a corrected *P* of <0.01. Besides *z* values for individual samples, we also applied the mean *z*-score enrichment values, which are the arithmetic means of *z*-scores for individual samples. To test for significant differences between the *z*-score means within each stratified group of samples we used the Mann-Whitney test implemented in Gitools. All heat maps were generated with Gitools.

Survival analysis. We used the log rank test ("survdiff") and Cox proportional hazards ("coxph") from R Bioconductor package (40) to calculate the significance and hazard ratios, respectively, and "survplot"

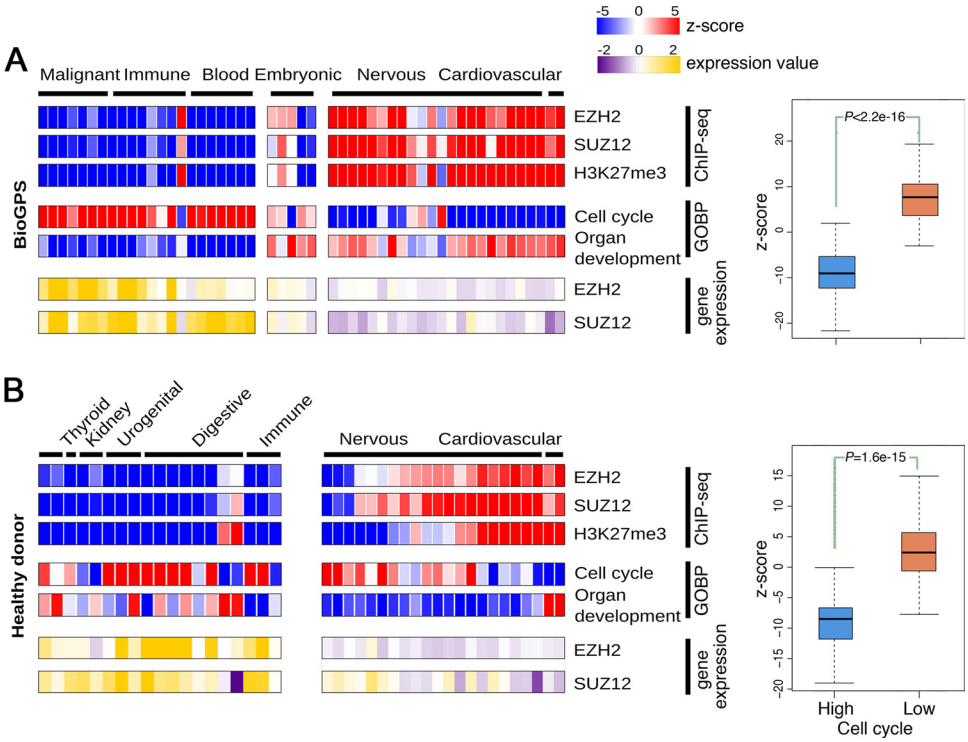

FIG 1 PRC2 module gene expression reflects the balance between cell cycle and differentiation. (A) SLEA analysis identifies tissues and cell lines from BioGPS with significantly changed expression of PRC2 modules. Heat maps of regulatory modules (rows) and tissue samples (columns). "Cell cycle" and "Organ development" correspond to genes annotated under the GO BP terms GO:0007049 and GO:0048513, respectively. Upregulation of a module in each sample is shown in red and downregulation in blue. The bottom rows depict median-centered probe-level expression, with yellow and purple colors indicating higher and lower expression, respectively. Consistent with the repressive effects of EZH2 and SUZ12 on gene expression, their expression level in tissue samples and cell lines inversely correlated with the expression level of the PRC2 modules. (Right panels) Box plots of the PRC2 modules' *z*-scores in groups of samples selected according to the overall enrichment of the cell cycle module; *P* values denote Mann-Whitney one-sided test significance. (B) SLEA analysis identifies healthy tissues with significantly changed expression of PRC2 modules.

(http://www.cbs.dtu.dk/~eklund/survplot) for the Kaplan-Meier curves. In the survival analysis, the survival data of the samples with positive enrichment for the module (*z*-scores with corrected *P* of <0.01) was compared to that of the samples with negative enrichment in the data set; the group size was at least 20 samples.

RESULTS

PRC2 module expression reflects the balance between proliferation and differentiation in normal tissues. Given the important role the PRC2 complex and its target genes play during cell-fate determination, we investigated how PRC2 modules are expressed in different cell types. PRC2 is placing the pivotal H3K27 trimethylation mark, with EZH2 as a catalytic component, and SUZ12 inducing EZH2 activity and interacting with the nucleosomes. We used whole-genome data from human embryonic stem cells (see Table S1 and Fig. S1 in the supplemental material) to define each of the three PRC2-related modules, EZH2, SUZ12, and H3K27me3. As a comparison to PRC2 modules, we also studied genes that have been ascribed gene ontology (GO) terms according to the biological processes (BP) or pathways (KEGG).

We analyzed expression of these gene modules in human tissue sample data sets (tissues and cell lines in BioGPS [34] and a collection of tissues dissected from 10 different donors [35], see Table S2 in the supplemental material) using sample-level enrichment analysis (SLEA). The analysis evaluates if the module is overexpressed (positive *z*-score) or underexpressed (negative *z*-score) and provides value of statistical significance of such changes compared to any random module. The cell types with high expression of cell cycle genes, i.e., immune system, blood, as well as all cancer cell lines, showed downregulation of the PRC2 modules (Fig. 1A). In contrast, more differentiated cells from the nervous system and cardiovascular system showed a significant upregulation of the PRC2 modules (Fig. 1B). Therefore, levels of expression of the PRC2 module are significantly different between various human tissues, correlating with their proliferative and differentiation states, and separating normal from cancer cells.

Molecular and phenotypic characteristics of PRC2 module-stratified breast tumors. Given the fact that EZH2 overexpression is common in breast, prostate, and other cancers, we asked if the PRC2 modules are changed in expression in human cancers. We performed SLEA on samples from the six largest breast cancer studies (see Table S2 in the supplemental material). Intriguingly, our analysis revealed many tumor samples that showed highly significant upregulation of the EZH2, SUZ12, and H3K27me3 modules as well as many samples that showed downregulation of these PRC2 modules. PRC2-independent targets of EZH2 (see 286 genes in Fig. S1B in the supplemental material) do not seem to contribute to the EZH2 module expression changes in breast cancer data sets, because we were unable to see significant expression changes in these genes when we analyzed them separately (see Fig. S2 in the supple-

mental material). In Fig. 2, we present SLEA results for two breast cancer studies. We set the significance level at $P < 0.01$ and show samples with positive *z*-score values in red and samples with negative *z*-score values in blue, while samples with values that don't meet the significance level are in gray. These criteria separated the samples into three different groups.

In order to better understand the underlying differences between the groups of samples sorted according to EZH2 module gene expression, we studied their clinical annotations (see Tables S3 and S4 in the supplemental material). With respect to the subtype classification and tumor grade, there were many tumors of the basal subtype and grade III tumors in the group of samples with PRC2 module downregulation, while the group with PRC2 module overexpression was enriched in the "normal-like" subtype (Fig. 2). Consistent with these data, when we analyzed breast cancer intrinsic subtype signatures (see Table S5 in the supplemental material), there was a direct correlation in the expression of the PRC2 module and signatures for normal and luminal A subtypes, which were expressed significantly differently between the two groups as determined by Mann-Whitney test (see Fig. S3 in the supplemental material).

Next, we analyzed genes grouped in pathways or GO terms. We found that genes related to cell cycle, RNA transport, the spliceosome, or the proteasome, known as the Achilles gene sets, in which the high level of activation is required for the proliferation of cancer cells (44), as well as genes involved in oxidative phosphorylation are expressed higher in tumors with low expression of PRC2 modules (Fig. 2 and 3; see Fig. S4 in the supplemental material). In contrast, genes involved in cell adhesion, organ development, anatomical structure morphogenesis, and neuroactive ligand-receptor interaction were expressed at lower levels in these samples. The opposite was true for the samples with a high level of PRC2 modules.

We evaluated how well the best known modules associated with breast cancer aggressiveness (see Table S6 in the supplemental material), such as the Rosetta data set (45), correlate in expression with the PRC2 modules. There was significant difference in *z*-scores of several cancer gene modules between the groups of samples sorted according to EZH2 module gene expression (see Fig. S5 in the supplemental material). In particular, we observed higher expression of modules of the stem cell-like and undifferentiated phenotypes, epithelial-mesenchymal transition (EMT), and metastasis in the group of samples that expressed low levels of the EZH2 module (Fig. 2). The same samples were also characterized by active BRCA1 network, which might be associated with decreased nuclear expression of phospho-BRCA1 (Ser1423) (46). In conclusion, breast cancer patients can be stratified according to the transcription status of the EZH2 module independently of clinical characteristics and known cancer gene signatures. We identified shared key properties of patient groups enriched for the EZH2 module related to their phenotype.

EZH2 depletion induces the epithelial-to-mesenchymal transition. As the key property of patients with low expression of EZH2 modules is increased expression of modules related to cell adhesion and the EMT, we investigated whether EZH2 is essential in EMT. We used inducible expression of short hairpin RNAs (shRNAs) designed to two different regions in the *EZH2* gene to knock down *EZH2* expression in breast cancer cells representing luminal and basal-like tumor subtypes. The level of shRNA production was essentially undetectable under the noninducing (Dox^+) condition, as evidenced by the absence of the RFP reporter gene expression that is linked to shRNA synthesis (Fig. 3A). In the Dox^- condition, RFP was induced proportionally to the increasing amount of the lentiviruses. In the cells transduced with *EZH2* shRNAs, the *EZH2* transcript levels were substantially reduced by 70% and expression of EZH2 protein was detected by immunoblotting at very low level (Fig. 3B; see Fig. S6A in the supplemental material). EZH2 expression remained unaltered in cells expressing a control shRNA. For the analysis of EMT, we grew cell colonies as epithelial acini in three-dimensional basement membrane cultures on Matrigel. The size of acini formed was smaller and the number was much lower in MCF7 cells with induced EZH2 knockdown compared to uninduced cells (Fig. 3C). This was consistent with lower proliferation when cells were cultured as monolayers (Fig. 3D). Similar results were obtained in basal-like MDA-MB-231 cells (data not shown). We analyzed expression of 14 different epithelial and mesenchymal markers. Upon *EZH2* depletion in MCF7 cells, we observed the upregulation of mRNAs of β-catenin and E-cadherin and downregulation of vimentin at both the mRNA and protein levels (Fig. 3E and F), indicating that the EMT is reversed. Overall, transcript level of five luminal epithelial markers was significantly increased, while the level of six mesenchymal markers was decreased. This effect was generally reproducible with the second *EZH2* shRNA construct in MCF7 cells and upon *EZH2* depletion in MDA-MB-231 cells (Fig. 3G; see Fig. S6B). The MCF10A showed a similar growth inhibition by *EZH2* shRNA (see Fig. S6C and D), indicating that *EZH2* depletion affects not only the growth of breast cancer cells but also the growth of benign mammary epithelial cells; however, the effects on EMT marker expression were lower than in MCF7 cells (see Fig. S6E and F in the supplemental material). These data indicate that *EZH2* knockdown restores epithelial markers and represses mesenchymal markers in breast cancer cells.

High expression of PRC2 modules predicts better breast cancer outcome. We asked whether the tumor of a patient with worse prognosis would have any significant changes in expression of the PRC2 modules compared to a patient with a better outcome. In a previous study, increased breast cancer invasiveness and metastasis have been associated with the recruitment of the PRC2 complex to new genomic loci and, as a result, altered H3K27 methylation and gene expression (47). Another study determined differentially expressed genes when EZH2 is depleted (48). Survival tests using SLEA stratification demonstrated similar or superior predictive power of any of PRC2-related modules compared to previously described modules as predictors of breast cancer survival (Fig. 4; see Tables S6 and S7 in the supplemental material). Importantly, when we performed analysis within only TCGA stage I tumors to assess the value of early detection in reducing mortality, we also observed a significant difference in survival rates. We were able to dissect the PRC2 module contribution to patient survival from the contribution of other modules, as there was no significant overlap between genes in the PRC2 modules with the multicancer gene modules (see Table S8 and Fig. S7 in the supplemental material). For example, to exclude the contribution from genes influencing proliferation (49), we removed these genes from the PRC2 modules, and the SLEA results were almost identical to those obtained from the full set of regulatory modules (see Table S7). In fact, the H3K27me3 module displayed a signifi-

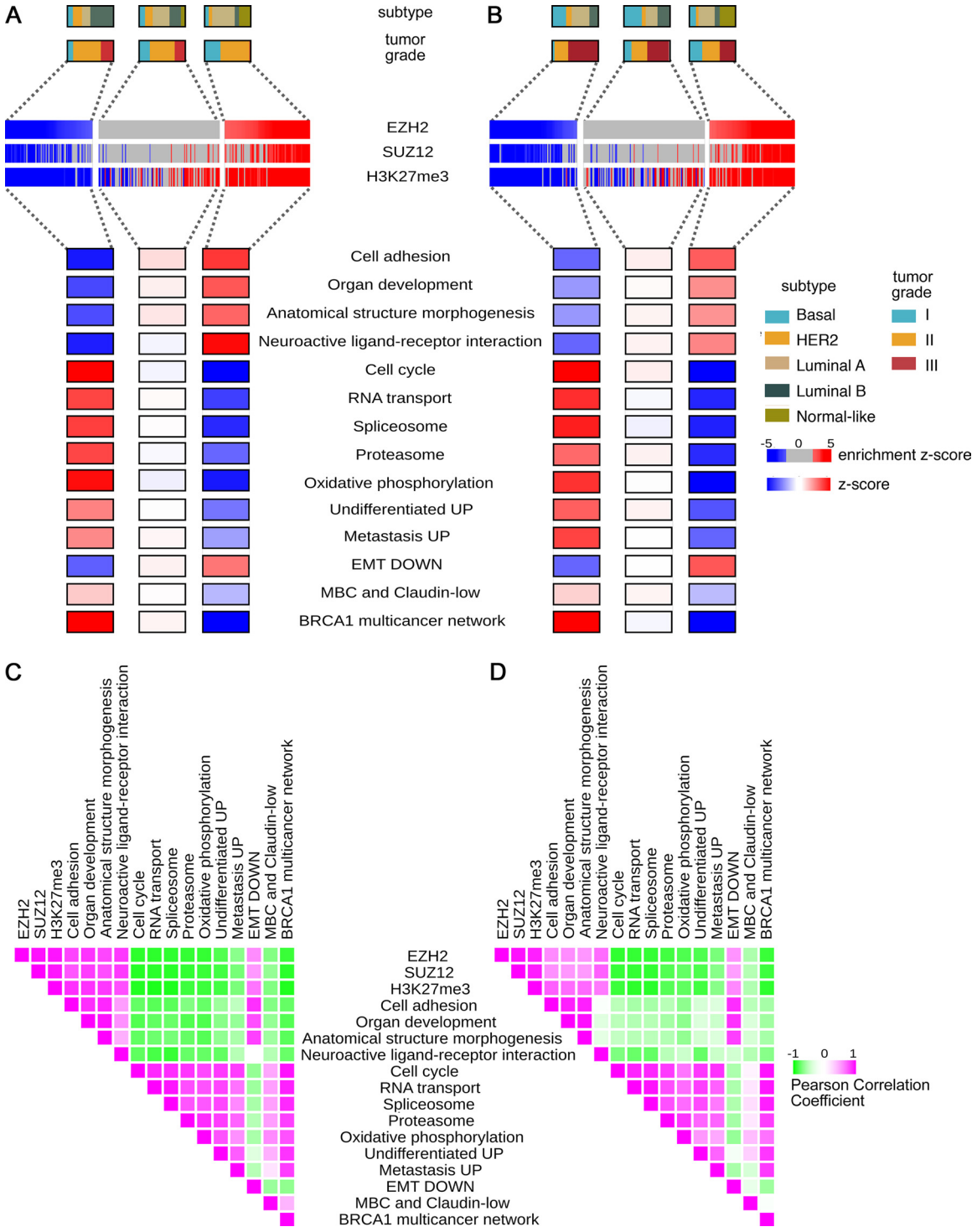

FIG 2 PRC2 module expression stratifies samples according to their molecular characteristics and aggressive tumor behavior. (A) PRC2 module stratification of breast tumor samples and enrichment of breast cancer prognosis signatures using gene expression data from Ivshina et al. (30). Samples were sorted according to the *z*-score value of the EZH2 module and divided into three groups with lower (in blue), nonsignificant (in gray), or higher (in red) expression of genes in the module (*z*-score significance level set at $P < 0.01$). (Upper panels) Color-coded annotations describe breast cancer subtypes and tumor grades taken from the clinical annotations of the patient samples. Note that the order of samples has been altered within each of the three groups to better represent the proportion of samples of each subtype or grade. (Middle panels) SLEA analysis of PRC2 modules is presented as heat maps for each tumor sample. (Lower panels) The mean *z*-score enrichment values of selected pathways, GO BP (see Fig. S4 in the supplemental material), and breast cancer prognostic gene modules (see Fig. S5 and Table S6 in the supplemental material) are presented for samples from each group stratified by EZH2 module enrichment. (B) The same analysis performed with samples from the study by Sabatier et al. (31). (C and D) Symmetric heat map representing the Pearson correlation coefficient of each pair of *z*-score vectors resulting from the enrichment analysis of PRC2 modules, GO terms and pathways, and prognostic modules. (C) Data set from Ivshina et al. (30); (D) data set from Sabatier et al. (31).

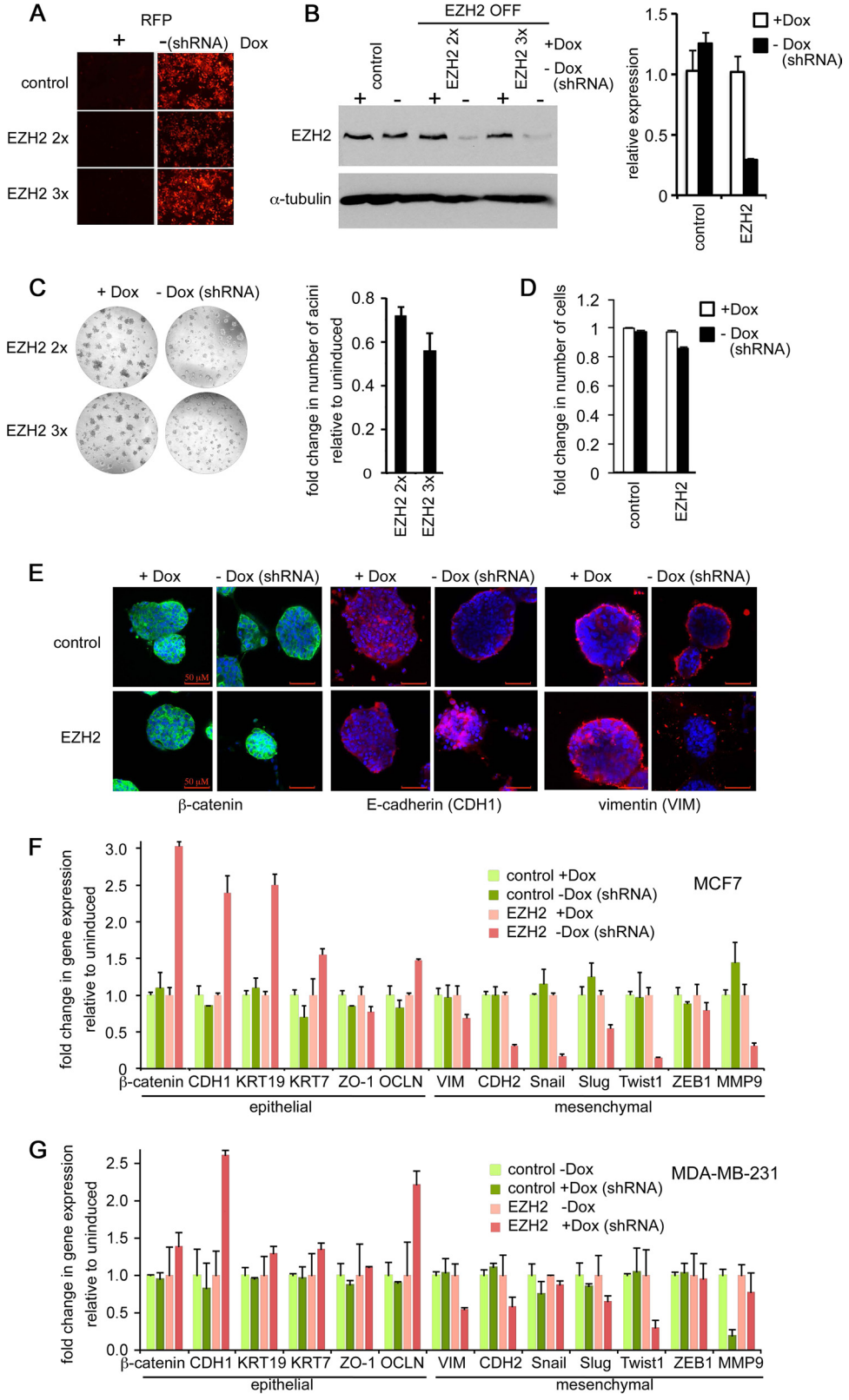

FIG 3 EZH2 depletion in breast cancer cells reduces cell proliferation and induces mesenchymal-to-epithelial transition. (A) Induction of shRNAs in the Tet-Off configuration is tightly regulated with doxycycline. MCF7 cells stably expressing the tTA protein were transduced with a lentivirus expressing the indicated shRNAs (nonsilencing control and pTRIPZ-OFF-EZH2-1) and treated with 100 ng/ml doxycycline (+Dox) for 6 days before analysis. The level of shRNA expression was monitored by the level of a TurboRFP reporter, which was induced proportionally to the amount of virus added to the cells (2× and 3×). (B) Difference in level of EZH2 expression upon Dox treatment as determined by RT-qPCR and immunoblot analyses. (C) Reduced proliferation of *EZH2* shRNA

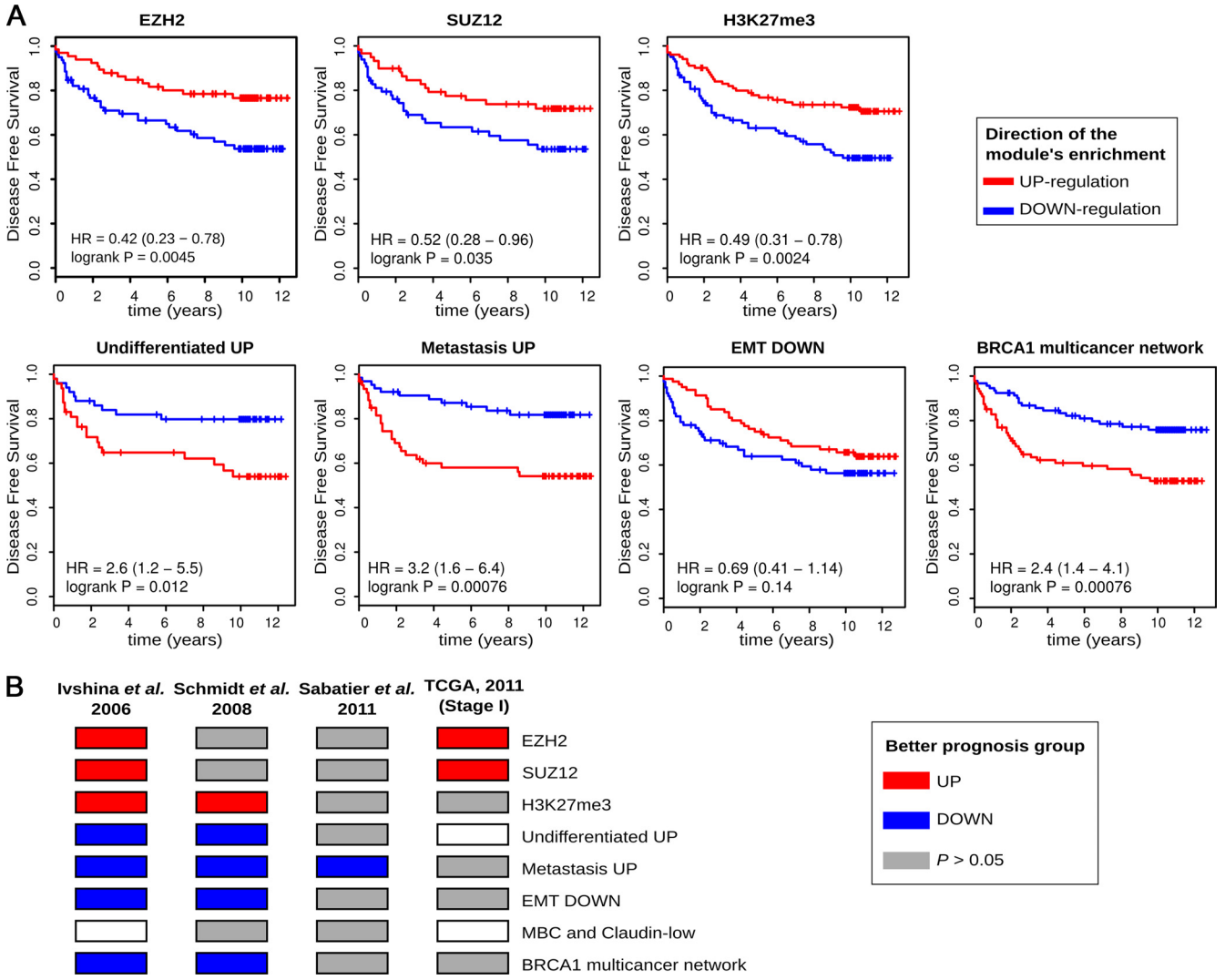

FIG 4 PRC2 regulatory modules predict outcome in breast cancer patients. (A) Kaplan-Meier curves for PRC2 modules and prognostic gene signatures in the data set from Ivshina et al. (30). HR, hazard ratio. (B) Survival analysis results after sample stratification according to PRC2 enrichment. The survival analysis results for the data set from Ivshina et al. (30) in panel A and three additional breast cancer cohorts (31, 56; TCGA, 2011) are presented for samples upregulated for the module (e.g., a PRC2 module or prognostic gene module) compared to those for samples that are downregulated for the module. Red and blue indicate that the up- or downregulation of that module, respectively, significantly predicts better outcome (log rank test, $P < 0.05$). Nonsignificant results are denoted in gray; missing analysis due to small sample group size is indicated in white. Prognostic gene signatures are as in Fig. 2.

cant overlap with developmental GO categories (see Fig. S7). Therefore, our comparative analysis shows that the PRC2 modules represent independent characteristics.

Genes that predict survival in breast cancer are directly regulated by EZH2. While the large group of PRC2 target genes collectively is linked to tumor characteristics, we have identified the genes that most contributed to the stratification of breast tumors and, thus, which differential expression could be a biomarker of patient outcome (see Table S9 in the supplemental material). The top genes represent six neuronal genes (*PDE4D*, *ADRA1A*, *POU4F2*, *SIM2*, *NEUROD2*, and *NTRK1*) and two cell-signaling genes (*PHOX2B* and *ULBP1*). RT-PCR experiments showed a high increase in the level of five out of eight genes, specifically in MCF7 cells depleted for *EZH2* (Fig. 5A), suggesting that the expression levels of these five genes directly depend on the level of EZH2. On the contrary, the transcript levels of these genes generally decreased in the patients with increased *EZH2* mRNA (Fig. 5B). Therefore, the genes that can be considered a proxy for the

MCF7 cells grown on Matrigel. Phase-contrast micrographs of *EZH2* shRNA acini are shown, with the relative number of acini (>20 cells) presented in a graph. Error bars, means + standard errors of the means (SEM), $n = 2$. (D) *EZH2* shRNA-mediated knockdown results in decreased proliferation compared to that in cells treated with control shRNA. Error bars, means + SEM, $n = 3$. (E) Control shRNA acini and *EZH2* shRNA acini examined by fluorescence microscope for expression of EMT markers. Scale bar, 50 μm. (F) *EZH2* depletion in MCF7 cells changes expression of EMT markers. RT-qPCR data are shown for cells grown in Dox^- versus Dox^+ media. (G) *EZH2* depletion in MDA-MB-231 cells changes expression of EMT markers. RT-qPCR data are shown for cells grown in Dox^+ versus Dox^- media (in the Tet-On configuration). Error bars for all RT-qPCR assays, means + SEM, $n = 3$.

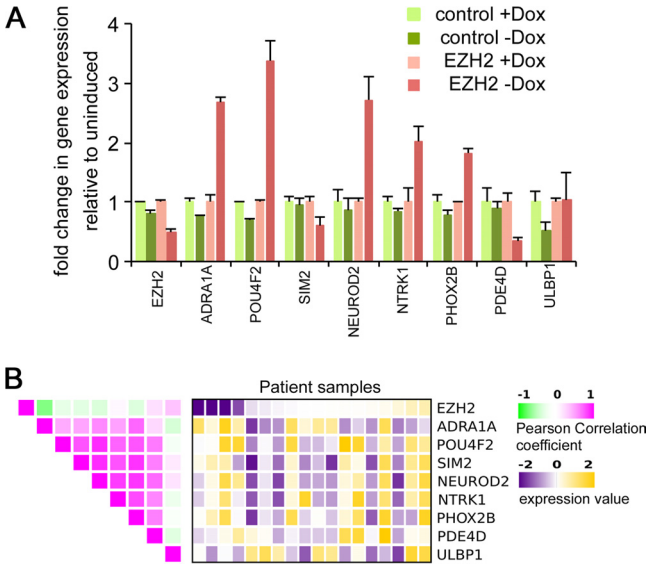

FIG 5 Genes from PRC2 modules the expression of which stratifies breast tumor samples are regulated by EZH2. (A) RT-qPCR was performed for *EZH2* and the top 8 PRC2 module genes on RNA from MCF7 cells and graphed relative to expression in cells without shRNA production. Error bars, means + SEM, $n = 3$. (B) Expression levels of *EZH2* and 8 top PRC2 module genes in breast tumors (TissueScan Array). The heat map shows expression values from RT-qPCR experiments in a panel of human tumor samples, normalized to the reference gene *B2M* and presented relative to the mean threshold cycle (ΔC_T) of each gene. Heat maps were produced using Gitools. Yellow and purple colors indicate higher and lower expression, respectively. Columns represent different patient samples. Samples are arranged according to the *EZH2* level. The correlation coefficient of the expression values of 8 genes (left panel) shows strong association with the level of 6 of these genes.

behavior of genes associated with PRC2 modules are likely to be under direct control of EZH2.

High EZH2 protein levels are associated with aggressive breast cancer phenotypes. The lower expression of PRC2 modules in more aggressive breast tumors led us to hypothesize that EZH2 protein level is increased in these tumors. To confirm this, we determined the expression of EZH2 in a diverse breast tumor sample cohort (95 patients, $n = 450$ samples) by immunohistochemistry using high-density tissue microarrays (Fig. 6A and B). Increased EZH2 expression was significantly associated with basal-like tumor subtypes ($P < 8e{-}07$, Fisher's two-tailed test) but not with tumor grade and patient survival. Consistent with the EZH2 IHC results, the samples with high EZH2 protein expression displayed increased *EZH2* transcript level (Fig. 6C). They generally had lower expression of the top genes (Fig. 6C), with the five genes that demonstrated derepression upon *EZH2* depletion (Fig. 5A) showing the lowest expression level.

DISCUSSION

The prevailing view of the cancer genome is that it arises through sequential genetic mutations, with each mutation accounting for specific tumor properties (i.e., increased cell proliferation, invasiveness, metastasis, and drug resistance) supporting selective outgrowth of a monoclonal cell population. Our analysis of publicly available ChIP-seq and microarray data, supported by the analysis of tissue microarrays and breast cancer cell lines with gene knockdowns, provides evidence that in aggressive breast cancer cases, a critical shift in gene regulation occurs in genes regulated by PRC2 activity. This may be contributing to changed demands in the cell cycle, developmental pathways, cell motility, and EMT. We propose that highly metastatic behavior of breast tumors is sustained by lower expression of PRC2 modules, as shRNA-mediated knockdown of EZH2 in breast cancer cells reactivated the top PRC2 module genes that predict cancer survival, as well as reduced cell proliferation, upregulated epithelial markers, and suppressed mesenchymal markers.

The rationale behind using gene expression to stratify patients for cancer prognosis came from the repeated observations that intrinsic biology seems to play a more important role than other variables, such as age or tumor stage, in determining the breast cancer phenotype (50). In our study, over a thousand of the genes that are targets of PRC2 stratified breast cancer patients into two clinically different groups and predicted survival. Their level of expression had high prognostic value independently of known molecular and pathological characteristics or cancer gene signatures and was changed as early as in stage I tumors, indicating its utility in predicting disease progression in patients with early stage cancer. Samples with relatively low expression of PRC2 modules had expression signatures common to embryonic tissues and adult tissues containing progenitor cells or moderately differentiated cells (compare Fig. 2 with Fig. 1). In an oncogenic context, overexpression of PRC2 complex components has been linked to a transition from a quiescent state into an actively dividing, phenotype more resembling the stem cell. Activation of the stem cell-like signature is common to tumors with low differentiation and along with the module associated with c-MYC predicts cancer outcome (51, 52). These studies proposed that the third module accounting for similarities between cancer cells and stem cells is the PRC2 module. Using SLEA, we were able to show that low expression of PRC2 module was associated with the most aggressive breast cancers.

While the large group of PRC2 target genes collectively is linked to tumor characteristics, we have identified eight genes for which differential expression could be a biomarker of patient outcome. The PRC2 complex is known to repress genes that contribute to cell differentiation (4), and the identification of neuronal genes as the top PRC2 module patient stratifiers is probably a reflection of the fact that a large proportion of neuronal differentiation genes are bound by PRC2 and remain repressed (14). Importantly, we detected a highly significant association of all studied PRC2 modules, not of the EZH2 protein itself, with tumor grade and patient survival. A plausible possibility is that dominant somatic mutations in the *EZH2* gene (53) contributed to lower expression of the EZH2 module in some samples. EZH2 was recently described as a transcriptional activator, where its function is independent of PRC2 (54). In the breast cancer data sets that we analyzed, we found only a few samples with significantly changed expression of the PRC2-independent EZH2 module, suggesting that this module cannot be used for patient stratification.

Our analysis has been limited by the public availability of large transcriptome data sets with survival information; the method is highly study dependent but provides a powerful means to investigate molecular subtypes in tumors without the need for prior knowledge. Since the SLEA method uses median-centered expression values, it is sensitive to the heterogeneity of the data set, and

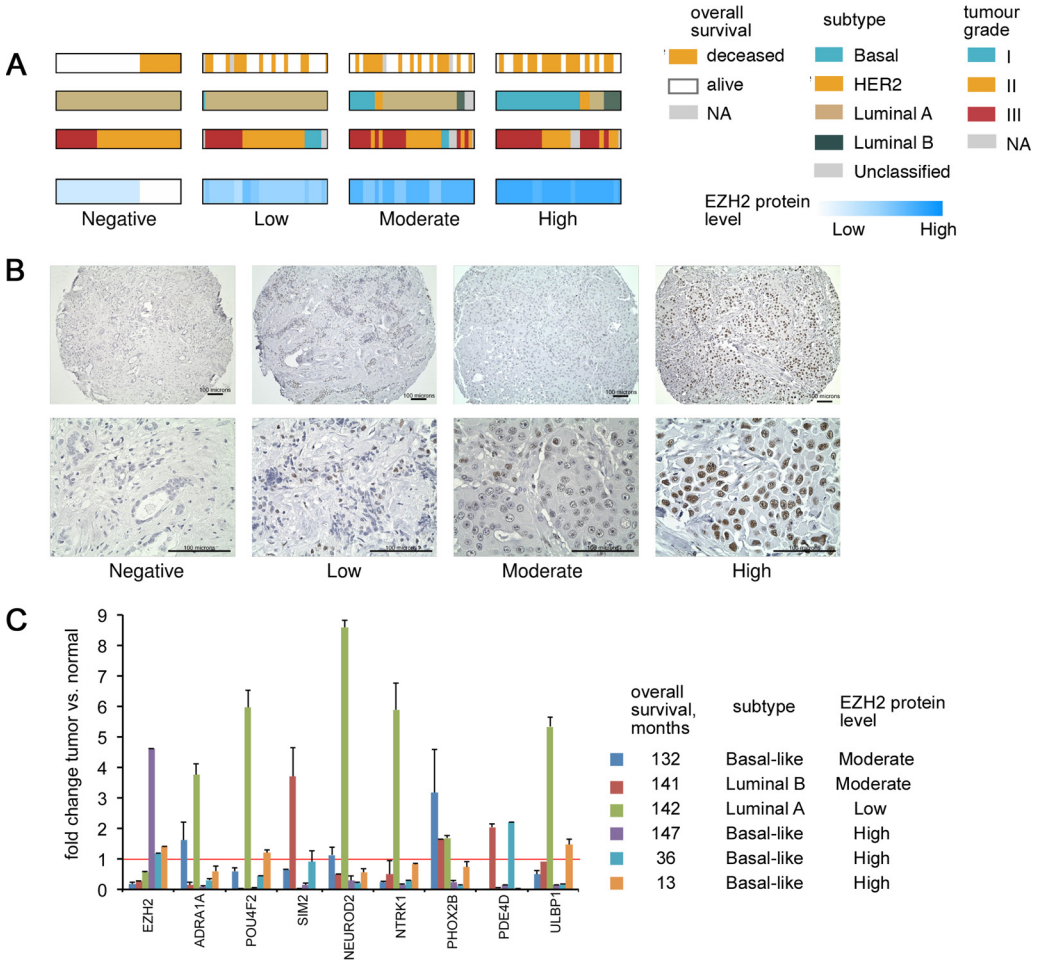

FIG 6 EZH2 protein levels in breast tumors correlate with the most aggressive intrinsic subtypes. (A) Analysis of tissue microarray data. High-density tissue microarray containing in total 450 spots of normal and cancerous (190 spots of ductal carcinoma *in situ* [DCIS] and invasive carcinomas) biopsy tissues was analyzed for EZH2 protein levels by immunohistochemistry. A total of 63.7% of the patients had estrogen receptor-positive (ER^+) tumors, 50.6% had progesterone receptor-positive tumors, and 21.2% had $HER2^+$ tumors. The distribution of breast cancer subtypes was basal-like (28.5%), $HER2^+/ER^-$ (4.2%), luminal A (58.9%), luminal B (6.3%), and unclassified (2.1%). A majority of the patients had high-grade tumors (46.0% grade II and 47.1% grade III). EZH2 protein levels (in blue) were divided in four categories and are shown with distribution of tumor subtypes, grade, and overall survival. (B) Representative images of biopsy specimens of invasive breast carcinomas with different EZH2 levels. Scale bar, 100 μm. (C) The transcript level of PRC2 module genes negatively correlates with the *EZH2* transcript and protein level. The expression level of *EZH2* and the 8 top PRC2 module genes was confirmed by RT-qPCR. Gene expression of each gene was normalized to three control genes. Error bars, means + SEM, $n = 3$ (PCR repeats). The table on the right shows EZH2 staining results and patient data.

the overrepresentation of high- or low-grade samples can potentially affect the results. As such, the PRC2 module was anticorrelated with survival in the data set from Ivshina et al. (30) but not in that from Sabatier et al. (31) nor that from TCGA, which are biased toward higher-grade tumors. We speculate that the PRC2 module has a predictive power in other cancers, since breast tumors tend to be diagnosed at earlier stages than many other solid tumors. It is interesting to note that the EMT, a process in which PRC2 may play a significant role, is a very early event in breast tumors, but it occurs significantly later in some other cancer subtypes (55). It is important also to note that our study reflects intratumor variability, since normal samples were removed from the original cohorts. Our aim was to determine the ability of loci associated with H3K27 methylation to stratify tumor subtypes, regardless of how altered was gene expression compared to that in a nontumorigenic reference. Due to the nature of SLEA, our analysis is sensitive to gene expression changes across samples but does not take into account absolute values. In particular, a set of tumors that are enriched for the overexpressed PRC2 module does not necessarily have high transcription rates for the genes in the PRC2 module.

It is likely that genes directly bound by the PRC2 components become underexpressed through acquisition of high H3K27 trimethylation. Importantly, as a function linked to the catalytic activity of the PRC2 complex, it may be targeted with a small molecule inhibitor. This study gives rise to important questions regarding the mechanisms by which the genes in the PRC2 module become coactivated or corepressed. An attractive model by which EZH2 can globally reprogram gene expression is by tethering its target gene loci into chromatin domains, which is assisted by other chromatin regulators as well as long intergenic noncoding RNAs (lincRNAs).

ACKNOWLEDGMENTS

The project was supported by grants SAF2009-06954 and SAF2012-36199 rom the Spanish Ministry of Science and Technology (N.L.-B.) and R01 CA138631 (E.V.B.), R01 CA142996, and P50 CA125183 (O.I.O.) from the National Institutes of Health.

REFERENCES

1. **Valk-Lingbeek ME, Bruggeman SWM, van Lohuizen M.** 2004. Stem cells and cancer; the Polycomb connection. Cell **118:**409–418.
2. **Feinberg AP, Ohlsson R, Henikoff S.** 2006. The epigenetic progenitor origin of human cancer. Nat. Rev. Genet. **7:**21–33.
3. **Sauvageau M, Sauvageau G.** 2010. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. Cell Stem Cell **7:**299–313.
4. **Sparmann A, van Lohuizen M.** 2006. Polycomb silencers control cell fate, development and cancer. Nat. Rev. Cancer **6:**846–856.
5. **Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, Ghosh D, Sewalt RG, Otte AP, Hayes DF, Sabel MS, Livant D, Weiss SJ, Rubin MA, Chinnaiyan AM.** 2003. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc. Natl. Acad. Sci. U. S. A. **100:**11606–11611.
6. **Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, Rubin MA, Chinnaiyan AM.** 2002. The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature **419:**624–629.
7. **Bachmann IM.** 2006. EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. J. Clin. Oncol. **24:**268–273.
8. **Niida A, Smith AD, Imoto S, Aburatani H, Zhang MQ, Akiyama T.** 2009. Gene set-based module discovery in the breast cancer transcriptome. BMC Bioinformatics **10:**71. doi:10.1186/1471-2105-10-71.
9. **Chang C-J, Yang J-Y, Xia W, Chen C-T, Xie X, Chao C-H, Woodward WA, Hsu J-M, Hortobagyi GN, Hung M-C.** 2011. EZH2 promotes expansion of breast tumor initiating cells through activation of RAF1-β-catenin signaling. Cancer Cell **19:**86–100.
10. **Alford SH, Toy K, Merajver SD, Kleer CG.** 2012. Increased risk for distant metastasis in patients with familial early-stage breast cancer and high EZH2 expression. Breast Cancer Res. Treat. **132:**429–437.
11. **Min J, Zaslavsky A, Fedele G, McLaughlin SK, Reczek EE, De Raedt T, Guney I, Strochlic DE, MacConaill LE, Beroukhim R, Bronson RT, Ryeom S, Hahn WC, Loda M, Cichowski K.** 2010. An oncogene-tumor suppressor cascade drives metastatic prostate cancer by coordinately activating Ras and nuclear factor-κB. Nat. Med. **16:**286–294.
12. **Ding L, Erdmann C, Chinnaiyan AM, Merajver SD, Kleer CG.** 2006. Identification of EZH2 as a molecular marker for a precancerous state in morphologically normal breast tissues. Cancer Res. **66:**4095–4099.
13. **Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Bell GW, Otte AP, Vidal M, Gifford DK, Young RA, Jaenisch R.** 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature **441:**349–353.
14. **Bracken AP.** 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. **20:**1123–1136.
15. **Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA.** 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell **122:**947–956.
16. **Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K-I, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA.** 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell **125:**301–313.
17. **McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, Liu Y, Graves AP, Pietra Della A, III, Diaz E, LaFrance LV, Mellinger M, Duquenne C, Tian X, Kruger RG, McHugh CF, Brandt M, Miller WH, Dhanak D, Verma SK, Tummino PJ, Creasy CL.** 2012. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature **492:**108–112.
18. **Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MCU, Nielsen TO, Moorman PG, Earp HS, Millikan RC.** 2006. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA **295:**2492–2502.
19. **Carr JR, Kiefer MM, Park HJ, Li J, Wang Z, Fontanarosa J, Dewaal D, Kopanja D, Benevolenskaya EV, Guzman G, Raychaudhuri P.** 2012. FoxM1 regulates mammary luminal cell fate. Cell Rep. **1:**715–729.
20. **Debnath J, Muthuswamy SK, Brugge JS.** 2003. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. Methods **30:**256–268.
21. **Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE.** 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. **4:**e1000242. doi:10.1371/journal.pgen.1000242.
22. **Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Stamatoyannopoulos JA, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SCJ, Sabo PJ, Sandstrom R, Shafer A, et al.** 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447:**799–816.
23. **Langmead B, Trapnell C, Pop M, Salzberg SL.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10:**R25. doi:10.1186/gb-2009-10-3-r25.
24. **Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Liu XS.** 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol. **9:**R137. doi:10.1186/gb-2008-9-9-r137.
25. **Althammer S, González-Vallinas J, BallarŽ C, Beato M, Eyras E.** 2011. Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. Bioinformatics **27:**3333–3340.
26. **Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W.** 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics **25:**1952–1958.
27. **Quinlan AR, Hall IM.** 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26:**841–842.
28. **Shin H, Liu T, Manrai AK, Liu XS.** 2009. CEAS: cis-regulatory element annotation system. Bioinformatics **25:**2605–2606.
29. **Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP.** 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. **102:**15545–15550.
30. **Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J, Lindahl T, Pawitan Y, Hall P, Nordgren H, Wong JEL, Liu ET, Bergh J, Kuznetsov VA, Miller LD.** 2006. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res **66:**10292–10301.
31. **Sabatier R, Finetti P, Cervera N, Lambaudie E, Esterni B, Mamessier E, Tallet A, Chabannon C, Extra J-M, Jacquemier J, Viens P, Birnbaum D, Bertucci F.** 2010. A gene expression signature identifies two prognostic subgroups of basal breast cancer. Breast Cancer Res. Treat. **126:**407–420.
32. **Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M.** 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. **40:**D109–D114.
33. **Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G.** 2000. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. **25:**25–29.
34. **Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB.** 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. U. S. A. **101:**6062–6067.
35. **Roth RB, Hevezi P, Lee J, Willhite D, Lechner SM, Foster AC, Zlotnik A.** 2006. Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics **7:**67–80.
36. **Schmidt M, Bohm D, von Torne C, Steiner E, Puhl A, Pilch H, Lehr**

HA, Hengstler JG, Kolbl H, Gehrmann M. 2008. The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res. **68:**5405–5413.

37. **Pawitan Y, Bjöhle J, Amler L, Borg A-L, Egyhazi S, Hall P, Han X, Holmberg L, Huang F, Klaar S, Liu ET, Miller L, Nordgren H, Ploner A, Sandelin K, Shaw PM, Smeds J, Skoog L, Wedrén S, Bergh J.** 2005. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res. **7:**R953. doi:10.1186/bcr1325.
38. **Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, Fulton LL, Dooling DJ, Ding L, Mardis ER, Wilson RK, Ally A, Balasundaram M, Butterfield YSN, Carlsen R, Carter C, Chu A, Chuah E, Chun H-JE, Coope RJN, Dhalla N, Guin R, Hirst C, Hirst M, Holt RA, Lee D, Li HI, Mayo M, Moore RA, Mungall AJ, Pleasance E, Gordon Robertson A, Schein JE, Shafiei A, Sipahimalani P, Slobodan JR, Stoll D, Tam A, Thiessen N, Varhol RJ, Wye N, Zeng T, Zhao Y, Birol I, Jones SJM, Marra MA, Cherniack AD, Saksena G, Onofrio RC, Pho NH, et al.** 2012. Comprehensive molecular portraits of human breast tumours. Nature **490:**61–70.
39. **Gundem G, Lopez-Bigas N.** 2012. Sample-level enrichment analysis unravels shared stress phenotypes among multiple cancer types. Genome Med. **4:**28. doi:10.1186/gm327.
40. **Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J.** 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. **5:**R80. doi:10.1186/gb-2004-5-10-r80.
41. **Perez-Llamas C, Lopez-Bigas N.** 2011. Gitools: analysis and visualisation of genomic data using interactive heat-maps. PLoS One **6:**e19541. doi:10.1371/journal.pone.0019541.
42. **Lopez-Bigas N, De S, Teichmann SA.** 2008. Functional protein divergence in the evolution of Homo sapiens. Genome Biol. **9:**R33. doi:10.1186/gb-2008-9-2-r33.
43. **Benjamini Y, Hochberg Y.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. Ser. B **57:**289–300.
44. **Nijhawan D, Zack TI, Ren Y, Strickland MR, Lamothe R, Schumacher SE, Tsherniak A, Besche HC, Rosenbluh J, Shehata S, Cowley GS, Weir BA, Goldberg AL, Mesirov JP, Root DE, Bhatia SN, Beroukhim R, Hahn WC.** 2012. Cancer vulnerabilities unveiled by genomic loss. Cell **150:**842–854.
45. **van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH.** 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature **415:**530–536.
46. **Gonzalez ME, DuPrie ML, Krueger H, Merajver SD, Ventura AC, Toy KA, Kleer CG.** 2011. Histone methyltransferase EZH2 induces Akt-dependent genomic instability and BRCA1 inhibition in breast cancer. Cancer Res. **71:**2360–2370.
47. **Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai M-C, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY.** 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature **464:**1071–1076.
48. **Lee ST, Li Z, Wu Z, Aau M, Guan P, Karuturi RKM, Liou YC, Yu Q.** 2011. Context-specific regulation of NF-κB target gene expression by EZH2 in breast cancers. Mol. Cell **43:**798–810.
49. **Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, Botstein D.** 2002. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol. Biol. Cell **13:**1977–2000.
50. **Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale A-L, Botstein D.** 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc. Natl. Acad. Sci. U. S. A. **100:**8418–8423.
51. **Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, Weinberg RA.** 2008. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat. Genet. **40:**499–507.
52. **Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG, Cantor AB, Orkin SH.** 2010. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell **143:**313–324.
53. **Yap DB, Chu J, Berg T, Schapira M, Cheng S-WG, Moradian A, Morin RD, Mungall AJ, Meissner B, Boyle M, Marquez VE, Marra MA, Gascoyne RD, Humphries RK, Arrowsmith CH, Morin GB, Aparicio SA.** 2011. Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. Blood **117:**2451–2459.
54. **Xu K, Wu ZJ, Groner AC, He HH, Cai C, Lis RT, Wu X, Stack EC, Loda M, Liu T, Xu H, Cato L, Thornton JE, Gregory RI, Morrissey C, Vessella RL, Montironi R, Magi-Galluzzi C, Kantoff PW, Balk SP, Liu XS, Brown M.** 2012. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. Science **338:**1465–1469.
55. **Kim H, Watkinson J, Varadan V, Anastassiou D.** 2010. Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving INHBA, THBS2 and COL11A1. BMC Med. Genomics **3:**51. doi:10.1186/1755-8794-3-51.
56. **Schmidt M, Bohm D, von Torne C, Steiner E, Puhl A, Pilch H, Lehr HA, Hengstler JG, Kolbl H, Gehrmann M.** 2008. The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res **68:**5405–5413.