

RESEARCH ARTICLE

Genetic and Epigenetic Regulation of TOX3 Expression in Breast Cancer

Yoo-Jeong Han^{1*}, Jing Zhang¹, Yonglan Zheng¹, Dezheng Huo², Olufunmilayo I. Olopade^{1*}

1 Center for Clinical Cancer Genetics and Global Health; and Section of Hematology and Oncology, Department of Medicine, University of Chicago, Chicago, IL 60637, United States of America, **2** Department of Public Health Sciences, University of Chicago, Chicago, IL 60637, United States of America

* folopade@medicine.bsd.uchicago.edu (OIO); yjhan@medicine.bsd.uchicago.edu (YJH)



OPEN ACCESS

Citation: Han Y-J, Zhang J, Zheng Y, Huo D, Olopade OI (2016) Genetic and Epigenetic Regulation of TOX3 Expression in Breast Cancer. PLoS ONE 11(11): e0165559. doi:10.1371/journal.pone.0165559

Editor: Hiromu Suzuki, Sapporo Ika Daigaku, JAPAN

Received: July 11, 2016

Accepted: October 13, 2016

Published: November 2, 2016

Copyright: © 2016 Han et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by NIH R01 CA142996-01 (OIO), NIH Specialized Program of Research Excellence in Breast Cancer P50 CA125183 (OIO), Breast Cancer Research Foundation (YH and OIO), and the Ralph and Marion Falk Medical Research Trust (OIO). OIO is American Cancer Society Clinical Research Professor.

Abstract

Genome wide association studies (GWAS) have identified low penetrance and high frequency single nucleotide polymorphisms (SNPs) that contribute to genetic susceptibility of breast cancer. The SNPs at 16q12, close to the *TOX3* and *CASC16* genes, represent one of the susceptibility loci identified by GWAS, showing strong evidence for breast cancer association across various populations. To examine molecular mechanisms of *TOX3* regulation in breast cancer, we investigated both genetic and epigenetic factors using cell lines and datasets derived from primary breast tumors available through The Cancer Genome Atlas (TCGA). *TOX3* expression is highly up-regulated in luminal subtype tumors compared to normal breast tissues or basal-like tumors. Expression quantitative trait loci (eQTL) analyses revealed significant associations of rs3803662 and rs4784227 genotypes with *TOX3* expression in breast tumors. Bisulfite sequencing of four CpG islands in the *TOX3* promoter showed a clear difference between luminal and basal-like cancer cell lines. 5-Aza-2'-deoxycytidine treatment of a basal-like cancer cell line increased expression of *TOX3*. TCGA dataset verified significantly lower levels of methylation of the promoter in luminal breast tumors with an inverse correlation between methylation and expression of *TOX3*. Methylation QTL (mQTL) analyses showed a weak or no correlation of rs3803662 or rs4784227 with *TOX3* promoter methylation in breast tumors, indicating an independent relationship between the genetic and epigenetic events. These data suggest a complex system of *TOX3* regulation in breast tumors, driven by germline variants and somatic epigenetic modifications in a subtype specific manner.

Introduction

Genetic factors play important roles in the etiology of breast cancer. Multiple breast cancer susceptibility genes such as *BRCA1* and *BRCA2* with high penetrant disease-associated mutations have been shown to segregate in families with breast cancer by linkage studies [1]. In

Competing Interests: The authors have declared that no competing interests exist.

addition, Genome Wide Association Studies (GWAS) have identified low penetrance, high frequency single nucleotide polymorphisms (SNPs), revealing more than 90 chromosomal regions that contribute to genetic susceptibility [2,3,4,5,6]. The SNP rs3803662 at 16q12 is one of the susceptibility loci identified by GWAS, for which the minor allele conferred increased risk of breast cancer in women of European ancestry [7]. This finding was also observed in women of Asian descent [8,9] with SNP rs4784227 at 16q12 as an additional risk variant for breast cancer in Asian populations [10]. Evaluation and fine mapping of the susceptibility loci in African Americans identified significant associations with markers at 16q12 SNPs that are independent of the index signal (rs3803662) but represent novel risk variants (rs3104793, rs3104788, rs3104778) with a perfect proxy for rs3112572 [11,12,13,14]. Thus data from diverse populations provide strong evidence for breast cancer association within SNPs at 16q12.

Revealing the molecular bases for the observed associations of SNPs and cancer risks remains challenging because the risk variants map predominantly to non-coding regions and the linked markers merely serve as indicators that a causal variant is present nearby. The 16q12 SNPs reside at introns of a non-protein coding hypothetical gene (*LOC643714*), which was recently named cancer-susceptibility candidate 16 (*CASC16*). *TOX3* is located ~5 kb downstream of rs3803622 and ~50 kb downstream of rs3104788. Because enhancers are long-range cis-regulatory elements and function over up to mega base-long genomic distances to regulate the expression patterns of their target gene(s), it has been suggested that risk variants control gene expression by altering the activity of enhancers [15]. For example, *in vitro* experimentation demonstrated that the SNP rs4784227, located 18.4 kb upstream of the *TOX3* gene, altered *TOX3* gene expression by disrupting enhancer function through FOXA1 affinity modulation [16].

TOX3 is a nuclear protein containing a nuclear localization signal and a high mobility group-box domain that can modify chromatin structure. *TOX3* binds to the BRCA1 promoter and negatively regulates BRCA1 expression [17]. Ectopic expression of *TOX3* increased breast cancer cell proliferation, migration, and survival after exposure to apoptotic stimuli and were associated with tumor progression in a mouse model of breast cancer. *TOX3* is often amplified and overexpressed in breast tumors, particularly in advanced breast tumors [17]. *TOX3* mutations were also found in breast tumors with an overall mutation frequency of 4.5% [18]. Patients with ER positive tumors and high levels of *TOX3* mRNA had shorter overall- and distant metastasis free-survival, an effect mostly attributable to patients with luminal B tumors [19].

CpG islands are short sequences of genomic DNA with the length of 0.5 to several kb [20] in which the frequency of the linear 5'-CpG-3' sequence is higher than at other regions of the gene [21,22]. Aberrant methylation of CpG islands resulting in gene silencing or overexpression of genes has been associated with development and progression of cancers. In this study, we tested if overexpression of *TOX3* is due to epigenetic factors that regulate *TOX3/CASC16* expression in breast cancer cells. We identified hypomethylation of the *TOX3* promoter as a significant contributor of *TOX3* upregulation in luminal subtype breast cancer. We also found a significant association of *TOX3* expression with rs3803662 and rs4784227, a risk variant identified in European and Asian women, respectively. These results suggest that both epigenetic and genetic factors contribute to the increased expression of *TOX3* in luminal cancer. Thus, our data support a plausible molecular mechanism integrating epigenetic modifications of the *TOX3* promoter and allele specific expression of SNPs in aggressive behavior of luminal breast tumors with high *TOX3* expression.

Results

Subtype-specific expression of *TOX3/CASC16* in breast cancer cells and tumors

The chromosomal region 16q12.1–12.2 contains two genes, a protein-coding gene (*TOX3*) and a long non-coding RNA (*CASC16/LOC643714*), which generates a long intergenic non-protein coding RNA 918 (Linc00918, RefSeq NR033920.1). We surveyed *TOX3* expression in 19 breast cell lines with different molecular subtypes, including basal-like (n = 10) and luminal subtype cells (n = 7) as well as non-malignant mammary cells (n = 2), human mammary epithelial primary cells (HMEC) and immortalized mammary cells (184A1). Comparison of *TOX3* expression in these cell lines revealed greater expression of *TOX3* in luminal subtype cells compared to normal breast and basal-like subtype cell lines (Fig 1A and 1B). The *TOX3* expression levels vary within luminal subtypes, ranging from a minimum of 172.7±13.4-fold increase (MCF-7 cells) to a maximum of 39,526±1,369-fold increase (ZR-75-30 cells) compared to HMEC. LincRNA 918 from *CASC16* exhibits a similar expression pattern with higher expression in luminal subtype cells compared to basal-like and normal breast cell lines (Fig 1C and 1D).

To validate *TOX3* expression in a large set of primary breast tumors, we analyzed three datasets, TCGA, the University of North Carolina (UNC), and University of Chicago (U of C) (Fig 2A–2C). The TCGA samples consist of normal breast tissues (n = 95), luminal A (n = 417), luminal B (n = 187), Her 2 amplified (n = 67) and basal-like (n = 141) breast tumors. Across all platforms, the data were consistent with greater expression of *TOX3* in luminal and Her2-positive subtypes, compared to normal breast tissues and basal-like tumors. Together, our data demonstrated higher expression of *TOX3* in luminal subtype cells/tumors compared to basal-like subtype or normal breast cells/tissues.

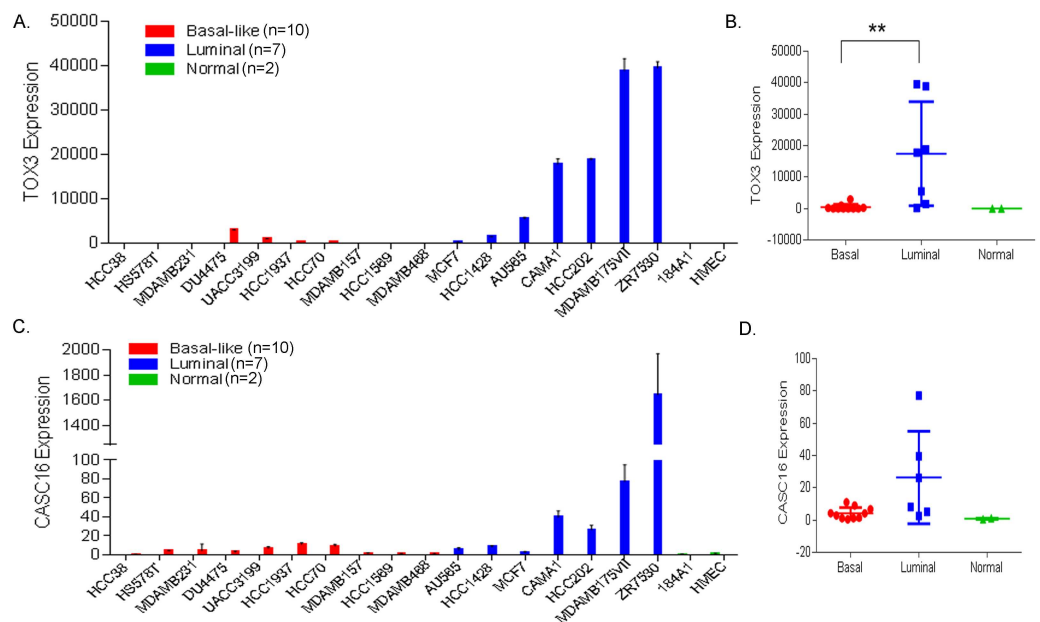


Fig 1. Increased expression of *TOX3* and *CASC16* in luminal breast cancer cell lines. (A and C) *TOX3* and *CASC16* expression was assessed in breast cancer cell lines using qRT-PCR relative to expression in normal breast epithelial cell line HMEC (RQ, relative quantity). (B and D) Expression of *TOX3* in luminal cell lines (blue) is significantly higher than expression in basal-like cells (red) and non-malignant breast epithelial cells (green) (***p*<0.01). *CASC16* also exhibits a similar expression pattern with higher expression in luminal cells but fails to reach statistical significance.

doi:10.1371/journal.pone.0165559.g001

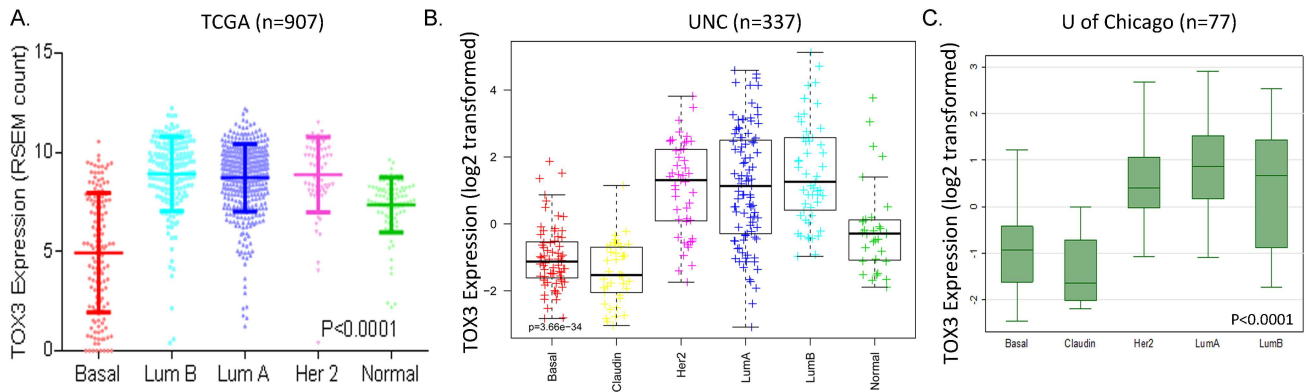


Fig 2. Subtype-specific expression of TOX3 in breast tumors. RNA sequencing (TCGA) and microarray results (UNC and U of Chicago) of TOX3 expression in breast tumors showed a significant difference in TOX3 expression among breast cancer subtypes, with TOX3 expression higher in luminal A, luminal B and Her 2 amplified tumors compared to normal breast epithelial tissues and basal-like tumors. (A) TCGA dataset ($p < 0.0001$), RNA-seq data presented by RSEM (RNA-Seq by Expectation Maximization) normalized counts; (B) the University of North Carolina (UNC) dataset ($p = 3.66e-34$), microarray data presented by log2 transformed values; (C) the University of Chicago (U of Chicago) dataset ($p < 0.0001$), microarray data presented by log2 transformed values.

doi:10.1371/journal.pone.0165559.g002

eQTL analyses showed correlations between SNP variants and TOX3 expression

To determine the effects of genetic variants associated with breast cancer on regulating TOX3 expression, we applied a linear regression model and conducted eQTL analyses on 345 SNPs genotyped 500 kb upstream and 250 kb downstream of the TOX3 gene, using TCGA dataset. We identified 77 SNPs associated with TOX3 gene expression which passed the multiple testing adjustments using a false discovery rate below 0.05 (S1 Table). A significant association with TOX3 expression was observed in rs3803662 ($p = 3.4e-7$), a risk variant of breast cancer in European women [7], with the minor A allele associated with lower TOX3 expression (Fig 3A). SNP rs4784227, a risk variant identified in Asian women [8,10], also showed a significant association with TOX3 expression ($p = 0.004$) with the minor T allele associated with increased risk of breast cancer and lower TOX3 expression (Fig 3B).

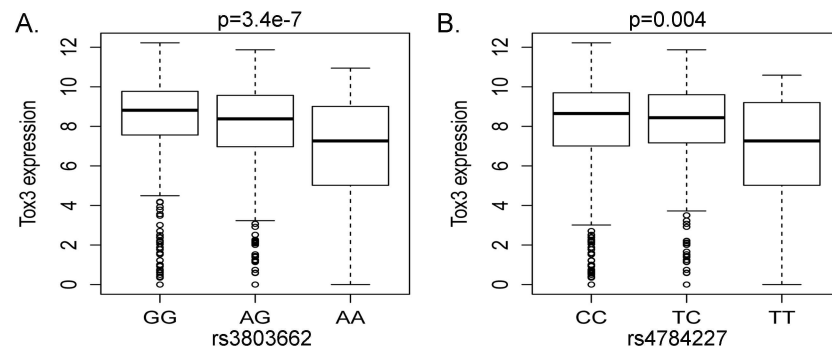


Fig 3. eQTL analyses between SNP genotypes and TOX3 expression. (A and B) eQTL analyses were performed on 345 SNPs located 500 kb upstream and 250 kb downstream of the TOX3 gene, using TCGA dataset. A significant association with TOX3 expression was observed in rs3803662 and rs4784227, a risk variant of breast cancer in European [7] and Asian women [8, 10], respectively.

doi:10.1371/journal.pone.0165559.g003

Our previous fine-mapping study in women of African ancestry revealed rs3104778 and rs3104788 as most-significant genotyped and imputed SNPs [11,12,13,14]. However, when we performed eQTL analyses of rs3112617 and rs3104788, we found no association between the variants and TOX3 expression (S1 Table). Together, our analyses confirmed rs3803662 and rs4784227 as eQTL SNPs significantly associated with TOX3 expression, whereas rs3112617 and rs3104788 have no influence on differential expression of TOX3 in luminal tumors.

Luminal breast tumors exhibits lower methylation levels in the TOX3 promoter

To determine if epigenetic regulation of the TOX3 promoter contributes to the subtype-specific expression of TOX3, we performed bisulfite sequencing in breast cancer cell lines. UCSC genome browser view shows location of CpG islands in the promoter region at chromosome 16q12 (Fig 4A). Four CpG dinucleotides in the promoter region (cg02709321, cg11410436, cg26648818, and cg01404163) were sequenced after bisulfite conversion. A representative chromatograph of a normal epithelial (HMEC), a luminal (ZR7530), and a basal-like cell line (HCC-1937) showed clear difference in sequence reading with bisulfite treatment (Fig 4B). The four sites are all unmethylated in ZR7530 cells and converted into TG (AC in the reverse strand) with bisulfite treatment, whereas they are all methylated in HMEC and HCC-1937 cells (GC in the reverse strand).

To validate CpG site methylation of TOX3 in a large set of breast tumors, we analyzed methylation levels of the four CpG dinucleotides using the TCGA HumanMethylation450 Array data. These results are expressed as beta values, which are continuous variables between 0 and 1. Hypomethylated regions have lower beta values while hypermethylated regions have higher

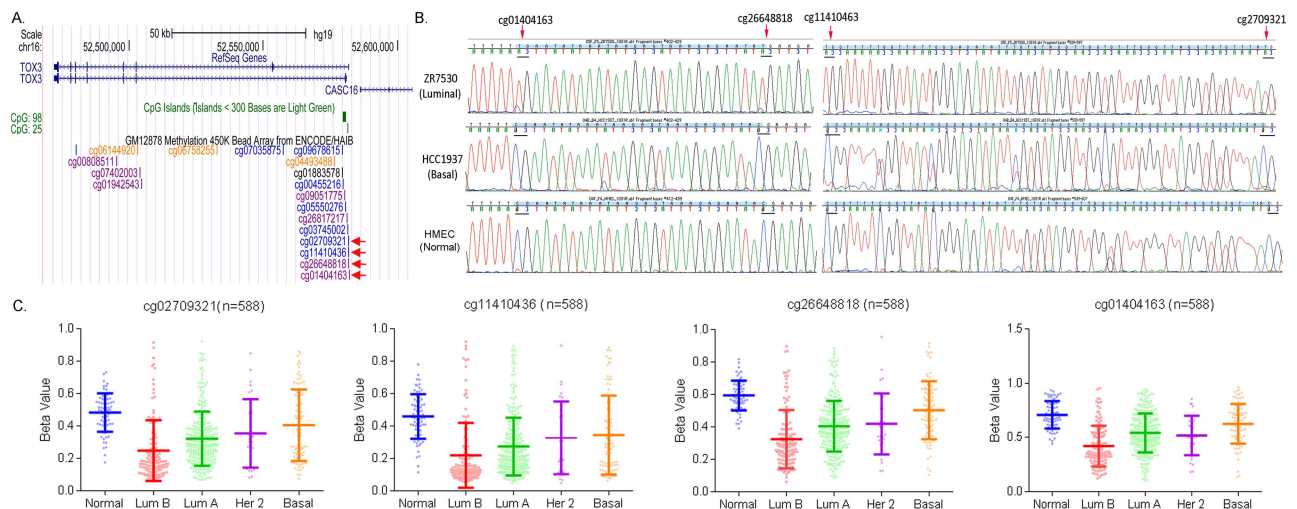


Fig 4. Subtype-specific methylation of the TOX3 promoter in breast tumors. (A) UCSC genome browser shows location of CpG islands in the TOX3 promoter at chromosome 16q12. Four CpG dinucleotides analyzed are indicated with red arrows. (B) Representative chromatographs in the TOX3 promoter after bisulfite treatment revealed unmethylated cytosines in the TOX3 promoter of ZR-75-30 (luminal subtype cells). DNA sequencing was performed on the reverse strand and four unmethylated CpG sites are underlined (AC, reverse complement of TG). In contrast, methylated cytosines (GC, reverse complement of CG) were detected in the promoter of HCC-1937 (basal-like subtype cells) and HMEC (primary epithelial normal cells) because methylation prohibited the effect of bisulfite treatment. (C) Methylation levels of four CpG dinucleotides in the promoter region of TOX3 were analyzed using the TCGA HumanMethylation450 Array data. One-way ANOVA and Tukey's multiple comparison tests showed a significant difference in methylation levels of the CpG sites between normal breast tissues and luminal A or B breast tumors ($p < 0.0001$) and between basal-like and luminal A or B breast tumors ($p < 0.01$).

doi:10.1371/journal.pone.0165559.g004

beta values. Out of 838 patient samples with DNA methylation profile, we selected 588 samples for the analysis, after excluding samples from male patients, samples with no PAM50 subtype information, or samples showing discrepancy between PAM50 subtypes and pathological analyses (sample types). The included samples were comprised of normal breast tissues (n = 70), luminal B (n = 127), luminal A (n = 276), Her 2 (n = 30) and basal-like (n = 85) breast tumors. One-way ANOVA and Tukey's multiple comparison tests at the four CpG islands showed significantly lower methylation in the CpG islands of the *TOX3* promoter in luminal A and B breast tumors compared to normal breast tissues and basal-like breast tumors (Fig 4C). Correlation analysis in the TCGA dataset showed a weak but statistically significant inverse correlation between methylation and expression levels of *TOX3* at the four CpG sites (Fig 5A). These results indicate that promoter methylation of *TOX3* may be a potential epigenetic modification resulting in the expression differences between the luminal and basal-like subtypes of breast cancer.

Inhibition of methylation increased *TOX3* expression in a basal-like cancer cell line

To determine if de-methylation of the promoter could increase expression of *TOX3*, we treated a basal-like breast cancer cell line (UACC-3199) with a DNA methylation inhibitor, 5-aza-2'-deoxycytidine (5-aza-CdR). We used *BRCA1* as an epigenetically regulated control gene, as we previously observed up-regulation of *BRCA1* driven by de-methylation of the promoter in UACC-3199 cells [23]. As expected, *BRCA1* expression was highly increased (a 5.52 ± 0.36 fold increase) in 5-aza-CdR treated cells compared to vehicle alone (DMSO) (Fig 5B). Our data also showed a 2.93 ± 0.27 fold increase in *TOX3* expression in 5-aza-CdR treated cells compared to control, indicating that CpG methylation of the *TOX3* promoter plays a significant role in the regulation of *TOX3* expression in the breast cancer cell line.

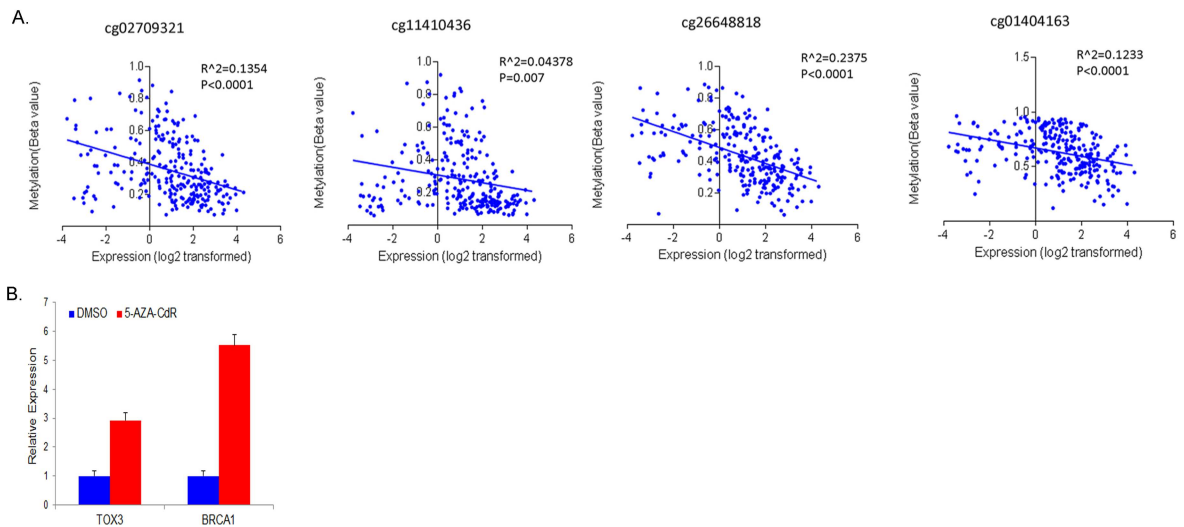


Fig 5. Expression of *TOX3* is correlated with the promoter methylation in breast tumors and increased by 5-aza-CdR treatment. (A) TCGA breast invasive carcinoma DNA methylation (Illumina Infinium HumanMethylation450 BeadChip, version 2014-05-02) and gene expression microarray (Agilent 244K, G4502A, version 2013-06-02) datasets were analyzed. Scatter plots show a significant inverse correlation between expression of *TOX3* and the promoter methylation at four CpG sites. (B) qRT-PCR results after UACC-3199 cells were treated for four days with 5-aza-CdR. Expression of *TOX3* and *BRCA1* were significantly increased after treatment.

doi:10.1371/journal.pone.0165559.g005

mQTL analysis showed a weak or no correlation between SNP variants and TOX3 methylation

Because our data showed that both genetic and epigenetic factors contributed to the molecular subtype-specific expression of TOX3 in breast tumors, we tested for a relation between the two factors. We conducted methylation quantitative trait loci (mQTL) and examined a correlation between SNP variants (rs3803662 and rs4784227) and methylation of the four CpG dinucleotides (cg02709321, cg11410436, cg26648818, and cg01404163). We found that cg02709321 was weakly associated with rs3803662 but not with rs4784227 (Fig 6). None of other CpG sites (cg11410436, cg26648818, and cg01404163) were associated with either rs3803662 or rs4784227. Our data suggest that there is a weak or no correlation between the SNP genotypes and TOX3 promoter methylation.

TOX3 copy number does not correlate with expression

To determine if subtype-specific expression of TOX3/CASC16 is due to deletions or amplifications of the genes, we performed FISH using two home-brewed probes, RP11-132F7 and RP11-748D7, which contain TOX3 and CASC16, respectively (Fig 7A). The chromosome 16-specific DNA repeat probe (CEP16) was used as a control [24] and the probe hybridization efficiency was validated on the normal lymphocyte cell line GM14667. The results showed a normal pattern of two copies of each signal with TOX3:CEP16 ratio of 1.0 (Fig 7B). The normal breast epithelium cell line HMEC also exhibited a normal pattern of 2:2 (TOX3:CEP16, seen in 88% of cells) (S2 Table). Gene copy number of TOX3 were examined in six breast cancer cell lines, including four luminal breast cancer cells (ZR-75-30, MDA-MB-175VII, HCC-202, and T-47D) and two basal-like breast cancer cells (HCC-70 and HCC-1500). Breast cancer cells displayed some degree of abnormal signal patterns including polysomy (ZR-75-30, HCC-70, and MDA-MB-175VII), but no actual gene amplification or deletion was observed (S2 Table).

We next examined TOX3 CNV in a large set of breast tumors using TCGA GISTIC2 CNV dataset and demonstrated similar levels of copy numbers across different subtypes with slightly lower levels in luminal tumors (Fig 7C). No correlation was observed between the copy number and expression of TOX3 ($P = 0.58$). The data suggested no influence of CNV on regulating gene expression of TOX3. Taken together, our data indicate the increased expression of TOX3 in luminal tumors is associated with CpG island methylation as well as SNP genotypes, but not with copy number alterations.

Discussion

In this study, we investigated both genetic and epigenetic factors contributing to TOX3 expression in breast tumors and cell lines. We have confirmed that TOX3 expression is highly up-regulated in luminal breast cancer compared to normal breast tissues or basal-like tumors. The molecular subtype-specific expression of TOX3 in breast tumors is significantly associated with epigenetic modifications at CpG islands of the promoter. Germline genotypes also contribute to TOX3 expression with a significant association of rs3803662 and rs4784227 with TOX3 expression in breast tumors. In contrast, no correlation between CNV and expression of TOX3 gene was observed.

The SNPs at 16q12, close to the TOX3 and CASC16 genes, represent one of the susceptibility loci identified by GWAS, showing strong evidence for breast cancer association across various populations. A functional study following GWAS showed that the risk allele of rs3803662 and the mRNA level of TOX3 predicted adverse outcomes for breast cancer patients [19]. The minor allele of SNP rs3803662 has been shown to correlate with increased breast cancer risk

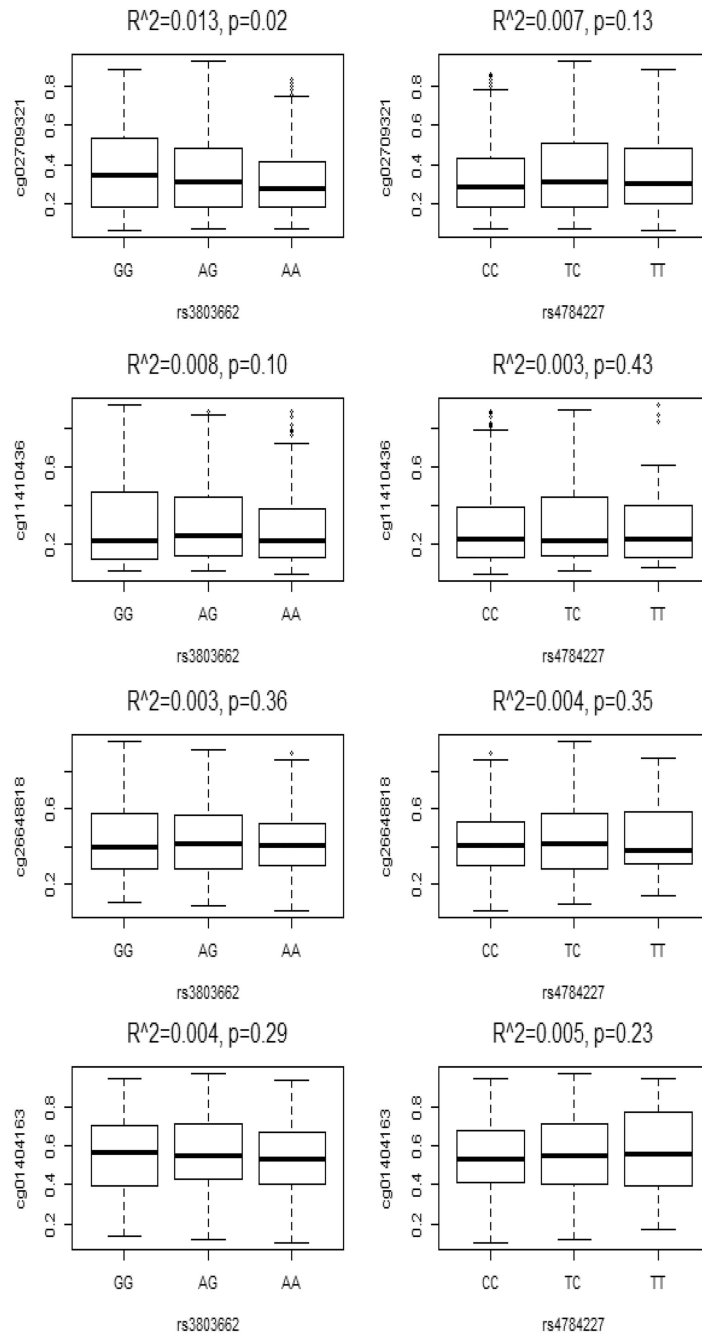


Fig 6. mQTL analyses between SNP genotypes and CpG site methylation of the TOX3 promoter. Linear regression model applied to mQTL analysis. Y-axis represents CpG sites (cg02709321, cg11410436, cg26648818, and cg01404163) and X-axis shows SNP variants (rs3803662 and rs4784227). R squared and p values shown on the top.

doi:10.1371/journal.pone.0165559.g006

and with lower expression of TOX3. The SNP rs4784227, a risk variant for breast cancer in Asian women, was also reported to alter TOX3 expression by disrupting enhancer function through FOXA1 affinity modulation [16]. Through TCGA dataset, we confirmed that rs3803662 and rs4784227 are associated with TOX3 expression in breast tumors. The association of rs3803662 with TOX3 expression has been previously observed, further demonstrating

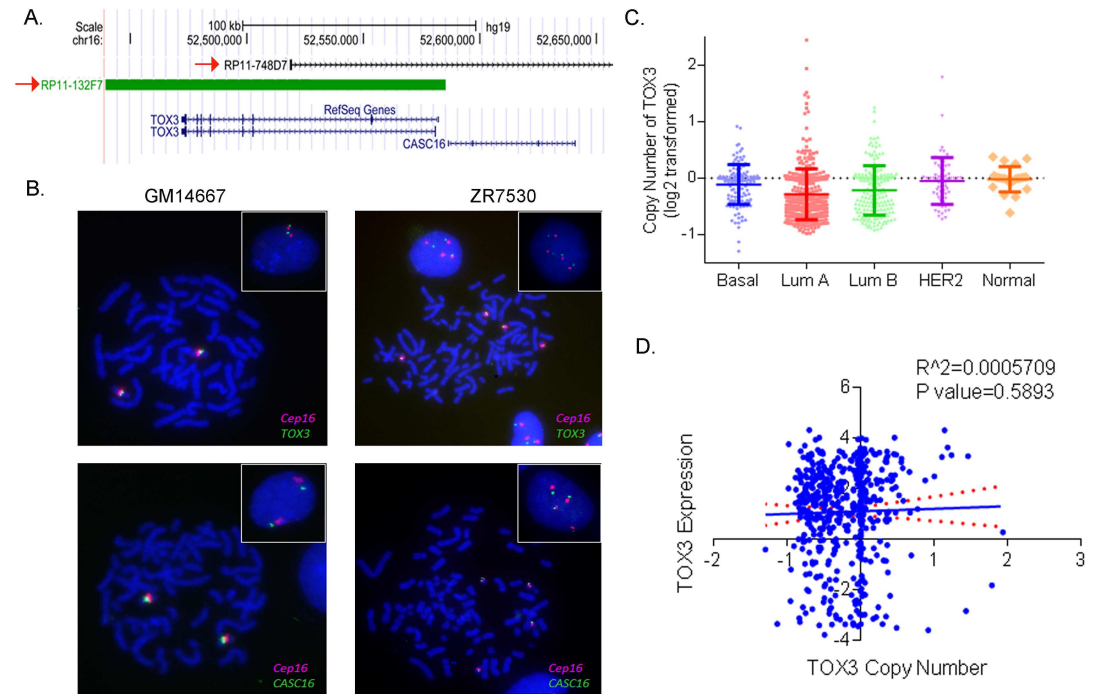


Fig 7. Copy number of *TOX3* in breast tumors. (A) Genomic positions of BAC RP11-132F7 and RP11-748D7 at chromosome 16q12. These clones were selected for homebrewed *TOX3* and *CASC16* FISH probes, respectively. (B) Representative FISH photomicrographs of *TOX3*:*Cep16* and *CASC16*:*Cep16* in GM14667 (control normal lymphocytes) and ZR-75-30 (luminal) cell lines. Cells were counterstained with DAPI (blue), while *TOX3* or *CASC16* is localized by green fluorescent signal, and *Cep16* is localized by a red fluorescent signal. Metaphase and interphase (insert) cells are shown. These results are summarized in S2 Table. (C) TCGA copy number data analysis shows copy number of *TOX3* in breast cancer subtypes. (D) Scatter plot shows no correlation between copy number and expression of *TOX3* in the breast tumors.

doi:10.1371/journal.pone.0165559.g007

the utility of publicly available data [25,26]. Interestingly, the risk variants in women of African ancestry (rs3103104788 or rs3112617) did not show an association with *TOX3* expression in breast tumors, suggesting the effects of risk variants on regulating *TOX3* expression might be risk allele-specific or ethnicity-specific.

In addition to the previous functional study of risk variants, we investigated epigenetic regulation of *TOX3* expression on breast tumors *in silico*, revealing an inverse correlation between subtype-specific expression of *TOX3* and CpG site methylation of the promoter. Transcription factors associated with ER (i.e. *FOXA1*) are also likely to modulate the *TOX3* promoter activity, based on the previous observation that SNP rs4784227 alter *TOX3* expression by disrupting enhancer function through *FOXA1* affinity modulation [16]. This suggests that methylation of *TOX3* may be one of several factors involved in the regulation of *TOX3* in breast cancer.

It is possible that SNPs rs3803662 and rs4784227 modulate CpG island methylation and thus contribute to subtype-specific expression of *TOX3* in breast cancer. We tested the possibility by performing mQTL between the SNP variants and *TOX3* promoter methylation in breast tumors. mQTL analyses showed a weak or no correlation of rs3803662 or rs4784227 with four CpG sites in breast tumors (cg02709321, cg11410436, cg26648818, and cg01404163). Our data suggest that the genetic and epigenetic regulations might be independent and perhaps represent no related events in regulating *TOX3* expression.

Functionally, TOX3 belongs to the high-mobility-group (HMG)-box family of proteins that modify chromatin structure [27]. It is expressed mainly in epithelial cells and targets both anti-apoptotic and proapoptotic transcripts [28]. High expression of TOX3 is correlated with ER and progesterone receptor expression as well as positive lymph nodes [19]. Patients with ER positive tumors and high levels of TOX3 mRNA had shorter overall- and distant metastasis free-survival. Although more mechanical studies are required to determine the regulation of TOX3 expression by epigenetic modifiers, our findings on the upregulation of TOX3 by cytosine methylation raise the possibility of new epigenetic biomarkers for prognosis of aggressive ER positive breast tumors.

Materials and Methods

Cell culture and total RNA and DNA isolation

Human mammary epithelial (HMEC) primary cells were purchased from Lonza and UACC3199 cells were obtained from University of Arizona Cancer Center. Other breast cancer cells were obtained from the American Type Culture Collection (Manassas, VA). Cells were tested negative for mycoplasma contamination and validated for species and unique DNA profile using short tandem repeat analysis by the provider or us. Twenty two breast cancer cell lines were cultured and maintained in the specified media. Breast epithelium cell lines 184A1 and 184B5 as well as primary cell line HMEC were cultured in MEGM medium (Lonza CC-3150). Breast cancer cell lines CAMA1 and MCF7 (with additional 0.01mg/ml bovine insulin) were cultured in MEM medium (Corning 10–010) with 10% FBS. HS578T was cultured in DMEM medium (Gibco 11995–065) with 10% FBS and 0.01mg/ml bovine insulin. All other cell lines were cultured in RPMI 1640 medium (Gibco 11875–093) with 10% FBS and 1% HEPES. Total RNA and DNA were isolated from cell lines using the RNeasy Mini Kit (Qiagen, Montgomery, MD, USA) and GentraPuregene Cell Kit (Qiagen, Montgomery, MD, USA), respectively. The integrity of RNAs was validated by bio-analyzer at the University of Chicago Genomics Core Facility. RNAs with minimum RNA integrity number of 8 were applied to cDNA synthesis.

qRT-PCR

cDNA was generated by SuperScript III First-Strand synthesis Super Mix for quantitative real-time polymerase chain reaction (qRT-PCR)(Life Technologies, Carlsbad, CA, USA). The assessment of mRNA levels was performed using a 7900HT Fast Real-Time PCR system (Life Technologies, Carlsbad, CA, USA) using Power SYBR Green PCR Master Mix (Life Technologies) or TaqMan Gene Expression Master Mix (Life Technologies). All real-time PCRs were performed in four replications, and the fold change in expression of mRNAs was calculated using the $\Delta\Delta C_t$ method, with rRNA18S as an internal control.

Fluorescent In Situ Hybridization (FISH) with home-brewed probes

To evaluate copy number of *TOX3* and *CASC16* in breast cancer cell lines, we carried out FISH using two home-brewed probes of RP11-132F7 and RP11-748D7. These BAC clones were selected from the UCSC genome browser as it contains the region of 16q12 that covers *TOX3* and *CASC16* gene respectively (<http://genome.ucsc.edu/>). The probes were prepared using the Abbott Nick Translation Kit (Abbott Laboratories, Abbott Park, IL) and labeled with Spectrum Green. Cell lines were scored and digital images were obtained using the Zeiss AXIO IMAGER Z2 microscope and Zeiss AxioCamMRm Rev 3 Monochromatic Camera. For each cell line, signals were counted in 60 metaphase and interphase cells with well-defined nuclei and averaged.

The ratio of the *TOX3* and *CASC16* signal to *CEP16* signal and percentage of each signal pattern was used to determine copy number in each cell line. Polysomy was defined as greater than three copies of a chromosome. The ratio ranges for deletion and amplification of *TOX3*: *CEP16* and *CASC16*:*CEP16* are as follows: deletion is < 0.8 , > 1 and < 2 is gain, and > 2 is amplification.

Bisulfite sequencing

Four CpG islands in the *TOX3* promoter were selected from TCGA data. Bisulfite modification of 0.5 μ g of DNA from breast cancer cell lines was carried out with the EZ DNA Methylation-Gold™ kit (Zymo Research, Irvine, CA) according to the manufacturer's protocol. The PCR reactions consisted of the bisulfite-modified DNA and ZymoTaq (Zymo Research, Irvine, CA) with forward (5' -TTAGGTTTTGGGTTAGTAAGGTGTG-3') and reverse (5' -ACCCCCTTCC TTCTTCATAAATAC-3') primers, at an annealing temperature optimized for each primer. The PCR products were run on a gel to confirm specificity. After cleaning PCR products with Exonuclease I (New England BioLabs Inc, Ipswich, MA) and Shrimp Alkaline Phosphatase (Affymetrix, Santa Clara, CA), Sanger sequencing was performed at the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility. Chromatograms of DNA sequencing were read on Chromas Lite 2.1.1 (Technelysium, South Brisbane, Australia).

Treatment of 5-Aza-2'-Deoxycytidine

UACC-3199 was seeded at a density of 3×10^5 per well of 6 well plate. One day after plating, cells were treated with 5-Aza-2'-Deoxycytidine (5-aza-CdR) at 8 μ M dose in DMSO and incubated for four days. Control sample was treated with same amount of DMSO. Nucleic acids were extracted using AllPrep DNA/RNA/Protein Mini Kit (Qiagen, Montgomery, MD) and subjected to qRT-PCR analysis.

TCGA data and statistical analysis

TCGA breast invasive carcinoma CNV (GISTIC2 method, version 2014-05-09), RNA-seq (Illumina HiseqV2, version 2014-05-09), gene expression microarray (Agilent 244K, G4502A, version 2013-06-02) and DNA methylation (Illumina Infinium HumanMethylation450 Bead-Chip, version 2014-05-02) datasets were extracted from UCSC Cancer Browser (<https://genome-cancer.ucsc.edu/>), along with the clinical-pathological phenotypes. In the analyses, we excluded samples from male patients, samples with no PAM50 subtype information, or samples showing discrepancy between PAM50 subtypes and pathological analyses (sample types). We performed the one-way ANOVA and Tukey's multiple comparison tests to compare the methylation level across the breast tumor subtype groups. Statistical analysis was conducted and plots were generated by GraphPad Prism 6.0 (GraphPad Software, Inc., La Jolla, CA), R or Stata.

Supporting Information

S1 Table. eQTL results.

(XLSX)

S2 Table. Gene copy number of *TOX3*.

(DOC)

Acknowledgments

The authors thank Dr. Charles M. Perou (University of North Carolina, Chapel Hill, NC) for generously providing the University of North Carolina dataset and Dr. Albert C. Yeh (University of Chicago) for critically reviewing the manuscript. We also thank Dr. Tatyana Grushko and Ms. Maria Gomez for assistance with FISH experiments and data interpretation, and Ms. Mariann Coyle for assistance with the cell culturing. This work was supported by NIH R01 CA142996-01 (OIO), NIH Specialized Program of Research Excellence in Breast Cancer P50 CA125183 (OIO), Breast Cancer Research Foundation (YH and OIO), and the Ralph and Marion Falk Medical Research Trust (OIO). OIO is American Cancer Society Clinical Research Professor.

Author Contributions

Conceptualization: YH DH OIO.

Data curation: YZ.

Formal analysis: YZ DH YH.

Funding acquisition: OIO YH.

Investigation: YH JZ YZ DH OIO.

Methodology: YH JZ YZ DH.

Project administration: YH OIO.

Resources: OIO.

Supervision: YH OIO.

Validation: YH JZ.

Visualization: YH.

Writing – original draft: YH.

Writing – review & editing: YH JZ YZ DH OIO.

References

1. Turnbull C, Rahman N (2008) Genetic predisposition to breast cancer: past, present, and future. *Annu Rev Genomics Hum Genet* 9: 321–345. doi: [10.1146/annurev.genom.9.081307.164339](https://doi.org/10.1146/annurev.genom.9.081307.164339) PMID: [18544032](https://pubmed.ncbi.nlm.nih.gov/18544032/)
2. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447: 1087–1093. doi: [10.1038/nature05887](https://doi.org/10.1038/nature05887) PMID: [17529967](https://pubmed.ncbi.nlm.nih.gov/17529967/)
3. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, et al. (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 39: 870–874. doi: [10.1038/ng2075](https://doi.org/10.1038/ng2075) PMID: [17529973](https://pubmed.ncbi.nlm.nih.gov/17529973/)
4. Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M (2010) Genetic susceptibility to breast cancer. *Mol Oncol* 4: 174–191. doi: [10.1016/j.molonc.2010.04.011](https://doi.org/10.1016/j.molonc.2010.04.011) PMID: [20542480](https://pubmed.ncbi.nlm.nih.gov/20542480/)
5. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, et al. (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 45: 353–361, 361e351–352. doi: [10.1038/ng.2563](https://doi.org/10.1038/ng.2563) PMID: [23535729](https://pubmed.ncbi.nlm.nih.gov/23535729/)
6. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, et al. (2015) Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 47: 373–380. doi: [10.1038/ng.3242](https://doi.org/10.1038/ng.3242) PMID: [25751625](https://pubmed.ncbi.nlm.nih.gov/25751625/)

7. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, et al. (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 39: 865–869. doi: [10.1038/ng2064](https://doi.org/10.1038/ng2064) PMID: [17529974](https://pubmed.ncbi.nlm.nih.gov/17529974/)
8. Han W, Woo JH, Yu JH, Lee MJ, Moon HG, et al. (2011) Common genetic variants associated with breast cancer in Korean women and differential susceptibility according to intrinsic subtype. *Cancer Epidemiol Biomarkers Prev* 20: 793–798. doi: [10.1158/1055-9965.EPI-10-1282](https://doi.org/10.1158/1055-9965.EPI-10-1282) PMID: [21415360](https://pubmed.ncbi.nlm.nih.gov/21415360/)
9. Long J, Shu XO, Cai Q, Gao YT, Zheng Y, et al. (2010) Evaluation of breast cancer susceptibility loci in Chinese women. *Cancer Epidemiol Biomarkers Prev* 19: 2357–2365. doi: [10.1158/1055-9965.EPI-10-0054](https://doi.org/10.1158/1055-9965.EPI-10-0054) PMID: [20699374](https://pubmed.ncbi.nlm.nih.gov/20699374/)
10. Long J, Cai Q, Shu XO, Qu S, Li C, et al. (2010) Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet* 6: e1001002. doi: [10.1371/journal.pgen.1001002](https://doi.org/10.1371/journal.pgen.1001002) PMID: [20585626](https://pubmed.ncbi.nlm.nih.gov/20585626/)
11. Huo D, Zheng Y, Ogundiran TO, Adebamowo C, Nathanson KL, et al. (2012) Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis* 33: 835–840. doi: [10.1093/carcin/bgs093](https://doi.org/10.1093/carcin/bgs093) PMID: [22357627](https://pubmed.ncbi.nlm.nih.gov/22357627/)
12. Zheng Y, Ogundiran TO, Falusi AG, Nathanson KL, John EM, et al. (2013) Fine mapping of breast cancer genome-wide association studies loci in women of African ancestry identifies novel susceptibility markers. *Carcinogenesis* 34: 1520–1528. doi: [10.1093/carcin/bgt090](https://doi.org/10.1093/carcin/bgt090) PMID: [23475944](https://pubmed.ncbi.nlm.nih.gov/23475944/)
13. Chen F, Chen GK, Millikan RC, John EM, Ambrosone CB, et al. (2011) Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. *Hum Mol Genet* 20: 4491–4503. doi: [10.1093/hmg/ddr367](https://doi.org/10.1093/hmg/ddr367) PMID: [21852243](https://pubmed.ncbi.nlm.nih.gov/21852243/)
14. Ruiz-Narvaez EA, Rosenberg L, Cozier YC, Cupples LA, Adams-Campbell LL, et al. (2010) Polymorphisms in the TOX3/LOC643714 locus and risk of breast cancer in African-American women. *Cancer Epidemiol Biomarkers Prev* 19: 1320–1327. doi: [10.1158/1055-9965.EPI-09-1250](https://doi.org/10.1158/1055-9965.EPI-09-1250) PMID: [20406955](https://pubmed.ncbi.nlm.nih.gov/20406955/)
15. Wasserman NF, Aneas I, Nobrega MA (2010) An 8q24 gene desert variant associated with prostate cancer risk confers differential *in vivo* activity to a MYC enhancer. *Genome Res* 20: 1191–1197. doi: [10.1101/gr.105361.110](https://doi.org/10.1101/gr.105361.110) PMID: [20627891](https://pubmed.ncbi.nlm.nih.gov/20627891/)
16. Cowper-Salari R, Zhang X, Wright JB, Bailey SD, Cole MD, et al. (2012) Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* 44: 1191–1198. doi: [10.1038/ng.2416](https://doi.org/10.1038/ng.2416) PMID: [23001124](https://pubmed.ncbi.nlm.nih.gov/23001124/)
17. Shan J, Dsouza SP, Bakhru S, Al-Azwani EK, Ascierto ML, et al. (2013) TNRC9 downregulates BRCA1 expression and promotes breast cancer aggressiveness. *Cancer Res* 73: 2840–2849. doi: [10.1158/0008-5472.CAN-12-4313](https://doi.org/10.1158/0008-5472.CAN-12-4313) PMID: [23447579](https://pubmed.ncbi.nlm.nih.gov/23447579/)
18. Jones JO, Chin SF, Wong-Taylor LA, Leaford D, Ponder BA, et al. (2013) TOX3 mutations in breast cancer. *PLoS One* 8: e74102. doi: [10.1371/journal.pone.0074102](https://doi.org/10.1371/journal.pone.0074102) PMID: [24069272](https://pubmed.ncbi.nlm.nih.gov/24069272/)
19. Gudmundsdottir ET, Barkardottir RB, Arason A, Gunnarsson H, Amundadottir LT, et al. (2012) The risk allele of SNP rs3803662 and the mRNA level of its closest genes TOX3 and LOC643714 predict adverse outcome for breast cancer patients. *BMC Cancer* 12: 621. doi: [10.1186/1471-2407-12-621](https://doi.org/10.1186/1471-2407-12-621) PMID: [23270421](https://pubmed.ncbi.nlm.nih.gov/23270421/)
20. Bird AP (1986) CpG-rich islands and the function of DNA methylation. *Nature* 321: 209–213. doi: [10.1038/321209a0](https://doi.org/10.1038/321209a0) PMID: [2423876](https://pubmed.ncbi.nlm.nih.gov/2423876/)
21. Millis RM (2011) Epigenetics and hypertension. *Curr Hypertens Rep* 13: 21–28. doi: [10.1007/s11906-010-0173-8](https://doi.org/10.1007/s11906-010-0173-8) PMID: [21125351](https://pubmed.ncbi.nlm.nih.gov/21125351/)
22. Vo AT, Millis RM (2012) Epigenetics and breast cancers. *Obstet Gynecol Int* 2012: 602720. doi: [10.1155/2012/602720](https://doi.org/10.1155/2012/602720) PMID: [22567014](https://pubmed.ncbi.nlm.nih.gov/22567014/)
23. Wei M, Grushko TA, Dignam J, Hagos F, Nanda R, et al. (2005) BRCA1 promoter methylation in sporadic breast cancer is associated with reduced BRCA1 copy number and chromosome 17 aneusomy. *Cancer Res* 65: 10692–10699. doi: [10.1158/0008-5472.CAN-05-1277](https://doi.org/10.1158/0008-5472.CAN-05-1277) PMID: [16322213](https://pubmed.ncbi.nlm.nih.gov/16322213/)
24. Baumgartner A, Weier JF, Weier HU (2006) Chromosome-specific DNA repeat probes. *J Histochem Cytochem* 54: 1363–1370. doi: [10.1369/jhc.6A6974.2006](https://doi.org/10.1369/jhc.6A6974.2006) PMID: [16924124](https://pubmed.ncbi.nlm.nih.gov/16924124/)
25. Riaz M, Berns EM, Sieuwerts AM, Ruijter R, de Weerd V, et al. (2012) Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival, and mRNA expression of the nearest genes. *Breast Cancer Res Treat* 133: 843–851. doi: [10.1007/s10549-011-1663-3](https://doi.org/10.1007/s10549-011-1663-3) PMID: [21748294](https://pubmed.ncbi.nlm.nih.gov/21748294/)
26. Li Q, Seo JH, Stranger B, McKenna A, Pe'er I, et al. (2013) Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* 152: 633–641. doi: [10.1016/j.cell.2012.12.034](https://doi.org/10.1016/j.cell.2012.12.034) PMID: [23374354](https://pubmed.ncbi.nlm.nih.gov/23374354/)

27. O'Flaherty E, Kaye J (2003) TOX defines a conserved subfamily of HMG-box proteins. *BMC Genomics* 4: 13. doi: [10.1186/1471-2164-4-13](https://doi.org/10.1186/1471-2164-4-13) PMID: [12697058](https://pubmed.ncbi.nlm.nih.gov/12697058/)
28. Dittmer S, Kovacs Z, Yuan SH, Siszler G, Kogl M, et al. (2011) TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex. *J Cell Sci* 124: 252–260. doi: [10.1242/jcs.068759](https://doi.org/10.1242/jcs.068759) PMID: [21172805](https://pubmed.ncbi.nlm.nih.gov/21172805/)