

SCIENTIFIC REPORTS



OPEN

Epigenetic alterations to Polycomb targets precede malignant transition in a mouse model of breast cancer

Ying Cai¹, Jhih-Rong Lin¹, Quanwei Zhang¹, Kelly O'Brien¹, Cristina Montagna^{1,2} & Zhengdong D. Zhang¹

Malignant breast cancer remains a major health threat to women of all ages worldwide and epigenetic variations on DNA methylation have been widely reported in cancers of different types. We profiled DNA methylation with ERRBS (Enhanced Reduced Representation Bisulfite Sequencing) across four main stages of tumor progression in the MMTV-PyMT mouse model (hyperplasia, adenoma/mammary intraepithelial neoplasia, early carcinoma and late carcinoma), during which malignant transition occurs. We identified a large number of differentially methylated cytosines (DMCs) in tumors relative to age-matched normal mammary glands from FVB mice. Despite similarities, the methylation differences of the premalignant stages were distinct from the malignant ones. Many differentially methylated loci were preserved from the first to the last stage throughout tumor progression. Genes affected by methylation gains were enriched in Polycomb repressive complex 2 (PRC2) targets, which may present biomarkers for early diagnosis and targets for treatment.

DNA methylation is one of the epigenetic marks that regulate gene expression¹. Alterations in DNA methylation are commonly observed in various cancer types^{2–5} and a well-known signature of the cancer genome is global hypomethylation accompanied by promoter hypermethylation of some tumor suppressor genes^{6–11}. Some genes in tumors showed distinct hypomethylation compared to normal cells¹⁰ and cancer progenitor cells with such DNA methylation alterations may help to predict cancer risks¹¹. Global hypomethylation has been proposed as a mechanism driving chromosomal instability and elevated mutation rates^{12,13} and hypermethylation caused gene suppression, such as *HAND2*, is suggested to contribute to cancer development¹⁴; hence DNA methylation abnormalities have been considered to play a causal role in tumorigenesis¹⁵.

Despite extensive profiling of gene expression and DNA methylation changes in breast cancer, not much is known about the dynamics of DNA methylation during cancer progression. It is almost impossible to monitor tumor progression by following the same patients, a task that can be easily fulfilled by using mouse models with almost identical genetic backgrounds. A better understanding of DNA methylation dynamics in breast carcinogenesis is not only vital to explain transcriptional deregulation of gene expression during tumor progression, but also adds to the understanding of tumor class and subtypes^{16,17} and prognosis¹⁸. Abnormal DNA methylation has been reported in breast carcinoma with subtype specific patterns¹⁹. In particular, the more aggressive luminal B subtype shows higher DNA methylation when compared with other subtypes and normal tissues^{19,20}; and reduced gene expression correspondingly²⁰.

Human breast cancer begins with premalignant atypical ductal hyperplasia (ADH), moves to the ductal carcinoma *in situ* (DCIS), and progresses to invasive ductal carcinoma (IDC)^{21,22}. Despite extensive evidences suggesting that ADH and DCIS are precursors of IDC, DNA methylation biomarkers identified at the early stages of breast cancer are limited. This represents a significant gap of knowledge in the field since such biomarkers may explain the biological basis of tumor progression and helps to predict outcomes. A recent study analyzing DCIS and IBC (invasive breast carcinoma) samples identified only 18 CpG loci associated with survival and prognosis of breast cancer patients¹⁸. Another study using FEA (flat epithelial atypia), ADH, DCIS, and IDC

¹Department of Genetics, Albert Einstein College of Medicine, Bronx, New York, USA. ²Department of Pathology, Albert Einstein College of Medicine, Bronx, New York, USA. Correspondence and requests for materials should be addressed to Z.D.Z. (email: zhengdong.zhang@einstein.yu.edu)

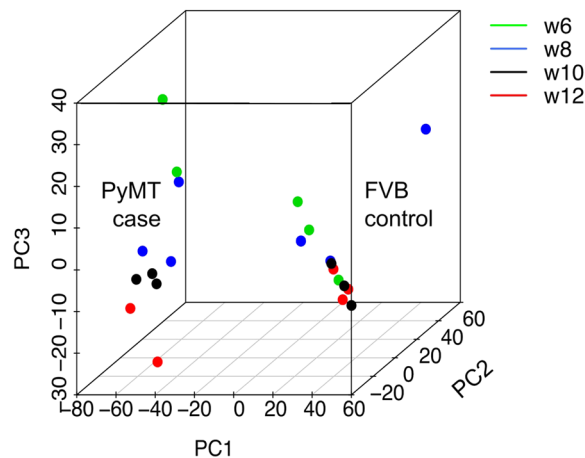


Figure 1. Principal component analysis of DNA methylation profiles of both PyMT and FVB control samples. Samples were taken at four time points: weeks 6, 8, 10, and 12 – shown as w6, w8, w10, and w12 – correspond to hyperplasia, adenoma/MIN, early carcinoma and late carcinoma in PyMT mice.

Stage	Hypomethylation	Hypermethylation	Total
Hyperplasia	70,323	30,187	100,510
Adenoma/MIN	79,147	34,900	114,047
Early carcinoma	97,436	48,665	146,101
Late carcinoma	100,148	49,570	149,718

Table 1. Total DMCs.

found increased number of genes with elevated promoter methylation during the progression²³. Although it has been discovered that normal tissue adjacent to breast cancer already exhibited numerous DNA methylation alterations²⁴, how those alterations change during progression is not well defined yet. Thus, it is apparent that early DNA methylation changes are relevant to breast carcinogenesis but their characterization is lacking. It is critical to comprehensively examine DNA methylation during the early stages of breast cancer progression.

The MMTV-PyMT transgenic mouse is an ideal model to study tumor progression because primary tumors evolve from pre-malignancy to an invasive malignant tumor through four stereotypical stages of progression (hyperplasia, adenoma/mammary intraepithelial neoplasia (MIN), early carcinoma and late carcinoma). Breast tumors in the PyMT mouse share both morphological and transcriptional similarities with human breast tumors²⁵, and generally cluster with the human luminal B subtype^{26,27}. Previous cytogenetic profiling also showed similarity of breast tumor from PyMT mice to human breast carcinoma on the level of chromosomal structural and numerical abnormalities^{28,29}.

For the first time, we report the dynamics of DNA methylation deregulation from early hyperplasia to late carcinoma across the four main stages of breast tumor progression. We performed ERRBS (Enhanced Reduced Representation Bisulfite Sequencing), which quantitatively measures DNA methylation at the base-pair resolution and increases coverage on CpG sites relative to RRBS³. We observed, as expected, a global hypomethylation and a shift towards hypermethylation as the tumors progressed. We found genes with promoter hypermethylation enriched in Polycomb repressive complex 2 (PRC2) targets throughout tumor progression. These shed lights on the mechanisms of abnormal DNA methylation, using the PyMT mouse model of breast cancer progression.

Results

DNA methylation profiles of the PyMT mouse model unveil DNA methylation changes precede malignant transition.

We first profiled DNA methylation alterations in hyperplasia, adenoma/MIN, early carcinoma and late carcinoma samples from the PyMT mice and normal tissues from FVB controls with principal component analysis (PCA). After removing two outlier samples, the PyMT samples clearly separated from the FVB controls, as shown in Fig. 1. In addition, the tumor samples separate across their temporal stages, whereas the controls remain within one cluster. This pattern is consistent with the PCA results from gene expression profiles of this mouse model as we reported previously³⁰. Only biological covariates but not technical ones contribute to PCs (Supplementary Fig. S1), suggesting the variance observed here was not due to batch effect but was of biological meanings.

As tumors progressed from hyperplasia to late carcinoma, the number of DMCs (methylation difference $\geq 10\%$ and q -value < 0.01 , Methods) increased for both hyper- and hypo-methylation, but hypomethylated loci were consistently more frequent (Table 1 and Fig. 2A). Interestingly, we observed a shift toward hypermethylation by roughly 3% (30.0% in hyperplasia to 33.1% in late carcinoma, DMC-hypermethylation

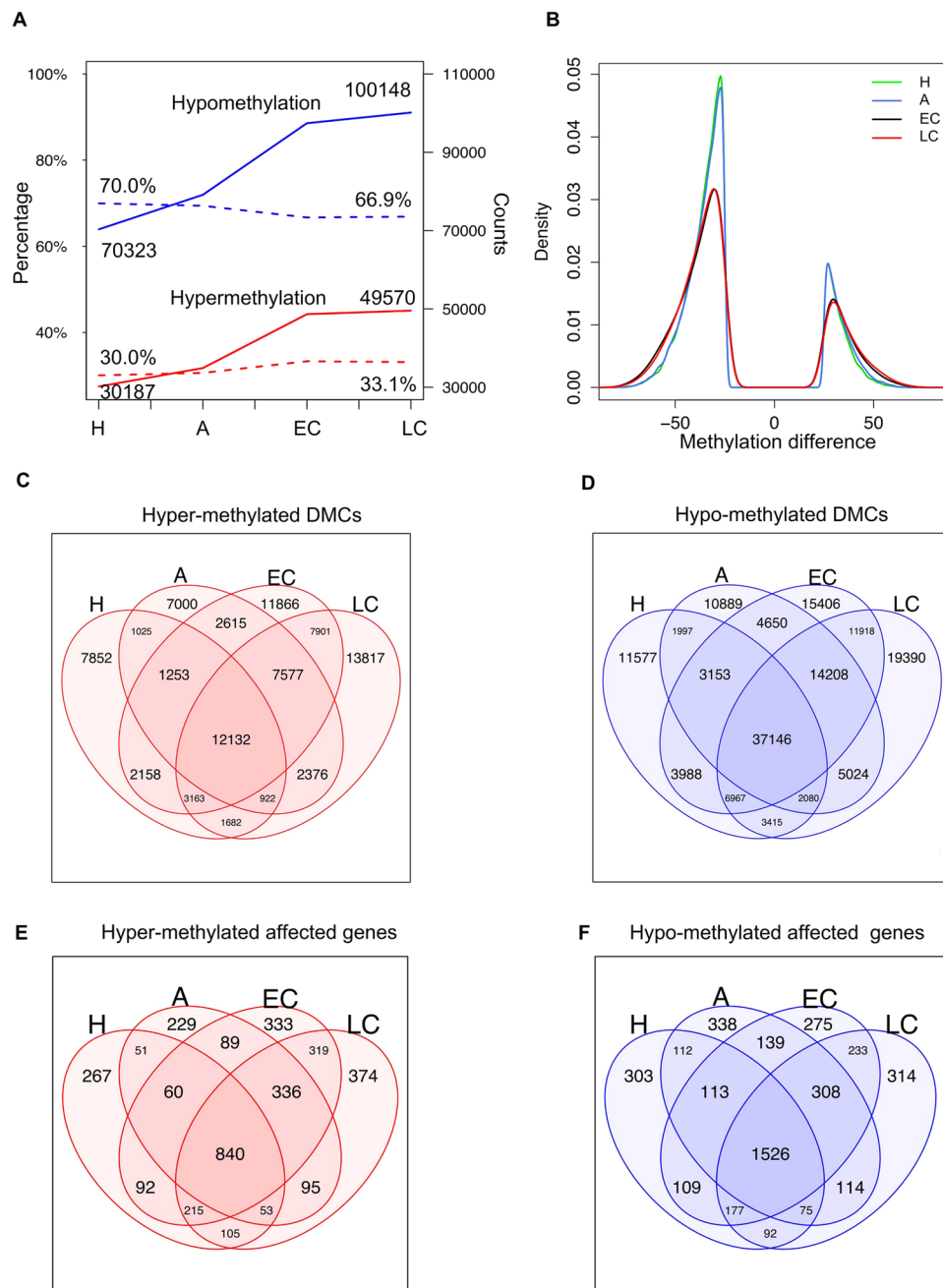


Figure 2. DMCs at different stages of tumor progression. **(A)** Numbers of DMCs. Counts and percentages are connected by solid and dashed lines, hypermethylation and hypomethylation are represented by red and blue lines; w6, w8, w10, w12 as week 6, week 8, week 10 and 12, corresponding to hyperplasia, adenoma/MIN, early carcinoma and late carcinoma. **(B)** DMCs (methylation difference $\geq 25\%$ and q -value < 0.01) methylation difference distribution at four time stages. **(C,D)** Intersections of DMCs. **(E,F)** Intersections of genes whose promoter contain DMCs. H = hyperplasia, A = adenoma/MIN, EC = early carcinoma, LC = late carcinoma.

percentage + DMC-hypomethylation percentage = 1) (Fig. 2A). The late stages (early carcinoma and late carcinoma) were similar on methylation patterns yet distinct from early stages (Fig. 2B).

To investigate impact of DNA methylation on gene expression and the possible gene affected, we mapped DMCs to the *cis*-regulatory elements of promoters (± 2 kb to RefSeq TSS) and enhancers (if the DMCs overlapped with predicted enhancer regions, Methods). Consistent with the high preservation of DMCs between neighboring stages and among all stages (Fig. 2C,D), we found that gene promoters overlapping with these DMCs were largely the same (Fig. 2E,F). These observations provide evidence that DNA methylation changes at late stages can be traced back to early lesions. Interestingly, using curated gene lists from MSigDB (Molecular Signatures Database) and GSEA (gene set enrichment analysis)³¹, we found an enrichment of PRC2 targets (FDR < 0.05) among the 374 genes with promoter hyper-DMCs unique to late carcinoma stage (*Barhl2*, *Cacna1e*, *Cntfr*, *Col4a6*,

Dkk2, Dok6, Elmod1, Fezf2, Fli1, Fzd10, Gabra2, Gdf7, Gria2, Gsc, Hlx, Hoxb7, Hoxb8, Hoxc12, Hoxd13, Hrk, Isl1, Lhx8, Neurod1, Nkx2-1, Nkx2-2, Nkx3-2, Nkx6-2, Otop3, Pdx1, Ripk3, Sctr, Tall1 and Wt1) and among the 840 genes with promoter hyper-DMCs common to all four stages (including *Alx3, Cited1, Comp, Coro6, Crlf1, Csmn1, Cyp26c1, Dgki, Dll4, Duoxa1, Eomes, Esam, Gdnf, Hes7, Hhex, Hhip, Hmx2, Hoxc5, Hoxc6, Hsf4, Il1rap12, Lrfn5, Ltk, Mab21l1, Nfix, Nkx2-8, Nrg1, Otx1, Pax7, Podn, Ptgdr, Rasl10a, Sidt1, Slc10a4, Slc1a2, Sox7, Tbr1, Tcea3, Tmem59l, Tmem88, Tp73, Ucn* and *Ush1g*). The former are potential biomarkers that distinguish late carcinoma from early stages and the latter can be potential biomarkers for early cancer detection. The distributions of CpGs captured and DMCs mapped to genomic compartments were similar across different stages. About 10% of all CpGs captured and about 26% (26% at hyperplasia, 23% at late carcinoma) of DMCs were mapped to putative enhancers (Supplementary Fig. S2A), suggesting DMCs were enriched in enhancers (permutation test p -value < 0.05 , Supplementary Fig. S2B, Methods). About 14% of DMCs and 28% of hypermethylated DMCs were mapped to promoters (Supplementary Fig. S2C), suggesting hypermethylation is biased toward promoters (permutation test p -value < 0.05 , Supplementary Fig. S2D). Not many hypomethylated DMCs (and DMCs overall) were detected in the promoter regions, possibly due to their low methylation in general.

We then analyzed the pathway enrichment among genes linked to promoters and enhancers with hypermethylated DMCs (Fig. 3A,B), assuming nearest genes within 10 kb of an enhancer as its targets. Genes with hypermethylated DMCs in promoters were enriched in targets of PRC2, EED, and SUZ12 (Fig. 3A, FDR < 0.05). Putative targets of hypermethylated enhancers were also enriched in PRC2 targets (Fig. 3B, FDR < 0.05). We further observed a significant reduced expression of such PRC2 target genes with promoter hyper-DMCs comparing to all PRC2 targets in all four stages (Fig. 3C, Wilcoxon-test p -value < 0.05). In addition, three core components of PRC2 (EZH2, EED and SUZ12) all exhibited higher mRNA expression in PyMT samples than in controls (Fig. 3D). Two core components of PRC2 (EZH2 and EED) are E2F targets and we observed higher expression of E2Fs in the PyMT mice as well (Fig. 3E).

To further study how gene expression is affected by DNA methylation in each cancer developmental stage, we analyzed together genes with promoter DMCs identified in this study and differentially expressed genes (FDR < 0.05) that we identified earlier in our study of mRNA expression of this PyMT mouse model³⁰. Comparing PyMT samples from the four stages with age-matched normal tissues from controls, we identified 101, 109, 159, and 198 genes, respectively, with both hypomethylated DMCs in promoter and increased gene expression and 131, 152, 216 and 249 genes with both hypermethylated DMCs in promoter and reduced gene expression. Down-regulated genes with promoter hypermethylation were enriched with genes involved in epithelial mesenchymal transition (EMT) in week 10 (FDR < 0.05 , *Tagln, Sgcb, Col4a2, Wipf1, Tpm2, Igfbp3, Mest, Col1a1, Tpm1, Mmp14, Fstl1, Fap, Plod1*) and week 12 (FDR < 0.05 , *Htra1, Mest, Slit3, Timp3, Tagln, Col4a1, Sgcb, Dpysl3, Tpm2, Col1a1, Wnt5a, Col1a2, Emp3, Fap, Nid2, Fstl1, Tpm1*). Text mining of PubMed literature with an R package RISmed revealed that most of them are with unclear roles in breast cancer except for *Igfbp3, Mmp14, Fap, Timp3* and *Wnt5a*.

Time-course analysis of DMRs reveals increasing methylation differences as tumor progress.

Since a dense region of differentially methylated loci may also have a regulatory impact on TF binding and gene expression, we identified differentially methylated regions (DMRs) – i.e., dense regions of DMCs – using an R package eDMR (with default settings, DMR mean methylation difference $\geq 20\%$)³². DMRs from various stages have a similar length distribution (median = ~185 bp) and the numbers of DMRs increased as tumor progressed (Table 2). To study chronological changes of DNA methylation, we analyzed the union of DMRs identified in the four stages (as long as a region was identified as DMR in one stage, we included that region). Unsupervised hierarchical clustering on the union of DMRs revealed three groups (Fig. 4A): group 1 (1,447) with hypermethylation, group 2 (1,050) with hypomethylation and group 3 (1,790) without significant methylation differences (methylation difference $< 20\%$) at the premalignant stages but with a moderate level of hypomethylation at the malignant stages. Then we applied GREAT (Genomic Regions Enrichment of Annotations Tool)³³ to associate DMRs with genes and test for enrichment. Because hypermethylation is often associated with *cis*-regulatory regions, we are more interested in Group 1. Angiogenesis was one of the significantly enriched terms in Group 1 (Fig. 4B).

Genes in Group 1 were enriched in disease ontology terms related to cancer, carcinoma, and neoplasm of breast (Fig. 4C). This suggests the DMRs identified here are of biological significance in breast cancer. We also confirmed the enrichment of PRC2 targets (FDR < 0.05) in this group. In Group 1, we found 30 regions connected to 24 genes in PANTHER angiogenesis pathway (*Fzd1, Angpt2, Wnt5a, Vegfa, Dok3, Grap, Fzd5, Pik3r2, Arhgap8, Wnt7b, Shc1, Notch2, Grb7, Fzd2, Rhoc, Axin2, Pxn, Grb2, Plcg2, Dll4, Pik3c2b, Ephb2, Hras1, and Prkcz*), including some *Fzd* genes. We found a hypermethylated DMR in the gene body of *Fzd5* (TSS + 2,655 bp) and overexpression of *Fzd5* in PyMT mice. *Fzd* genes encode Frizzled class receptors, which initiate Wnt signaling cascade when activated. The COSMIC (Catalogue of Somatic Mutations in Cancer)³⁴ database reported overexpression of *Fzd5* in 475 TCGA cancer samples. Previous studies also observed hypermethylation in the gene bodies of *Fzd1, Fzd2, Fzd7, and Fzd10* in pancreatic adenocarcinoma³⁵.

In addition, our DNA methylation profiles showed a clear transition from the premalignant to the malignant invasive stages (Fig. 2A,B). To further investigate DNA methylation changes in the malignant transition, we compared all premalignant PyMT samples (weeks 6 and 8) with all malignant samples (weeks 10 and 12) and mapped DMCs (methylation difference $\geq 10\%$ and q -value < 0.01) identified to gene promoters. To minimize effects due to mammary gland development, we filtered out genes that were also identified by a similar comparison using control samples. We found 496 and 943 genes with promoter hypomethylation and hypermethylation, respectively, at the malignant stages compared with early lesions. Those 943 genes were enriched in PRC2 targets (Fig. 5A). The results are consistent with DMC and DMR analysis that genes with increased methylation during cancer progression were enriched in PRC2 targets.

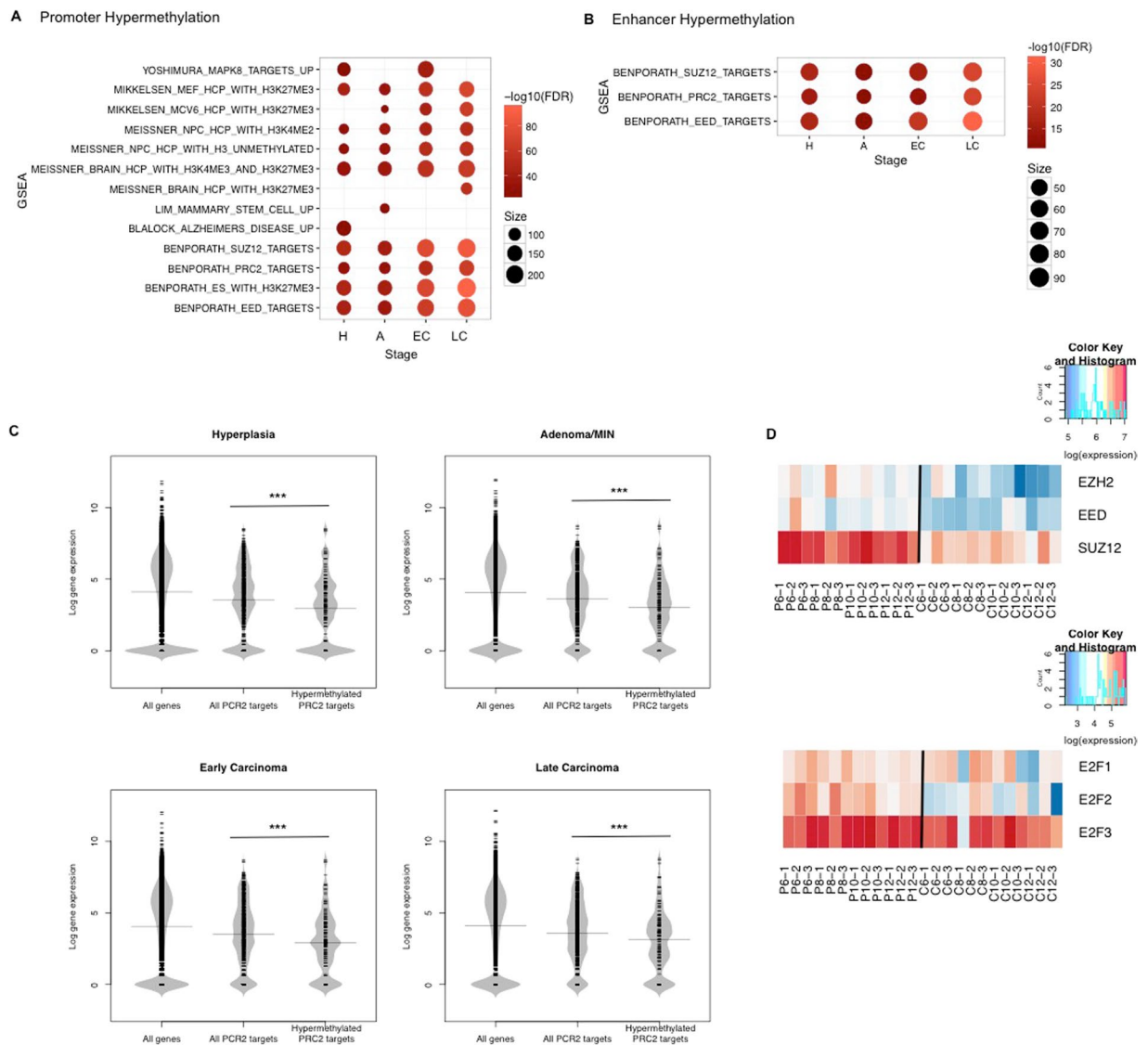


Figure 3. PRC2 targets enrichment and expression. **(A)** Gene set enrichment analysis with GSEA on genes with promoter hypermethylation. The top 10 terms with most significant FDR at each stage were plotted. H = hyperplasia, A = adenoma/MIN, EC = early carcinoma, LC = late carcinoma. Color represents $-\log_{10}(\text{FDR})$, dot size represents the number of genes overlapping each term. **(B)** Gene set enrichment analysis with GSEA on hypermethylated enhancer target genes. **(C)** Log transformed gene expression of PRC2 targets with promoter hypermethylation (at week 12) compared with all PRC2 targets and all genes. (***) Wilcoxon-test p -value < 0.05 **(D)** Log transformed gene expression of EZH2, EED and SUZ12. **(E)** Log transformed gene expression of E2Fs. P = PyMT, C = controls, 1–3 represent 3 biological replicates.

Stage	Hypomethylation	Hypermethylation	Total
Hyperplasia	1,447	471	1,918
Adenoma/MIN	1,410	594	2,004
Early carcinoma	2,054	1,004	3,058
Late carcinoma	2,221	1,156	3,377

Table 2. Total DMRs.

Co-methylation network analysis. In addition to the differential methylation analyses, we built and explored the co-methylation gene network to investigate gene functions at a higher systems level. We discovered 31 modules, including a module that was highly preserved between PyMT samples and controls as measured by the z-summary score (colored in pink in Fig. 4D) and also showed increased methylation (Wilcoxon test, adjusted p -value < 0.05) in PyMT samples. After mapping DMCs in this module to gene promoters, we discovered an

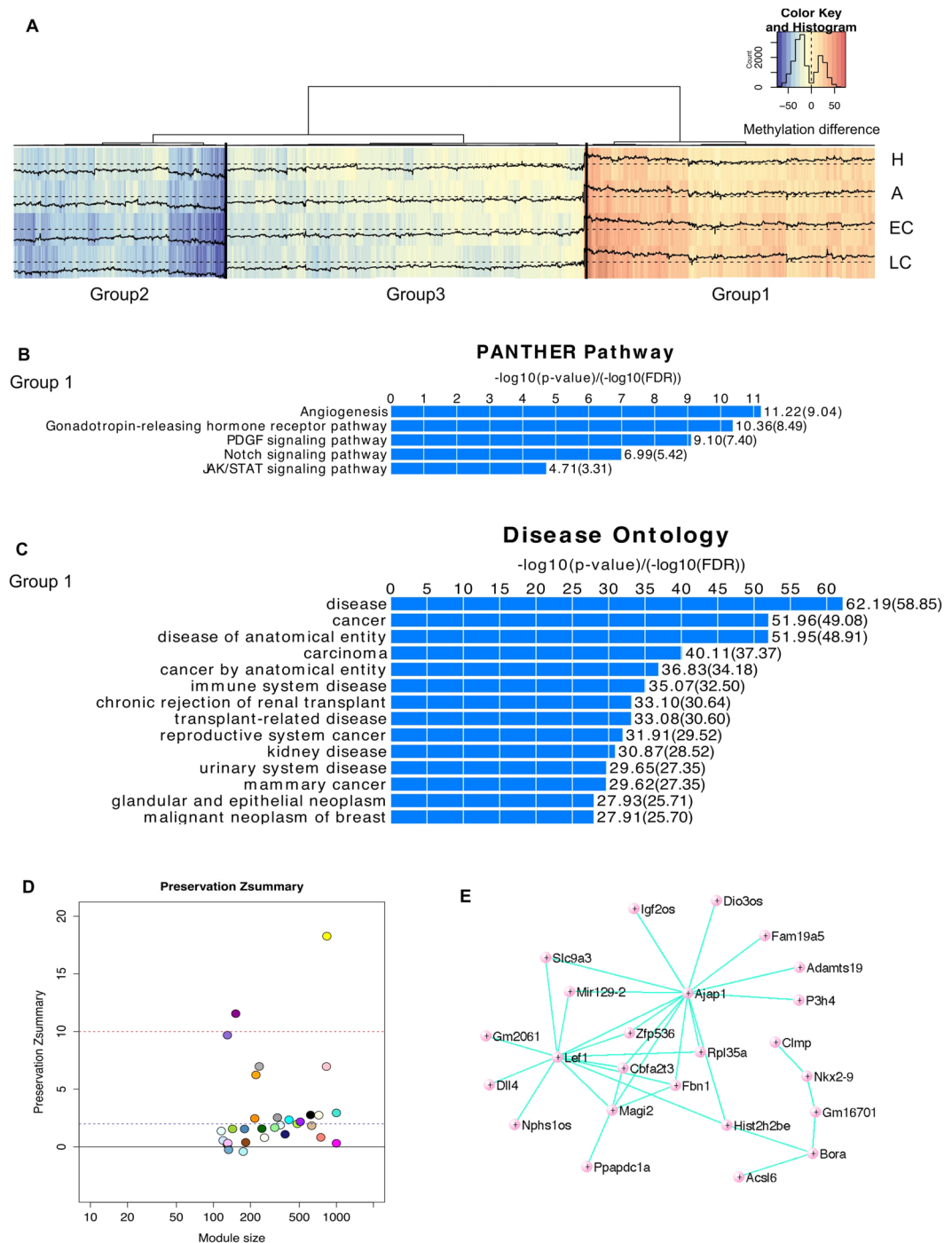


Figure 4. Clusters and enrichment of DMRs. (A) Hierarchical clustering of all DMRs detected during cancer progression. Color represents mean methylation differences between PyMT samples and controls. Dashed black lines represent 0 values, solid lines are mean values of methylation difference. (B) PANTHER pathway enrichment of Group 1. (C) Disease ontology enrichment of Group 1. H = hyperplasia, A = adenoma/MIN, EC = early carcinoma, LC = late carcinoma. (D) Preservation z-summary scores for all modules detected. (E) Some hub genes of the pink module, visualized with VisANT⁸⁸.

enrichment of SUZ12, EED and PRC2 targets (FDR < 0.05) among genes in this module. Moreover, we identified genes – e.g., *Ajap1* and its network neighbors – with strong co-methylation connections in this module (Fig. 4E). Interestingly, the putative tumor suppressor AJAP1 has been suggested to be epigenetically silenced by DNA methylation in many glioblastoma³⁶ and its reduced expression might be associated with patients' better survival outcome³⁷.

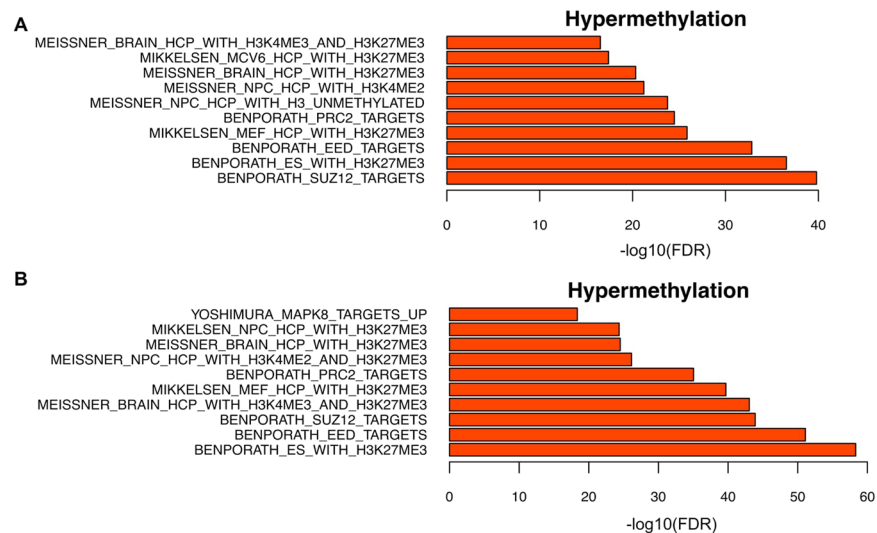


Figure 5. GSEA gene set enrichment. **(A)** Genes with promoter hypermethylation during malignant transition enriched in PRC2 targets. **(B)** Genes with promoter hypermethylation (DMR methylation difference $\geq 40\%$ and q -value < 0.05) in luminal B TCGA samples enriched in PRC2 targets.

TCGA breast invasive carcinoma samples exhibit DNA methylation dysregulation in PRC2 targets. To translate our findings from the PyMT mice to human, we studied TCGA breast carcinoma data of gene expression (RNA-sequencing) and DNA methylation (HumanMethylation450 BeadChip). After subtyping all human breast cancer cases into molecular types using the AIMS method³⁸, we end up with 269 luminal A, 189 luminal B, 111 HER2-enriched, and 142 basal-like samples. We used the bump-hunting method implemented in the R package minfi to identify DMRs (methylation difference $\geq 25\%$ and FDR < 0.01 or 0.05) by comparing tumor samples with normal mammary gland controls in each subtype. We then carried out enrichment analysis on genes with hypo- or hyper-methylated promoter in the luminal B subtype, which is most closely recapitulated by the PyMT mice^{26,27}. Genes with promoter hypermethylated DMRs were enriched in PRC2 targets (Fig. 5B). There were 313 PRC2 targets and their expression was significantly reduced when compared with all PRC2 targets (Supplementary Fig. S3A, Wilcoxon-test p -value < 0.05). The enrichment of PRC2 target genes was a common signature for other subtypes: we found 271, 206 and 43 PRC2 targets with promoter hypermethylation in luminal A, HER2-enriched and basal-like subtypes, with a significant overlap among these genes (Supplementary Fig. S3B).

To narrow down the candidate gene list, we used 12 PRC2 target genes overlapped between 60 PRC2 targets (DMR methylation difference $\geq 40\%$ and FDR < 0.01) and 130 PRC2 targets with significantly reduced expression (FDR < 0.05) as a gene panel to test if they can predict metastasis risk in human patients. We applied the gene list to Gene Expression-Based Outcome for Breast Cancer (GOBO)³⁹ and found there was a significant difference on distant metastasis-free survival (DMFS) between patients with high and low expression of those genes (Supplementary Fig. S3C). This suggests potential biological importance of DNA methylation mediated gene silencing in breast cancer. Because many PRC2 targets function in cell differentiation/differentiation, we further explored six such genes (*SLIT2*, *PDGFRA*, *GATA6*, *TAL1*, *FLII*, *KLF4*) among the 12 genes; only *KLF4* is known to be involved in breast cancer based on text mining. They also presented prognostic value in breast cancer patients (Supplementary Fig. S3D), hence can be potential novel biomarkers that worth further investigation.

Discussion

We observed, for the first time, a global dysregulation of DNA methylation during malignant breast cancer progression in the MMTV-PyMT mouse model. In addition to a global hypomethylation and possible preservation of DMCs, we observed an increase of both DMCs number and degree of methylation difference during tumor development. Such shift accompanied the malignant transition, with a slight overall shift towards hypermethylation. PRC2 target genes were the primary targets of this epigenetic field defects, these findings may contribute to explain the regulatory mechanisms underlying in breast cancer of PyMT mice and humans.

Based on the status of hormone receptors such as estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor-2 (HER2), breast cancers can be largely clustered into five subtypes: luminal A, luminal B (represented by the PyMT mouse model), HER2-enriched, basal-like, and normal-like⁴⁰. The breast cancer hormone receptor status could be connection to DNA methylation^{41,42}. Thus, in addition to these hormone receptors, DNA methylation also exhibit subtype-specific patterns, which has been shown in studies of cell lines, tissue samples, and TCGA data. For example, Park *et al.* reported distinct promoter CpG island methylation among breast cancer subtypes at 12 loci related to tumor progression, including *APC*, *DLEC1*, *GRIN2B*, *GSTP1*, *HOXA1*, *MT1G*, *RARB*, *RASSF1A*, *RUNX3*, *SCGB3A1*, *SFRP1*, and *TMEFF2*⁴³. Dedeurwaerder *et al.* profiled DNA methylation in 248 breast tissues and characterized new breast cancer subtypes associated with T lymphocyte infiltration beyond the widely-used cancer classification based on gene expression profiles⁴⁴.

In the initial TCGA BRCA study, epigenetic changes in more than 800 breast tumors were detected¹⁹. Hierarchical clustering of methylation values of all samples revealed several groups corresponding to intrinsic subtypes, with one group showing hypermethylation significantly enriched for luminal B subtype samples and with low frequency of PIK3CA, MAP3K1 and MAP2K4 mutations. Some 490 genes, enriched in Wnt signaling, exhibited both DNA hypermethylation and reduced expression.

The higher methylation in luminal B samples has been also observed previously. In one study⁴⁵, DNA methylation difference among luminal A, luminal B and basal-like subtype was readily observed in 189 breast tumor samples with basal-like ones being least frequently methylated. In particular, targets of PRC2 (Polycomb repressive complex) showed higher methylation in luminal B samples. In another¹⁷, 15 CpG loci were found differentially methylated among various subtypes and novel epigenotypes that distinguish basal-like tumors from HER2-overexpressing tumors. In a more recent study⁴⁶, the unsupervised clustering of 188 breast tumor samples revealed seven DNA methylation epigenetic subgroups as epitypes. Only hypermethylation patterns in luminal subtypes might play roles in tumor progression but not that in basal-like subtype. In addition, studies of breast cancer cell lines have also identified distinct methylation patterns on CpG island shores among subtypes as well as hypomethylation in the basal-like B subtype^{47,48}.

The study by Fang *et al.*⁴⁹ has established foundations of epigenetic changes in the metastatic process of breast cancer and showed that breast cancer can be characterized by B-CIMP (breast CpG island methylator phenotype). The aforementioned study as well showed elevated methylation levels and frequencies in some gene promoters as well as a gradual increase in the number of methylated genes from normal mammary tissue to FEA, ADH and DCIS²³. Here we report a global dysregulation of DNA methylation starting as early as the hyperplasia stage. Moreover, our previous analysis showed that genes such as *Dnmt1* implicated in DNA methylation were differentially expressed between PyMT samples and controls³⁰. Although no significant DNA methylation changes was found in DNMTs and TETs genes, which are all key players in DNA methylation process, we observed an increased expression of DNMT3B in late carcinoma stage (log₂ fold change is 1.28, FDR = 0.02) from a previous study of mRNA expression in this mouse model³⁰. Together all provided strong evidence for interplay between dysregulation of DNA methylation and abnormal gene expression in tumorigenesis of breast cancer.

We also observed a similar enrichment of PRC2 targets among genes with promoter hypermethylation in the PyMT mice as well as in human breast cancer. This is very much in consistent with the findings in normal tissue adjacent to breast cancer exhibiting DNA methylation alterations, which are also associated with PRC2 targets²⁴. Previously, cancer specific promoter DNA hypermethylation associated with Polycomb targets was observed^{50–52}. In our study, we extended the observations to early stage lesions. It has been found that DNA methylation affects Polycomb target genes prior to cervical neoplastic transformation, and such risk can be predicted⁵³. Recently, another study also found genes with promoter methylation alterations starting in early stage of cervical intraepithelial neoplasia were enriched in PRC2 targets⁵⁴.

We hypothesized that there is a regulatory cascade from E2Fs to PRC2-mediated gene silencing, which connected our findings from both expression profiles and DNA methylation profiles. Two core components of PRC2 – EED and EZH2 – are targets of E2Fs^{55,56}. Together with the E2Fs, they showed increased mRNA expression in PyMT samples. As discussed before, we also found elevated expression of *Dnmt1*, which is an E2F target as well^{55,57}. The elevated E2Fs levels may result in increased expression of DNMT1, EED and EZH2, leading to reduced expression of PRC2 targets.

SUZ12, EED, and EZH2 are three core components of PRC2. EZH2 catalyzes histone H3 trimethylation of lysine 27 (H3K27me₃), leading to transcriptional repression. EZH2 and SUZ12 are known to initiate tumorigenesis, and are overexpressed in various human cancers^{58–60}. Increased expression of *EED* and *EZH2* was detected in human breast cancer lymph node metastases⁶¹. Overexpression of *EZH2* was as well reported in prostate cancer and is suggested to be involved in tumor progression⁶². It is also considered a marker of tumor aggressiveness in breast cancer⁶³. Increased expression of PRC2 core subunits may promote malignant progression through EMT (epithelial–mesenchymal transition) by repressing E-cadherin^{64–66}. The results of EZH2 high activation seem to be context specific: the molecule can transcriptionally silence DNA-damage repair genes, pRB tumor suppressor and lineage specification genes, all leading to cancer as a common consequence⁶⁷. Many PRC2 targets are involved in differentiation, and the fact that EZH2 suppresses expression of some lineage specification genes suggests that EZH2 may promote transformation by repressing differentiation⁶⁷. In our study, we also observed reduced expression of some PRC2 targets functioning in differentiation in PyMT mice.

It is known that DNA methylation and H3K27me₃ are both involved in epigenetic gene silencing. In general, H3K27me₃ and DNA methylation are mutually exclusive at CpG islands⁶⁸, but the two repressive marks are more likely to co-occur at CpG islands and TSS of silenced genes in cancer⁶⁹. Although the relationship between PRC2-mediated gene silencing and DNA methylation remains elusive, it seems to be more than a dual repression. In one study, epigenetic repression occurs first through gain of H3K27me₃ and then through gain of DNA methylation⁷⁰. In another study with mouse embryonic stem cells (ESCs), PRC2 is required for DNA methylation at some genes, while DNA methylation represses H3K27me₃ placement globally⁷¹. Although EZH2 may recruit DNMTs⁷², it is suggested that the recruitment of DNMT3A to specific sites by EZH2 alone may not be sufficient for *de novo* DNA methylation⁷³. Multiple factors, such as transcription factors and histone modification are involved in recruiting PRC2 to targeted genomic regions⁷⁴. Studies also suggested that the long non-coding RNA (lncRNA) *HOTAIR* interacts with PRC2 and is required to target its occupancy⁷⁵. In breast cancer, *HOTAIR* showed increased expression in primary tumors as well as metastases⁷⁶. In addition to lncRNA, another type of non-coding RNAs: microRNA, also contribute to breast cancer progression in this PyMT mice as suggested by our previous study⁷⁷. Taken together, there seems to be a synergy of epigenetic regulation and gene expression in the PyMT mouse model of breast cancer.

Although our current findings need further validation and more studies to understand the functional links between E2Fs-(PRC2/DNMTs)-PRC2 targets repression, it is important to point out that EZH2 and DNMT1

are targeted by multiple drugs (EZH2: E11, EPZ-6438, GSK126; DNMT1: PROCAINAMIDE, IFOSFAMIDE, DECITABINE, CISPLATINUM) according to DGIdb (Drug-Gene interaction database)⁷⁸. The elevated expression of both genes may warrant a combinatory use for breast cancer early treatment and metastasis prevention. The findings from this study help to improve our understanding of the molecular mechanisms of breast cancer progression in the MMTV-PyMT mouse model as well as in humans. Combining it with our previous mRNA study, we found that gene expression and DNA methylation changes precede tumor malignant transition in this mouse model. New breast cancer candidate genes discovered in both studies warrant further investigation and may serve as cancer biomarkers for prognosis.

Methods

Animals and tissue collection. This study was approved by the Institutional Animal Care and Use Committee (IACUC) of Albert Einstein College of Medicine. All procedures involving mice were conducted in accordance with the National Institutes of Health guidelines concerning the use and care of experimental animals. Male PyMT mice (FVB/N-Tg (MMTV-PyVT) 634Mul/J mice, Stock Number: 002374, the Jackson Laboratory) were randomly bred with homozygous FVB females to obtain F1 female mice (PyMT mice hereafter). They were heterozygous for the PyMT transgene, they developed breast cancer spontaneously and were used as cases. Age matched homozygous FVB females were used as controls.

We selected four time points, corresponding to the main stages of tumor progression: most PyMT mice develop hyperplasia at week 6, adenoma/MIN at week 8, early carcinoma at week 10, and late carcinoma with lung metastasis at week 12²⁵. At each time point, three PyMT mice and three age-matched controls were sacrificed, bulk mammary tumors and normal mammary glands were collected, snap froze and stored at -80°C until use. The histological changes undergoing in the PyMT samples dissected for DNA methylation analysis was evaluated by a pathologist after H&E staining. Typical features of the tumor developmental stage based on the mouse age were observed (see our previous paper³⁰ for details. Samples used for RNA-sequencing and ERRBS were not necessarily from the same tumor piece.)

Enhanced Reduced Representation Bisulphite Sequencing (ERRBS). Total genomic DNA was extracted from the snap frozen samples using a phenol-chloroform method⁷⁹. DNA samples passing Nanodrop quality control were used for ERRBS following the standard protocol^{3,80}.

Data preprocessing and detection of differential methylation. Raw reads passing Illumina's purity filter were preprocessed by an ERRBS pipeline⁸⁰, which uses Bismark⁸¹ to align reads to the mm10 reference genome. All statistical analysis were carried out by R v3.2.1⁸². After assessing samples quality with reads coverage, methylation percentage, and principal component analysis, we removed two outliers from the dataset.

Using the methylKit R package⁸³ we analyzed DMCs. We first retained in the analysis reads with ≥ 10 coverage and captured, on average, ~ 1.4 million confident loci. For each locus, we then calculated a methylation score as percentage of methylation (number of methylated cytosines/(number of methylated cytosines + unmethylated cytosines)) and got the methylation difference between PyMT samples and age-matched FVB control samples. We mapped DMCs (methylation difference $\geq 10\%$ and q -value < 0.01) to genes, using annotation from the UCSC genome browser (<https://genome.ucsc.edu>). Promoters are defined as RefSeq transcription start site (TSS) ± 2 kb. Biological pathways and gene sets enrichment were performed using GSEA³¹. Predicted mouse enhancer regions were obtained from a study using the mouse ENCODE data⁸⁴; enhancer prediction was based on ChIP-seq data of multiple tissues and cell types⁸⁴. Data originally mapped to mm9 was lifted over to mm10 using the R package rtracklayer. Permutation tests were used to demonstrate the significance of DMCs enrichment. For permutation tests, we randomly sampled n loci of interest (for example, number of DMCs), and count how many mapped to annotation X (for example, enhancers); the process was repeated 1,000 times to generate an empirical distribution of overlap counts and determine a p -value.

Co-methylation network analysis. We adopted methods from the WGCNA (weighted gene co-expression network analysis)⁸⁵ to study gene co-methylation network. We first retained significant DMCs (q -value < 0.01) from previous step. Next, for computational efficiency, we further restricted to DMCs in promoter regions 2 kb upstream to TSS. We focused on upstream promoter because the GETx eQTL study⁸⁶ showed an upstream bias (more eQTLs located upstream of TSS), suggesting upstream TSS regions possibly have greater influence on expression. After these two steps, over 21k CpG loci remained and were used to construct a signed co-methylaiton network using an adjacency matrix based on Pearson's correlation coefficient on every pair of CpG loci. Modules were identified using a "dynamic tree cut" method⁸⁷. Preservation of module connectivity patterns between cases and controls were gauged by a z -summary statistic. Modules with high preservation showed significant difference on module eigengene (first principal component of a module) between PyMT samples and controls (Wilcoxon-test adjusted p -value < 0.05), meaning methylation values varied in the two groups.

Bisulphite MassArray verification assays. We performed Bisulfite MassArray to verify our ERRBS results using samples from carcinoma stage. We bisulfite-converted DNA using the Zymo EZ-Methylation-Gold Kit. Primers were designed to span loci with high, intermediate and low levels of methylation (Supplementary Table S3). We performed MassArray (EpiTyper) and proved a high correlation between MassArray and ERRBS results (Supplementary Fig. S4).

Availability of Data. The PyMT ERRBS data is available in the Gene Expression Omnibus (GEO) database as GSE83623.

Declarations and Ethics Approval. This study of cancer in mice was approved by the Institutional Animal Care and Use Committee (IACUC) of Albert Einstein College of Medicine. All procedures involving mice were conducted in accordance with the National Institutes of Health guidelines concerning the use and care of experimental animals.

References

- Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484–92 (2012).
- Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* **45**, 1127–1133 (2013).
- Akalin, A. *et al.* Base-Pair Resolution DNA Methylation Sequencing Reveals Profoundly Divergent Epigenetic Landscapes in Acute Myeloid Leukemia. *Plos Genetics* **8** (2012).
- Ehrlich, M. DNA methylation in cancer: too much, but also too little. *Oncogene* **21**, 5400–13 (2002).
- Baylin, S. B. DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol* **2**(Suppl 1), S4–11 (2005).
- Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726–34 (2011).
- Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683–92 (2007).
- Hon, G. C. *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res* **22**, 246–58 (2012).
- Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1**, 239–59 (2009).
- Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89–92 (1983).
- Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet* **7**, 21–33 (2006).
- Eden, A., Gaudet, F., Waghmare, A. & Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300**, 455 (2003).
- Chen, R. Z., Pettersson, U., Beard, C., Jackson-Grusby, L. & Jaenisch, R. DNA hypomethylation leads to elevated mutation rates. *Nature* **395**, 89–93 (1998).
- Jones, A. *et al.* Role of DNA methylation and epigenetic silencing of HAND2 in endometrial cancer development. *PLoS Med* **10**, e1001551 (2013).
- Gaudet, F. *et al.* Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489–92 (2003).
- Adorjan, P. *et al.* Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res* **30**, e21 (2002).
- Bediaga, N. G. *et al.* DNA methylation epigenotypes in breast cancer molecular subtypes. *Breast Cancer Res* **12**, R77 (2010).
- Fleischer, T. *et al.* Genome-wide DNA methylation profiles in progression to *in situ* and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol* **15**, 435 (2014).
- Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
- Gao, Y. *et al.* The integrative epigenomic-transcriptomic landscape of ER positive breast cancer. *Clin Epigenetics* **7**, 126 (2015).
- Ma, X. J. *et al.* Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci USA* **100**, 5974–9 (2003).
- Allred, D. C., Mohsin, S. K. & Fuqua, S. A. Histological and biological evolution of human premalignant breast disease. *Endocr Relat Cancer* **8**, 47–61 (2001).
- Park, S. Y. *et al.* Promoter CpG island hypermethylation during breast cancer progression. *Virchows Arch* **458**, 73–84 (2011).
- Teschendorff, A. E. *et al.* DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun* **7**, 10478 (2016).
- Lin, E. Y. *et al.* Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *American Journal of Pathology* **163**, 2113–2126 (2003).
- Pfefferle, A. D. *et al.* Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biology* **14** (2013).
- Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biology* **8** (2007).
- Acosta, D. *et al.* DNA methylation changes in murine breast adenocarcinomas allow the identification of candidate genes for human breast carcinogenesis. *Mamm Genome* **22**, 249–59 (2011).
- Ried, T. *et al.* Molecular cytogenetics of mouse models of breast cancer. *Breast Dis* **19**, 59–67 (2004).
- Cai, Y. *et al.* Transcriptomic dynamics of breast cancer progression in the MMTV-PyMT mouse model. *BMC Genomics* **18**, 185 (2017).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–50 (2005).
- Li, S. *et al.* An optimized algorithm for detecting and annotating regional differential methylation. *BMC Bioinformatics* **14**(Suppl 5), S10 (2013).
- McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495–501 (2010).
- Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**, D805–11 (2015).
- Vincent, A. *et al.* Genome-wide analysis of promoter methylation associated with gene expression profile in pancreatic adenocarcinoma. *Clin Cancer Res* **17**, 4341–54 (2011).
- Lin, N. *et al.* Deletion or epigenetic silencing of AJAP1 on 1p36 in glioblastoma. *Mol Cancer Res* **10**, 208–17 (2012).
- Cogdell, D. *et al.* Tumor-associated methylation of the putative tumor suppressor AJAP1 gene and association between decreased AJAP1 expression and shorter survival in patients with glioma. *Chin J Cancer* **30**, 247–53 (2011).
- Paquet, E. R. & Hallett, M. T. Absolute assignment of breast cancer intrinsic molecular subtype. *J Natl Cancer Inst* **107**, 357 (2015).
- Ringner, M., Fredlund, E., Hakkinen, J., Borg, A. & Staaf, J. GOBO: gene expression-based outcome for breast cancer online. *PLoS One* **6**, e17911 (2011).
- Carey, L. A. *et al.* Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *Jama* **295**, 2492–502 (2006).
- Ottaviano, Y. L. *et al.* Methylation of the estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast cancer cells. *Cancer Res* **54**, 2552–5 (1994).
- Fackler, M. J. *et al.* Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer Res* **71**, 6195–207 (2011).
- Park, S. Y. *et al.* Distinct patterns of promoter CpG island methylation of breast cancer subtypes are associated with stem cell phenotypes. *Mod Pathol* **25**, 185–96 (2012).
- Dedeurwaerder, S. *et al.* DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol Med* **3**, 726–41 (2011).
- Holm, K. *et al.* Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res* **12**, R36 (2010).
- Holm, K. *et al.* An integrated genomics analysis of epigenetic subtypes in human breast tumors links DNA methylation patterns to chromatin states in normal mammary cells. *Breast Cancer Res* **18**, 27 (2016).

47. Rhee, J. K. *et al.* Integrated analysis of genome-wide DNA methylation and gene expression profiles in molecular subtypes of breast cancer. *Nucleic Acids Res* **41**, 8464–74 (2013).
48. Chae, H., Lee, S., Nephew, K. P. & Kim, S. Subtype-specific CpG island shore methylation and mutation patterns in 30 breast cancer cell lines. *BMC Syst Biol* **10**, 116 (2016).
49. Fang, F. *et al.* Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci Transl Med* **3**, 75ra25 (2011).
50. Gal-Yam, E. N. *et al.* Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc Natl Acad Sci USA* **105**, 12979–84 (2008).
51. Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nat Genet* **39**, 232–6 (2007).
52. Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat Genet* **39**, 157–8 (2007).
53. Teschendorff, A. E. *et al.* Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. *Genome Med* **4**, 24 (2012).
54. Wijetunga, N. A. *et al.* A polycomb-mediated epigenetic field defect precedes invasive cervical carcinoma. *Oncotarget* **7**, 62133–62143 (2016).
55. Muller, H. *et al.* E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev* **15**, 267–85 (2001).
56. Bracken, A. P. *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *Embo j* **22**, 5323–35 (2003).
57. McCabe, M. T., Davis, J. N. & Day, M. L. Regulation of DNA methyltransferase 1 by the pRb/E2F1 pathway. *Cancer Res* **65**, 3624–32 (2005).
58. Jene-Sanz, A. *et al.* Expression of polycomb targets predicts breast cancer prognosis. *Mol Cell Biol* **33**, 3951–61 (2013).
59. Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* **6**, 846–56 (2006).
60. Sauvageau, M. & Sauvageau, G. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell* **7**, 299–313 (2010).
61. Yu, H. *et al.* PRC2/EED-EZH2 complex is up-regulated in breast cancer lymph node metastasis compared to primary tumor and correlates with tumor proliferation *in situ*. *PLoS One* **7**, e51239 (2012).
62. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–9 (2002).
63. Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci USA* **100**, 11606–11 (2003).
64. Cao, Q. *et al.* Repression of E-cadherin by the polycomb group protein EZH2 in cancer. *Oncogene* **27**, 7274–84 (2008).
65. Herranz, N. *et al.* Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. *Mol Cell Biol* **28**, 4772–81 (2008).
66. Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat Med* **19**, 1438–49 (2013).
67. Kim, K. H. & Roberts, C. W. Targeting EZH2 in cancer. *Nat Med* **22**, 128–34 (2016).
68. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res* **22**, 1128–38 (2012).
69. Statham, A. L. *et al.* Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res* **22**, 1120–7 (2012).
70. Tellez, C. S. *et al.* EMT and stem cell-like properties associated with miR-205 and miR-200 epigenetic silencing are early manifestations during carcinogen-induced transformation of human lung epithelial cells. *Cancer Res* **71**, 3087–97 (2011).
71. Hagarman, J. A., Motley, M. P., Kristjansdottir, K. & Soloway, P. D. Coordinate regulation of DNA methylation and H3K27me3 in mouse embryonic stem cells. *PLoS One* **8**, e53880 (2013).
72. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (2005).
73. Rush, M. *et al.* Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not *de novo* DNA methylation. *Epigenetics* **4**, 404–14 (2009).
74. van Kruijsbergen, I., Hontelez, S. & Veenstra, G. J. Recruiting polycomb to chromatin. *Int J Biochem Cell Biol* **67**, 177–87 (2015).
75. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–23 (2007).
76. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–6 (2010).
77. Nogales-Cadenas, R. *et al.* MicroRNA expression and gene regulation drive breast cancer progression and metastasis in PyMT mice. *Breast Cancer Res* **18**, 75 (2016).
78. Griffith, M. *et al.* DGIdb: mining the druggable genome. *Nat Methods* **10**, 1209–10 (2013).
79. Garcia, A. M. *et al.* Detection and analysis of somatic mutations at a lacZ reporter locus in higher organisms: application to *Mus musculus* and *Drosophila melanogaster*. *Methods Mol Biol* **371**, 267–87 (2007).
80. Garrett-Bakelman, F.E. *et al.* Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. *J Vis Exp* e52246 (2015).
81. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–2 (2011).
82. Team, R.C. R: A Language and Environment for Statistical Computing (2014).
83. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* **13**, R87 (2012).
84. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–64 (2014).
85. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics* **9** (2008).
86. Consortium, G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–60 (2015).
87. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
88. Hu, Z., Snitkin, E. S. & DeLisi, C. VisANT: an integrative framework for networks in systems biology. *Brief Bioinform* **9**, 317–25 (2008).

Acknowledgements

This work was supported by NIH grants R00 LM009770 from National Library of Medicine and R01 HG008153 from the National Human Genome Research Institute to Z.D.Z. We thank our colleagues at Albert Einstein College of Medicine – Drs John Grealley, Jan Vijg, Yousin Suh, and Deyou Zheng – for their expert advice. We also thank the Einstein Center for Epigenomics, the Epigenomics Shared Facility, the Histotechnology and Comparative Pathology Facility (supported by the Albert Einstein Cancer Center Support Grant from the National Institutes of Health under award number P30 CA013330), the Genomics Shared Facility and the Computational Epigenomics Group for technical support.

Author Contributions

Z.D.Z. conceived and designed this study. Y.C. carried out the analyses and performed some experiments. Y.C. and Z.D.Z. wrote the manuscript. K.B. performed animal experiments and sample preparation. J.R.L., Q.Z., and C.M. provided useful input for the analyses and helped edit the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-24005-x>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018