Published in final edited form as:

Cancer Discov. 2013 July ; 3(7): 812–825. doi:10.1158/2159-8290.CD-13-0038.

Selective Estrogen Receptor Modulators and Pharmacogenomic Variation in ZNF423 Regulation of BRCA1 Expression: Individualized Breast Cancer Prevention

James N. Ingle^{1,*}, Mohan Liu^{2,*}, D. Lawrence Wickerham^{3,4}, Daniel J. Schaid⁵, Liewei Wang², Taisei Mushioka⁶, Michiaki Kubo⁶, Joseph P. Costantino^{4,7}, Victor G. Vogel⁸, Soonmyung Paik⁴, Matthew P. Goetz¹, Matthew M. Ames², Gregory D. Jenkins⁵, Anthony Batzler⁵, Erin E. Carlson⁵, David A. Flockhart⁹, Norman Wolmark^{4,10}, Yusuke Nakamura^{6,11}, and Richard M. Weinshilboum²

¹Division of Medical Oncology, Mayo Clinic, Rochester, MN

²Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN

³Section of Cancer Genetics and Prevention, Allegheny General Hospital, Pittsburgh, PA

⁴National Surgical Adjuvant Breast and Bowel Project (NSABP), Pittsburgh, PA

⁵Department of Health Sciences Research, Mayo Clinic, Rochester, MN

⁶RIKEN Center for Genomic Medicine, Tokyo, Japan

⁷NSABP Biostatistical Center and the University of Pittsburgh, Graduate School of Public Health, Department of Biostatistics, Pittsburgh, PA

⁸Department of Medical Oncology, University of Pittsburgh, Pittsburgh, PA, Current address: Cancer Institute, Geisinger Health System. Danville, PA

⁹Indiana University, Indianapolis, IN

¹⁰Department of Human Oncology, Allegheny General Hospital, Pittsburgh PA

Abstract

The selective estrogen receptor modulators (SERMs) tamoxifen and raloxifene can reduce the occurrence of breast cancer in high risk women by 50%, but this FDA-approved prevention therapy is not often used. We attempted to identify genetic factors that contribute to variation in SERM breast cancer prevention using DNA from the NSABP P-1 and P-2 breast cancer prevention trials. An initial discovery genome-wide association study identified common single nucleotide polymorphisms (SNPs) in or near the *ZNF423* and *CTSO* genes that were associated with breast cancer risk during SERM therapy. We then showed that both *ZNF423* and *CTSO* participated in the estrogen-dependent induction of *BRCA1* expression, in both cases with SNP-

Reprint requests: James N. Ingle, M.D., Mayo Clinic, 200 First Street S.W., Rochester, MN 55905, ingle.james@mayo.edu, phone: 507-284-4857, fax 507-284-1803.

¹¹Current address: University of Chicago, Chicago, IL.

*Drs. Ingle and Liu contributed equally to this work.

Potential conflicts of interest: The authors declare no potential conflicts of interest

The data from this genome-wide association study have been deposited in dbGaP and the study title is "A Genome-Wide Association Study in Participants Experiencing Breast Cancer Events in High-Risk Postmenopausal Women Receiving Selective Estrogen Receptor Modulators on NSABP Trials P-1 and P-2. A Collaboration Between the NIH Pharmacogenetics Research Network and the RIKEN Yokohama Institute Center for Genomic Medicine." The dbGaP Study Accession number is phs000305.v1.p1, and the URL is http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000305.v1.p1

dependent variation in induction. *ZNF423* appeared to be an estrogen-inducible *BRCA1* transcription factor. The odds ratio for differences in breast cancer risk during SERM therapy for subjects homozygous for both protective or both risk alleles for *ZNF423* and *CTSO* was 5.71.

Keywords

tamoxifen; raloxifene; breast cancer prevention; *ZNF423*; *CTSO*; *BRCA1*; single nucleotide polymorphisms; genome-wide association study

INTRODUCTION

Worldwide, breast cancer is the most frequently diagnosed cancer and most frequent cause of cancer death in women (1). In high risk women, breast cancer can be prevented by treatment with selective estrogen receptor modulators (SERMs), drugs that compete with estrogens for binding to the estrogen receptor (ER). Five years of SERM therapy with tamoxifen or raloxifene can reduce the occurrence of breast cancer in these women by one-half. The largest and most influential of a series of SERM breast cancer prevention trials were the double-blind, placebo-controlled National Surgical Adjuvant Breast and Bowel Project (NSABP) P-1 trial of tamoxifen (2) and the double-blind NSABP P-2 trial that compared raloxifene with tamoxifen (3, 4). Combined, these studies involved over 33,000 women (2–4) and were the basis for US Food and Drug Administration (FDA) approval of both drugs for breast cancer prevention. However, in spite of the striking results of P-1 and P-2, as well as other SERM trials (5), these drugs are not widely prescribed to prevent breast cancer (6, 7). That is true both because of the large number (about 51) of patients who must be exposed to five years of SERM therapy to prevent a single case of breast cancer and because of the occurrence of rare, but serious SERM-related adverse drug responses including deep vein thrombosis with pulmonary embolus and endometrial carcinoma (8). Acceptance of SERM prevention would be expected to improve if we could develop a more highly individualized approach to SERM-based breast cancer prevention, thus resulting in a more favorable benefit/risk ratio.

The present study was performed in an attempt to apply genome-wide techniques to identify single nucleotide polymorphisms (SNPs) associated with risk for the occurrence of breast cancer in women who were treated with tamoxifen or raloxifene during the P-1 and P-2 breast cancer prevention trials and then, of equal importance, to attempt to understand mechanisms associated with the function of those SNPs and related genes. It should also be emphasized that the participants entered on the P-1 and P-2 trials represent approximately 60% of all those world-wide who have participated in SERM breast cancer prevention trials.

The series of studies described subsequently began with a nested discovery case-control genome-wide association study (GWAS). That study identified SNPs in the *ZNF423* gene on chromosome 16 that were associated with decreased risk for the occurrence of breast cancer during SERM therapy, as well as SNPs near the *CTSO* gene on chromosome 4 that were associated with increased risk. There had been no prior reports that either of these genes might be related to SERM effect, to estrogens or to breast cancer risk. Because SERMs interact with the ER, we then used several different cell lines to show that estradiol (E2) induced both *ZNF423* and *CTSO* expression, but only for wild type (WT), not for variant SNP genotypes. These experiments, as well as several other studies described subsequently, were made possible by our use of a panel of 300 lymphoblastoid cell lines (LCLs) for which we had generated dense SNP and mRNA expression data – a model system that has already demonstrated its power for both generating pharmacogenomic hypotheses (9–11) and for testing hypotheses that arise from clinical GWAS (12–14).

Because the phenotype in the present study was the occurrence of breast cancer, we next focused our attention on the possible relationship of the estrogen-dependent induction of ZNF423 and CTSO expression to that of BRCA1, an important breast cancer risk gene (15, 16). BRCA1 expression is known to be estrogen-inducible, but that process is not thought to result from direct interaction of liganded ER α with the *BRCA1* promoter, and mechanisms underlying the estrogen-dependent expression of BRCA1 are not well understood (17). Our experiments showed that the estrogen-dependent induction of the expression of both ZNF423 and CTSO also induced the expression of BRCA1. However, surprisingly, in the presence of tamoxifen or raloxifene, the SNP- and estrogen-dependent induction of ZNF423 was reversed and it occurred only for variant but not WT *ZNF423* SNP genotypes, which is compatible with the observed decrease in breast cancer risk in P-1 and P-2 participants who carried variant *ZNF423* SNP sequences. The discovery of these novel mechanisms for the estrogen-dependent induction of *BRCA1* and for SNP-dependent variation in SERM effect has obvious implications for the individualization of SERM-dependent breast cancer prevention.

RESULTS

GWAS Genotyping and Analysis

The initial step in this series of experiments was performance of a discovery GWAS that involved 592 cases (i.e., participants who developed breast cancer while on SERM therapy) and 1171 matched controls selected from the 33,000 participants enrolled in the NSABP P-1 and P-2 breast cancer prevention trials. As shown in Supplementary Table S1, characteristics of the cases and controls were balanced. A total of 592,236 SNPs were genotyped by the RIKEN Center for Genomic Medicine using the Illumina Human610-Quad BeadChip and 547,356 SNPs were carried forward for analysis after appropriate quality control (as described in the Methods section). The quantile-quantile plot for the conditional logistic regression results (Supplementary Fig. S1) showed that the inflation factor, lambda, was 1.021, indicating little influence of population stratification (18). To avoid local genomic linkage disequilibrium (LD) (19–21), 7,606 SNPs that were uncorrelated with each other (Pearson correlation < 0.063) were selected for use in the EigenStrat analyses (19, 21). Nine eigenvalues with Tracy-Widom p-values < 0.05 were identified (22, 23), and none differed significantly between cases and controls (i.e., all p-values > 0.10 by Wilcoxon rank sum test).

The Manhattan Plot in Fig. 1A shows results for conditional logistic analyses for the development of breast cancer, and the 11 Illumina platform genotyped SNPs with the lowest p-values are listed in Table 1. The association of these 11 SNPs with breast cancer risk did not differ according to trial (P-1 vs. P-2), treatment (tamoxifen vs. raloxifene), type of breast cancer (Supplementary Table S2) or time to the development of breast cancer (Supplementary Table S3). The three SNPs with the lowest p-values were on chromosome 16 (rs8060157, $P=2.12E-06$), chromosome 13 (rs9510351, $P=2.76E-06$), and chromosome 4 (rs10030044, $P=3.63E-06$). Although none of these SNPs met criteria for genome-wide significance, since we had used the largest sample set available, as well as the majority of samples available worldwide, to perform the GWAS, we pursued the results functionally and mechanistically. That decision was supported by the observations described subsequently that provided novel insight into the estrogen-dependent regulation of an important breast cancer risk gene, *BRCA1*, as well as a novel genetic mechanism for variation in SERM effect.

We chose to focus on the functional implications of the SNP signals on chromosomes 16 and 4 because both of those sets of SNPs were either in or near a gene, while the chromosome 13 SNP was not close to any gene (Table 1). The results for the imputation of

ungenotyped SNPs located within 200 kb of the lowest p-value SNPs on chromosomes 16 and 4 are shown in Fig. 1B and 1C, and the top SNPs after imputation, initially using HapMap2 (24) and later 1000 Genomes project data (25), are listed in Supplementary Table S4. The HapMap2 imputed SNPs with low p-values were also genotyped (fine mapped) (see Supplementary Methods) to ensure their validity. Finally, the cytochrome P450 2D6 gene, *CYP2D6*, a gene encoding an enzyme involved in tamoxifen metabolism, has been reported to influence tamoxifen effect (26, 27). Therefore, we also determined *CYP2D6* genotypes and inferred *CYP2D6* phenotypes for the participants included in our study, but neither was significantly associated with the occurrence of breast cancer in these participants (28). Therefore, *CYP2D6* status was not included in the analysis as a covariate.

Functional Genomics of Chromosome 16 SNPs

The minor alleles for the rs8060157 and rs11076499 SNPs on chromosome 16 were associated with decreased risk for the development of breast cancer (odds ratio of 0.70, Table 2), and both SNPs mapped to intron 2 of the *ZNF423* gene, a gene that encodes a putative zinc finger protein. Imputation and subsequent genotyping of imputed SNPs in this region of chromosome 16 identified eight additional SNPs (rs6500258, rs9940645, rs7499405, rs60841334, seq5958, rs1861343, rs5816658, and rs12446233) with low p-values ($p=2.12\text{E-}06$ to $4.22\text{E-}06$), all of which mapped to the second intron of *ZNF423* (Supplementary Table S4D). All of these SNPs were in tight LD (Pearson correlation $r^2>0.9$), with minor allele frequencies (MAFs) that varied from 0.39 to 0.41.

Since SERMs interact with the ER, we next determined whether the expression of *ZNF423* might be estrogen-dependent. Specifically, we incubated ZR75-30 and ZR75-1 breast cancer cells with 0.1 nM estradiol E2 and observed that *ZNF423* mRNA expression was induced 75- and 13-fold after 32 and 36 hours (h) in the two cell lines, respectively (Supplementary Fig. S2). These experiments served to link the gene containing the chromosome 16 SNPs that we identified during the GWAS to the mechanism of action of the drugs used to treat these women. The next question was the possible relationship of SNP genotypes to this phenomenon. To answer that question, we took advantage of cell lines selected on the basis of *ZNF423* genotype from a genomic data-rich panel of LCLs that has already proven to be a powerful tool for generating and testing pharmacogenomic hypotheses (9–13). To perform those experiments, we stably transfected LCLs homozygous for either WT or variant *ZNF423* SNP genotypes with ER α , and then exposed them to increasing concentrations of E2. Cell lines homozygous for WT *ZNF423* SNP sequences showed induction of *ZNF423* expression in response to estrogen exposure, while LCLs homozygous for variant SNP sequences did not (Fig. 2A). These studies linked *ZNF423* SNP genotypes to the effect of E2 on the gene containing the SNPs. The next question was how the estrogen- and SNP-dependent induction of *ZNF423* might be related to breast cancer risk. That question led us to BRCA1.

It has been known for over a decade that *BRCA1* is estrogen inducible, although the mechanism has remained unclear (17). Since the phenotype for our study was breast cancer risk during five years of SERM therapy, we tested the hypothesis that *ZNF423* might influence the expression of *BRCA1*, a gene that can play an important role in breast cancer risk (29). Specifically, we performed qRT-PCR for *BRCA1* (Fig. 2B) using RNA preparations from the LCLs for which data are shown in Fig. 2A. We found that *BRCA1* expression increased in the cell lines with WT sequences for the *ZNF423* SNPs after exposure to increasing concentrations of E2, but not in LCLs homozygous for variant SNP sequences (Fig. 2B). Those observations suggested that *ZNF423*, a zinc finger protein, might function as a transcription factor and, after estrogen induction, might induce *BRCA1* mRNA expression.

ZNF423 and BRCA1 Transcription—As a first step in testing the hypothesis that ZNF423 might influence *BRCA1* transcription, we cloned approximately 2000 bp of the *BRCA1* 5′-flanking region (5′-FR) into a pGL3-Basic reporter plasmid (Promega, Madison, WI). This area of *BRCA1* has been reported to include the promoter for the gene (30). After knockdown (KD) of *ZNF423* to 24% of baseline in Hs578T-ERα breast cancer cells (Fig. 2C, upper left), *BRCA1* 5′-FR transcriptional activity in the reporter construct decreased to 2% of baseline (Fig. 2C, bottom left). Conversely, when ZNF423 was overexpressed 3-fold over baseline (Fig. 2C, upper right), *BRCA1* 5′-FR reporter gene transcriptional activity increased 3.6-fold (Fig. 2C, lower right). These observations were compatible with the conclusion that ZNF423 can regulate BRCA1 transcriptional activity, either directly or indirectly.

We next tested the possibility of a direct interaction between ZNF423 and the *BRCA1* promoter by performing a chromatin immunoprecipitation (ChIP) assay using Hs578T-ERα cells. ZNF423 has previously been reported to interact with a 5′-CCGCCC-3′ DNA binding sequence (31). Four of those sequence motifs were present in the 5′-FR of *BRCA1*. Therefore, we designed primer pairs surrounding each of these regions and performed ChIP assays after 24 h exposure of the cells to 0.1 nM E2 to induce the expression of *ZNF423*. We found that a 217 bp PCR amplicon containing one of the putative ZNF423 binding motifs located 1735 to 1518 bp upstream from the ATG in *BRCA1* could bind *ZNF423* (Fig. 2D). Similar results were obtained when these experiments were repeated using U2OS-ERα cells (data not shown). This series of experiments demonstrated that estrogen exposure could induce ZNF423 expression and that ZNF423, in turn, could induce the expression of BRCA1. However, since the estrogen-dependent induction of ZNF423 occurred in cells with WT but not variant SNP genotypes (Fig. 2A), these observations appeared to be at odds with the clinical GWAS data which indicated that variant, not WT, SNP genotypes were associated with decreased risk for breast cancer. Therefore, we once again turned to the LCL model system to determine the effect of SERMs on the stepwise process of estrogen-dependent ZNF423 induction, followed by ZNF423-dependent induction of BRCA1 expression.

ZNF423 SNP Effects Differ in the Presence of SERMs—In the next series of experiments, LCLs stably transfected with ERα that had differing genotypes for the *ZNF423* SNPs were treated with increasing concentrations of E2 alone and then with increasing concentrations of 4-hydroxytamoxifen (an active tamoxifen metabolite) in the presence of E2 in an attempt to mimic *in vitro* the treatment of P-1 and P-2 participants. After exposure of LCLs homozygous for WT SNP genotypes to increasing concentrations of E2, average ZNF423 expression increased approximately 3-fold (Fig. 3A) and, in parallel, BRCA1 expression also increased approximately 3-fold (Fig. 3B). In LCLs homozygous for variant SNP genotypes, average ZNF423 expression was virtually unchanged, as was BRCA1 expression during exposure to increasing E2 concentrations. However, ERα blockade with increasing concentrations of 4-hydroxytamoxifen, always in the presence of 0.01 nM E2, showed a striking reversal of the ZNF423 and BRCA1 expression patterns. Specifically, in LCLs homozygous for variant SNP genotypes, ZNF423 expression increased almost 2-fold and BRCA1 expression increased almost 3-fold when 4-hydroxytamoxifen was present, while, in contrast, expression in cells with WT genotypes decreased to levels at or below baseline (Fig. 3A, 3B). Of importance, similar experiments performed with raloxifene showed very similar patterns of response (Supplementary Fig. S3). These results, which showed enhanced BRCA1 induction in cells with variant genotypes in the presence of SERMs, were compatible with our GWAS observations that variant *ZNF423* SNP genotypes were associated with decreased breast cancer risk during five years of SERM therapy for those enrolled in P-1 and P-2. However, these results also raised the question of the

mechanism underlying the reversal of SNP effects on both *ZNF423* and *BRCA1*, the subject of the next set of experiments.

ZNF423 SNPs and ER α Binding—To determine whether the SNPs in intron 2 of *ZNF423* might alter ER α binding to estrogen response elements (EREs) present in intron 2 of the *ZNF423* gene, we performed ChIP assays with ER α antibody for all five *ZNF423* intron 2 SNPs that had canonical ERE sequence motifs located within 200 bp of the SNP. None of those SNPs either created or disrupted an ERE motif based on predictions of the Transfac transcription factor database. We found that the rs9940645 SNP (Fig. 3C) showed striking differential binding, with greater binding for the WT SNP genotype in the presence of E2, but with a reversal of the binding pattern, with greater binding to the ERE near the variant SNP when 4-hydroxytamoxifen was present (Fig. 3D). Several EREs were located within approximately 240 bp of rs9940645 (Fig. 3C), and the ChIP data shown in Fig. 3D involved amplicons that included both the rs9940645 SNP and the closest ERE. When we repeated the ChIP assay, but did not include the SNP in the amplicon, the SERM-dependent differences in binding were not observed (data not shown).

Finally, to determine whether the differences in binding shown in Fig. 3D might have functional consequences, we performed dual luciferase reporter assays (Fig. 3E). Specifically, we cloned a 500 bp DNA sequence that included either WT or variant sequence for the rs9940645 SNP as well as all of the EREs shown in Fig. 3C into the pGL3 basic reporter plasmid. We also cloned a 1500 bp region of the *ZNF423* promoter downstream of the 500 bp segment, and used these dual insert constructs to transfect IGROV-1 cells, a cell line in which *ZNF423* is highly expressed. After treatment with 0.1 nM E2, the cell lines with the WT SNP sequence displayed 3-fold greater luciferase activity than did the variant SNP sequence. However, when the cells were incubated with 0.1 nM E2 and 10^{-7} 4-hydroxytamoxifen was added, the activity of the variant SNP construct increased while the WT SNP activity decreased, with the variant SNP construct displaying approximately twice the WT sequence activity (Fig. 3E, left panel). We obtained very similar results when we repeated these experiments with 10^{-7} M raloxifene rather than 4-hydroxytamoxifen (Fig. 3E, right panel).

These results moved beyond the binding data shown in Fig. 3D to show that the rs9940645 SNP was also functional in its effects on transcription, at least in these reporter gene constructs. The results of this series of studies indicated that both the functional effect of E2 exposure (Fig. 3A and 3E), and ER α binding in this setting, as determined by ChIP assay (Fig. 3D), could be reversed in the presence of the active tamoxifen metabolite, 4-hydroxytamoxifen, resulting in increased *BRCA1* expression in the presence of 4-hydroxytamoxifen in cells homozygous for variant *ZNF423* SNP genotypes, but with the opposite effect for cells homozygous for WT *ZNF423* SNP genotypes. We next tested the functional implications of these observations using DNA double strand break repair, a process in which *BRCA1* plays an important role, as a phenotype.

ZNF423 is Upstream of *BRCA1* in DNA Double-Strand Break Repair—To determine the possible functional implications of *ZNF423*-dependent regulation of *BRCA1* expression, we used DNA double-strand break repair, an important function of *BRCA1*, as a phenotype (32). In these experiments, we used ZR75-30 cells to perform comet assays after the KD of *ZNF423* and/or *BRCA1* (Fig. 4A). As anticipated, there was evidence of DNA double-strand breaks after the KD of *BRCA1*, but the same was true after the KD of endogenous *ZNF423*. However, *BRCA1* overexpression could overcome the effect of *ZNF423* KD and reverse the comet assay evidence of DNA double-strand breaks. These results indicated that *ZNF423* is located upstream of *BRCA1*, at least in terms of DNA double-strand break repair, and were compatible with a model in which the induction of

ZNF423 results in increased BRCA1 expression. The next question that we addressed was the possible mechanism by which the SNPs on chromosome 4 near the *CTSO* gene, SNPs that were also among the top hits in our discovery GWAS (see Fig. 1A and Table 1) might be related to variation in breast cancer risk during clinical SERM prevention therapy.

CTSO SNPs and Estrogen-Dependent BRCA1 Expression—As pointed out previously, a group of SNPs that mapped to chromosome 4 also had low p-values in the GWAS (Fig. 1, Table 1). The most significant chromosome 4 SNP genotyped on the Illumina GWAS platform was rs10030044 ($p = 2.76 \times 10^{-6}$) but, in this case, variant SNP alleles (MAF=0.45) were associated with increased risk for developing breast cancer, with an odds ratio of 1.42. Imputation revealed two SNPs with lower p-values than that of rs10030044 (rs6835859 and rs4550865) (Fig. 1C and Supplementary Table S4B) that were confirmed by genotyping. We searched the ENCODE database for the top *CTSO* SNPs (Supplementary Table S4B) and only the rs126399271 SNP mapped to an ENCODE transcription factor binding site, in this case STAT3 and CTCF binding.

The rs10030044 SNP, the SNP in this area genotyped in the Illumina platform, was located 5' of the *CTSO* gene, a gene encoding cathepsin O. We began our functional genomic studies of these SNPs by performing a series of experiments similar to those that we performed with ZNF423, starting with a determination of whether *CTSO* might, like ZNF423, be estrogen inducible and, if so, whether it also might be related to *BRCA1* expression.

CTSO mRNA expression was induced by exposure to increasing concentrations of E2 in LCLs stably transfected with ER α (Fig. 4B). Furthermore, this process was SNP-genotype-dependent, with estrogen-dependent induction occurring in cells with WT genotypes, but not in those with variant SNP genotypes (Fig. 4B). Of greatest interest was the fact that, just as was true for ZNF423, BRCA1 was also induced with increasing E2 concentrations, but only in cell lines homozygous for the WT genotype (Fig. 4C). The LCL model system was used, once again, because it provided access to a number of cell lines with known genotypes for the SNP signals detected during the GWAS. Specifically, cells containing the variant chromosome 4 SNP genotypes, a genotype associated with increased risk for developing breast cancer, were minimally induced by estrogen exposure and, furthermore, there was not induction of BRCA1 expression in these cell lines (Fig. 4B). However, cells containing the WT SNP alleles demonstrated increased expression of both *CTSO* and *BRCA1* in response to increasing concentrations of E2 (Fig. 4B). We also examined the SNP-dependent expression of both *CTSO* and *BRCA1* in the presence of E2 and 4-hydroxytamoxifen in a manner similar to that done for ZNF423 (Figs. 3A and B). In the case of *CTSO*, we did not see the pattern of a striking reversal of expression with 4-hydroxytamoxifen that was seen with the ZNF423 SNPs (Supplementary Fig. S4), but this was not unexpected given that the rs6810983 SNP near *CTSO* disrupts an ERE.

A search of the TRANSFAC database for the top *CTSO* SNPs (Table S4B) indicated that only the rs6810983 SNP was predicted to be involved in transcription factor binding. As mentioned previously, the variant genotype for this SNP, disrupted an ERE that we showed by ChIP assay to be functional if the WT sequence was present, but no ER α binding was seen for the variant SNP genotype (Supplementary Fig. S5). Finally, in a fashion similar to ZNF423, *CTSO* appeared to be upstream of BRCA1 because *CTSO* KD resulted in DNA double-strand breaks with comet tails in ZR75-1 breast cancer cells, but this effect could be reversed by BRCA1 overexpression (Fig. 4D).

ZNF423 and *CTSO* SNP Genotypes and Breast Cancer Risk during SERM Therapy—Since our initial GWAS had identified SNPs associated with decreased

(*ZNF423*) and increased (*CTSO*) risk, both of which appeared to be associated with the estrogen-dependent induction of *BRCA1*, we examined their joint effect in the women enrolled in P-1 and P-2. The joint odds ratios for these two sets of SNPs are listed in Table 2 which shows a broad range of relative odds ratios for the development of breast cancer while on SERM therapy for five years, values that ranged from 1.00 (set as the baseline) for women homozygous for both sets of favorable (i.e., low risk) alleles, to 5.71 for women homozygous for unfavorable (i.e., high risk) alleles for both *ZNF423* and *CTSO*. In addition, we examined the top SNPs from chromosomes 4 and 16 (Table 1) and found no differences in their MAFs, whether the invasive cancer was ER positive or negative (data not shown).

DISCUSSION

This study began with a discovery GWAS performed in an attempt to identify genetic factors that might contribute to risk for the development of breast cancer during five years of breast cancer prevention therapy with tamoxifen or raloxifene. We observed SNPs in the *ZNF423* gene on chromosome 16 with variant alleles that were associated with decreased risk and SNPs on chromosome 4, near the *CTSO* gene that were associated with increased risk (Fig. 1 and Table 1). *ZNF423* is a zinc finger transcription factor (33) and *CTSO* encodes cathepsin O, a cysteine proteinase of the papain superfamily, (34). As noted previously, there had been no prior reports of a relationship of either of these genes to SERM effect, to estrogens or to breast cancer risk. Although the p-values for these SNPs did not reach genome-wide significance, (i.e., $p < E-07$), they were near that threshold and, as a result, we felt it important to investigate these signals functionally and mechanistically, especially since our study included the majority of patients worldwide who have been treated with SERMs in prospective breast cancer prevention trials. The fact that the phenotype that we tested was breast cancer risk and that both of these SNP signals were found to be functionally related to the estrogen-dependent expression of an important breast cancer risk gene, *BRCA1*, supports our decision to proceed with functional genomic studies, especially since those studies resulted in novel, biomedically relevant insights.

Specifically, our functional genomic studies showed that the expression of *ZNF423* in cells with WT *ZNF423* SNP sequences was induced by exposure to increasing concentrations of estrogen, but little change was seen for variant SNP genotypes (Fig. 2A). We next examined the possible relationship between *ZNF423* expression and that of *BRCA1*, a gene known to be induced by estrogen exposure through a mechanism that has remained unclear (17), and found that *BRCA1* expression paralleled the SNP-dependent expression that we had observed for *ZNF423* (Fig. 2B), i.e., the SNPs in *ZNF423* were related to the E2-dependent induction of *BRCA1*. We also showed that *ZNF423* overexpression and KD altered the transcriptional activity of the *BRCA1* 5'-flanking region and that *ZNF423*, a zinc finger protein, bound to the 5'-flanking region of *BRCA1*, as demonstrated by ChIP assay (Fig. 2C and 2D). Finally, we examined the effect of SERM exposure on this estrogen and SNP-dependent process for the regulation of *BRCA1* expression and found that the presence of SERMs reversed the expression responses for both *ZNF423* and *BRCA1* in a SNP-dependent manner (Fig. 3A and 3B). Specifically, variant SNP sequences were associated with increased expression of *ZNF423* and *BRCA1* in the presence of the active tamoxifen metabolite, 4-hydroxytamoxifen, and of raloxifene, while the reverse was true for WT SNP genotypes, which is compatible with the association of variant *ZNF423* SNP sequences with protection from breast cancer while subjects were being treated with SERMs. These studies also uncovered a SNP-dependent mechanism for variation in SERM effect, a mechanism involving SNPs near, but not within an ERE.

Another top hit (i.e., lowest p-value) set of SNPs observed during the GWAS mapped near the *CTSO* gene – a gene that we also showed to be induced by estrogen exposure in a SNP-dependent fashion and which also resulted in the downstream induction of *BRCA1* expression (Fig. 4B and 4C). In this case, a variant SNP sequence disrupted and inactivated an ERE. The fact that two of the lowest p-value SNP signals in this study of breast cancer risk were associated with altered expression of *BRCA1* enhanced confidence in the validity of the observed associations. It should be emphasized once again that, although we used a variety of cell lines, predominantly breast cancer cell lines, to perform these experiments, this set of novel mechanistic observations would not have been possible without our use of a genomic-data rich LCL model system that has repeatedly demonstrated its value in both the generation of pharmacogenomic hypotheses and the testing of hypothesis arising from GWAS, especially for antineoplastic therapy (9–14).

The observations reported here have both mechanistic and clinical implications, particularly for individualized breast cancer prevention by SERMs. The differential effects of the *ZNF423* SNPs during SERM exposure in women at high risk for breast cancer also provide novel insights into mechanisms responsible for individual variation in the effect of tamoxifen and raloxifene in breast cancer prevention as well as mechanisms underlying the estrogen-dependent regulation of *BRCA1* expression. Finally, beyond the novel mechanistic insights gained as a result of functional pursuit of SNP signals observed during our discovery GWAS, these observations have potential implications for the selection of women for SERM breast cancer prevention therapy and for the possible reduction of exposure to these drugs of women who have less likelihood of benefit, i.e., implications for individualized breast cancer prevention.

METHODS

Study Design

A nested matched case-control design was used, with matching on the following factors: 1) trial and treatment arm (P-1 tamoxifen, P-2 tamoxifen, P-2 raloxifene); 2) age at trial entry (when controls could not be exactly matched on age, we incremented the age of matching by ± 1 year, in sequence until a match was obtained without exceeding ± 5 years); 3) 5-year predicted breast cancer risk based on the Gail model ($<2.00\%$, $2.01\text{--}3.00$, $3.01\text{--}5.00$, >5.01); 4) history of lobular carcinoma in situ (yes vs. no); 5) history of atypical hyperplasia in breast (yes vs. no); 6) time on study (controls had to be on study at least as long as the time to diagnosis of the breast event for the case). Because 94.2% of the participants on P-1 and P-2 treated with tamoxifen or raloxifene were Caucasian, this study was restricted to only Caucasian subjects to minimize issues with population stratification.

A total of 596 cases and 1176 matched controls were identified by NSABP for inclusion in the GWAS. Six samples had insufficient DNA (two cases and four controls), and an additional three patients (two cases and one control) were excluded after a series of quality control steps, as noted below.

Genotype Quality Control—Two cases and two controls were chosen randomly for use as duplicates for quality control of genotype concordance. The DNA samples were plated into 96-well plates, with cases and controls randomly distributed across the plates. A Caucasian parent-child Centre d'Etude du Polymorphisme Humain (CEPH) trio from the HapMap project was included to check for Mendelian transmission of alleles. For quality control purposes, we included four of our study subjects who had duplicate samples and 38 replicate CEPH samples (child-parent trio). This resulted in a total of 1808 samples genotyped for this study (592 cases, 1174 controls, four duplicates, and 38 CEPH).

For the four duplicate NSABP samples, there were only 15 discordant SNP genotypes (6.5E-6) among the 2,312,687 pairs of SNPs evaluated. For the CEPH trios that were included as controls, there were 284 Mendelian errors (4.9E-4) out of 578,217 genotypes evaluated. These results showed extremely high genotype quality.

Initial quality control analyses suggested that there were either sample mix-ups or race/gender ambiguity for three participants. Using the software STRUCTURE (19, 20), two of the NSABP participants clustered tightly with African-American subjects. Furthermore, one participant could possibly be male or a subject with chromosomal anomalies. This was suggested by using the software PLINK (35) to estimate the fraction of homozygous genotypes on the X chromosome, found to be 0.97 for one participant, suggesting either sample mix-up, or gender ambiguity. We excluded these three participants from the analyses, resulting in a final total of 592 cases and 1171 controls.

Primary Analysis

The primary analysis was based on conditional logistic regression to account for the matched design. However, 42 controls had no matched cases and 2 cases had no matched controls, so those subjects were excluded from the primary analyses. Therefore, the matched analyses included 539 cases, each matched to two controls, and 51 cases, each matched to one control. SNP genotypes were coded as additive effects on the log odds ratio by coding as 0, 1, or 2 for minor allele count. This resulted in a likelihood ratio test with one degree of freedom for each SNP. The primary covariates used to match cases and controls were implicitly controlled in the conditional logistic regression.

Although none of the nine eigenvectors identified, as noted in the RESULTS, differed significantly between cases and controls, we performed sensitivity analyses by including the eigenvectors as covariates (no other covariates were included, although the matched conditional analyses control for the matched covariates). Table 1 and the Tables in the supplementary information present both adjusted and unadjusted analyses for the conditional logistic regression models. As a further sensitivity analysis, we used unconditional logistic regression in order to include the 42 controls who had no matched cases and 2 cases had no matched controls, again adjusting for the 9 eigenvectors. Because adjustment for eigenvectors had little effect on our results, our primary analysis was based on conditional logistic regression, not adjusted for the eigenvectors.

Imputation and Joint SNP Analyses—The regions of interest on chromosomes 4 and 16 were both submitted to imputation. Detailed descriptions of imputation, genotyping after imputation, and joint SNP analyses are included in the Supplementary Material under the heading Supplementary Analyses.

Functional Genomic Studies

Cell Culture: Detailed descriptions of the cell culture experiments are included in the Supplementary Material under the heading Functional Genomic Methods. The LCLs used in this study were obtained from the Coriell Institute and all other cell lines came from the ATCC. All of these cells had been tested and authenticated by the Coriell Institute and ATCC. Original stocks of cells used in our studies had been stored in liquid nitrogen and had not been passaged.

BRCA1 Reporter Plasmids and Transcriptional Activity Assays: U2OS-ER α genomic DNA was used as a template for PCR reactions performed with the following two pairs of primers: BRCA1 Forward, 5'-CTCGAGTGGGGT GAATCTAACATGGCGGACA-3'; BRCA1 Reverse 5'-AGATCTGGAGTGACTGACC GGCTAGGTGG-3'. The amplicons

were cloned directly into the luciferase reporter plasmid pGL3 basic (Promega) after digestion with *NheI* and *BglII* to create a pGL3-BRCA1 construct. The sequences of the inserted amplicons were verified by DNA sequencing.

Hs578T-ER α cells were then seeded in triplicate in 12-well cell culture plates at a concentration of 10^5 cells/well. After 24 h, the cells were transfected using Lipofectamine 2000 (Invitrogen, Grand Island, NY) with 2 μ g of the pGL3-BRCA1 construct and 500 ng pRL-CMV encoding a CMV-driven Renilla luciferase vector (Promega) plus carrier DNA (pGL3 basic). Luciferase assays were performed 48 h after transfection using a luciferase reporter assay system (Promega). The Renilla activity was used to correct for possible variation in transfection efficiency.

LCL Culture and Exposure to E2, 4-Hydroxytamoxifen and Raloxifene: Eight LCLs homozygous for variant genotypes for the *ZNF423* SNPs; seven cell lines with heterozygous genotypes; and eight cell lines homozygous for WT sequences at those SNPs, all stably transfected with ER α as described previously, were used in these studies. These stably transfected LCLs were cultured in RPMI media containing 15% (vol/vol) FBS with 200 μ g/ml Zeocin. Prior to E2 treatment, 2×10^7 cells from each cell line were cultured for 24 h in RPMI media containing 5% (vol/vol) charcoal stripped FBS with 200 μ g/ml Zeocin, followed by culture in the same medium without FBS for another 24 h. All cells were then cultured for 24 h in 6-well plates, with RPMI 1640 media that contained 0, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1.0 and 10 nM E2. 4-hydroxytamoxifen (Sigma-Aldrich, St. Louis, MO) or raloxifene (Sigma-Aldrich) were then added to the same medium containing 0.01 nM E2 at final 4-hydroxytamoxifen concentrations of 10^{-8} , 10^{-7} , 10^{-6} and 10^{-5} μ M, or raloxifene concentrations of 10^{-6} to 10^{-3} μ M, and the cells were cultured for an additional 24 h. Total RNA was isolated from the cells with the RNeasy mini kit (Qiagen, Valencia, CA). Two hundred ng of total RNA was then used to perform qRT-PCR with *ZNF423*, *BRCA1*, and ER α primers. *ZNF423* and *BRCA1* expression levels were normalized on the basis of ER α expression in each cell line.

Chromatin Immunoprecipitation (ChIP) and Reporter Gene Assays: Detailed descriptions of the ChIP and reporter gene assays are included in the Supplementary Material under the heading Functional Genomic Methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: These studies were supported in part by NIH grants U19 GM61388 (JNI, DJS, LW, GDJ, AB, EEC, RMW) (The Pharmacogenomics Research Network), P50CA116201 (Mayo Clinic Breast Cancer Specialized Program of Research Excellence) (JNI, MPG), U10CA37377 (NW) and U10CA69974 (JPC), U24CA114732 (NW), U01GM63173 (DAF), and the RIKEN Center for Genomic Medicine and the Biobank Japan Project funded by the Ministry of Education, Culture, Sports, Science and Technology, Japan (TM, MK, YN).

The authors acknowledge the women who participated in the NSABP P-1 and P-2 clinical trials and provided DNA and consent for its use in genetic studies.

Abbreviation list

5'FR	5' flanking region
CEPH	Centre d'Etude du Polymorphisme Humain

ChIP	chromatin immunoprecipitation
CTSO	cathepsin O
CYP2D6	cytochrome P450 2D6
E2	estradiol
ER	estrogen receptor
ERE	estrogen response element
GWAS	genome-wide association study
KD	knock down
LCLs	lymphoblastoid cell lines
LD	linkage disequilibrium
MAF	minor allele frequency
NSABP	National Surgical Adjuvant Breast and Bowel Project
SERM	selective estrogen receptor modulator
SNP	single nucleotide polymorphism
WT	wild type

References

1. globocan.iarc.fr/factsheets/cancers/breast.asp aJ, 2013.
2. Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for prevention of breast cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst.* 1998; 90:1371–88. [PubMed: 9747868]
3. Vogel VG, Costantino JP, Wickerham D, Cronin WM, Cecchini RS, Atkins JN, et al. Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: The NSABP Study of Tamoxifen and Raloxifene (STAR) P-2 Trial. *JAMA.* 2006; 295:2727–41. [PubMed: 16754727]
4. Vogel VG, Costantino JP, Wickerham DL, Cronin WM, Cecchini RS, Atkins JN, et al. Update of the National Surgical Adjuvant Breast and Bowel Project Study of Tamoxifen and Raloxifene (STAR) P-2 Trial: preventing breast cancer. *Cancer Prev Res.* 2010; 3:696–706.
5. Cuzick J, Powles T, Veronesi U, Forbes J, Edwards R, Ashley S, et al. Overview of the main outcomes in breast-cancer prevention trials. *The Lancet.* 2003; 361:296–300.
6. Ropka ME, Keim J, Philbrick JT. Patient decisions about breast cancer chemoprevention: a systematic review and meta-analysis. *J Clin Oncol.* 2010; 28:3090–5. [PubMed: 20458026]
7. Armstrong K, Quistberg D, Micco E, Domchek S, Guerra C. Prescription of tamoxifen for breast cancer prevention by primary care physicians. *Archiv Int Med.* 2006; 166:2260–5.
8. Santen RJ, Boyd NF, Chlebowski RT, Cummings S, Cuzick J, Dowsett M, et al. Critical assessment of new risk factors for breast cancer: considerations for development of an improved risk prediction model. *Endocrine-Related Cancer.* 2007; 14:169–87. [PubMed: 17639036]
9. Niu N, Qin Y, Fridley BL, Hou J, Kalari KR, Zhu M, et al. Radiation pharmacogenomics: a genome-wide association approach to identify radiation response biomarkers using human lymphoblastoid cell lines. *Genome Res.* 2010; 20:1482–92. [PubMed: 20923822]
10. Li L, Fridley B, Kalari K, Jenkins G, Batzler A, Safgren S, et al. Gemcitabine and cytosine arabinoside cytotoxicity: association with lymphoblastoid cell expression. *Cancer Res.* 2008; 68:7050–8. [PubMed: 18757419]

11. Pei H, Li L, Fridley BL, Jenkins GD, Kalari KR, Lingle W, et al. FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt. *Cancer Cell*. 2009; 16:259–66. [PubMed: 19732725]
12. Ingle JN, Schaid DJ, Goss PE, Liu M, Mushiroda T, Chapman JA, et al. Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. *J Clin Oncol*. 2010; 28:4674–82. [PubMed: 20876420]
13. Liu M, Wang L, Bongartz T, Hawse J, Markovic S, Schaid D, et al. Aromatase inhibitors, estrogens and musculoskeletal pain: estrogen-dependent T-cell leukemia 1A (TCL1A) gene-mediated regulation of cytokine expression. *Breast Cancer Res*. 2012; 14:R41. [PubMed: 22405131]
14. Liu M, Ingle JN, Fridley BL, Buzdar AU, Robson ME, Kubo M, et al. TSPYL5 SNPs: association with plasma estradiol concentrations and aromatase expression. *Mol Endocrinol*. 2013 Mar 21. (Epub ahead of print).
15. Antoniou A, Pharoah PDP, Narod SA, Risch HA, Risch HA, Eyfjord JE, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet*. 2003; 72:1117–30. [PubMed: 12677558]
16. Chen S, Parmigiani G. Meta-analysis of *BRCA1* and *BRCA2* penetrance. *J Clin Oncol*. 2007; 25:1329–33. [PubMed: 17416853]
17. Marks JR, Huper G, Vaughn JP, Davis PL, Norris J, McDonnell DP, et al. BRCA1 expression is not directly responsive to estrogen. *Oncogene*. 1997; 14:115–21. [PubMed: 9010238]
18. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999; 55:997–1004. [PubMed: 11315092]
19. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006; 2:e190. [PubMed: 17194218]
20. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38:904–9. [PubMed: 16862161]
21. Tian C, Gregersen PK, Seldin MF. Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet*. 2008; 17:R143–R50. [PubMed: 18852203]
22. Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, Selmi C, et al. Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet*. 2008; 4:e4. [PubMed: 18208329]
23. Yu K, Wang Z, Li Q, Wacholder S, Hunter DJ, Hoover RN, et al. Population substructure and control selection in genome-wide association studies. *PLoS One*. 2008; 3:e2551. [PubMed: 18596976]
24. The International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005; 437:1299–320. [PubMed: 16255080]
25. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65. [PubMed: 23128226]
26. Goetz M, Knox S, Suman V, Rae J, Safgren S, Ames M, et al. The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res Treat*. 2007; 101:113–21. [PubMed: 17115111]
27. Goetz MP, Suman VJ, Hoskin TL, Gnant M, Filipits M, Safgren SL, et al. CYP2D6 metabolism and patient outcome in the Austrian Breast and Colorectal Cancer Study Group Trial (ABCSG) 8. *Clin Cancer Res*. 2013; 19:500–7. [PubMed: 23213055]
28. Goetz MP, Schaid DJ, Wickerham DL, Safgren S, Mushiroda T, Kubo M, et al. Evaluation of CYP2D6 and efficacy of tamoxifen and raloxifene in women treated for breast cancer chemoprevention: results from the NSABP P1 and P2 clinical trials. *Clin Cancer Res*. 2011; 17:6944–51. [PubMed: 21880792]
29. Narod SA, Salmena L. BRCA1 and BRCA2 mutations and breast cancer. *Discov Med*. 2011; 12:445–53. [PubMed: 22127115]

30. Rice JC, Ozelik H, Maxeiner P, Andrulis I, Futscher BW. Methylation of the BRCA1 promoter is associated with decreased BRCA1 mRNA levels in clinical breast cancer specimens. *Carcinogenesis*. 2000; 21:1761–5. [PubMed: 10964110]
31. Hata A, Seoane J, Lagna G, Montalvo E, Hemmati-Brivanlou A, Massagué J. OAZ uses distinct DNA- and protein-binding zinc fingers in separate BMP-Smad and Olf signaling pathways. *Cell*. 2000; 100:229–40. [PubMed: 10660046]
32. Starita LM, Parvin JD. The multiple nuclear functions of BRCA1: transcription, ubiquitination and DNA repair. *Curr Opin Cell Biol*. 2003; 15:345–50. [PubMed: 12787778]
33. Turner J, Crossley M. Mammalian Krüppel-like transcription factors: more than just a pretty finger. *Trends Biochem Sci*. 1999; 24:236–40. [PubMed: 10366853]
34. Santamaria I, Pendas AM, Velasco G, López Otín C. Genomic structure and chromosomal localization of the human cathepsin O gene (CTSO). *Genomics*. 1998; 53:231–4. [PubMed: 9790772]
35. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–75. [PubMed: 17701901]

Statement of Significance

This study identified novel, functionally polymorphic genes involved in the estrogen-dependent regulation of BRCA1 expression, as well as a novel mechanism for genetic variation in SERM therapeutic effect. These observations, and definition of their underlying mechanisms, represent steps toward pharmacogenomically individualized SERM breast cancer prevention.

Figure 1.

Panel A. Manhattan plot of p-values for conditional logistic regression adjusted for nine eigenvectors. (Black: p-value $\geq 1E-4$, Blue: p-value $< 1E-4$ to $1E-5$, Red: p-value $< 1E-5$). Panel B. Chromosome 16 region of interest: Manhattan plot of $-\log_{10}$ p-values from conditional logistic regression adjusted for nine eigenvectors for Platform SNPs (genotyped using Illumina Human610-Quad BeadChip) (blue circles), Fine mapped SNPs (imputed using HapMap2 and then genotyped) (red circles), HapMap2 imputed SNPs (green circles) and 1K Imputed SNPs (imputed using 1000 Genomes database) (black circles). Panel C. Chromosome 4 region of interest: Manhattan plot of $-\log_{10}$ p-values from conditional logistic regression adjusted for nine eigenvectors for Platform SNPs (genotyped using Illumina Human610-Quad BeadChip) (blue circles), Fine mapped SNPs (imputed using HapMap2 and then genotyped) (red circles), HapMap2 imputed SNPs (green circles) and 1K Imputed SNPs (imputed using 1000 Genomes database) (black circles).

Figure 2.

Panels A and B. mRNA expression relative to ER α (in %) for ZNF423 (panel A) and BRCA1 (panel B) for lymphoblastoid cell lines (LCLs) of known genotypes for variant (V) and wild type (WT) chromosome 16 *ZNF423* SNPs (3 cell lines each) incubated for 24 h with estradiol (E2) concentrations that varied from 0.00001 to 0.01 nM. Error bars represent standard error of the mean (SEM). Panel C. ZNF423 mRNA expression relative to baseline in Hs578T cells after ZNF423 overexpression (OE) (upper left graph) or ZNF423 knockdown (KD) (upper right graph) together with corresponding BRCA1 transcriptional activity measured by luciferase activity at 48 h (lower graphs). Error bars represent SEM. Panel D. Chromatin immunoprecipitation assay using human Hs578T cells stably transfected with ER α showing binding of ZNF423 to a region containing a ZNF423 consensus binding sequence located 1735 to 1518 bp from the ATG in *BRCA1*.

Figure 3.

Panels A and B. mRNA expression for ZNF423 (panel A) and BRCA1 (panel B) for lymphoblastoid cell lines with WT/WT (8 cell lines), WT/V (7 cell lines), and V/V (8 cell lines) genotypes for the chromosome 16 *ZNF423* SNPs after exposure to estradiol (E2) alone or E2 with increasing concentrations of 4-hydroxytamoxifen (4-OH-TAM). Error bars represent standard error of the mean (SEM). Panel C. Schematic depiction of the *ZNF423* intron 2 area near the rs9940645 SNP. The locations of estrogen response elements (EREs) are shown as boxes, and arrows show the locations of primers used to perform the ChIP amplifications. Panel D. Chromatin immunoprecipitation assay for the area of *ZNF423* that contains the rs9940645 SNP. WT, wild type; V, variant; E2, estradiol; 4-OH TAM, 4-

hydroxytamoxifen. Panel E. Dual luciferase reporter-gene assays showing the effect of 4-hydroxytamoxifen (4-OH-TAM) (left graph) and raloxifene (right graph). A 500 bp DNA sequence that included either WT or variant SNP sequence for rs9940645 and all of the EREs shown in Figure 4C was cloned into the pGL3 basic reporter plasmid. A 1500 bp region of the *ZNF423* promoter was then cloned downstream of the 500 bp segment, and these constructs were used to transfect IGROV-1 cells that were treated with 0.1 nM E2, with or without 10^{-7} μ M 4-hydroxytamoxifen (left graph) or 10^{-7} M raloxifene (right graph). Error bars represent SEM.

Figure 4.

Panel A. Comet assays for knockdown (KD) of ZNF423 with or without BRCA1 overexpression (OE) and for BRCA1 KD. Panel B. CTSO mRNA expression relative to ERα (in %) in lymphoblastoid cell lines (LCLs) of known genotypes for variant (V) and wild type (WT) chromosome 4 SNPs (3 cell lines each) incubated for 24 hours at varying estradiol (E2) concentrations from 0.0001 to 0.01 nM. Error bars represent standard error of the mean (SEM). Panel C. BRCA1 mRNA expression relative to ERα (in %) in the same LCLs shown in panel B. Error bars represent SEM. Panel D. Comet assays for KD of CTSO with or without BRCA1 OE and BRCA1 KD.

Table 1
SNPs With Lowest P-Values Identified During GWAS Genotyping: NSABP P-1 and P-2

SNP	Chrom	SNP Position (bp)	Gene Name (±50kb)	Gene Position (bp)	Minor Allele		Conditional Logistic Regression for Matched Design				Unconditional Logistic Regression	
							Adjusted for 9 eigenvectors	OR	OR 95% CI	Unadjusted		
					Frequency							
						Cases	Controls		Lower	Upper	p-value	p-value
rs060157	16	49830204	ZNF423	49521435–49891830		0.39	0.47	0.7	0.6	0.81	2.12E-06	2.86E-06
rs0510351	13	23384332	N/A	N/A		0.12	0.19	0.61	0.5	0.75	2.76E-06	5.11E-06
rs0030044	4	157011923	CTSO	156845270–156875063		0.45	0.36	1.42	1.23	1.65	3.63E-06	3.16E-06
rs11076499	16	49828405	ZNF423	49521435–49891830		0.39	0.47	0.71	0.61	0.82	3.70E-06	4.46E-06
rs11999029	9	11860446	N/A	N/A		0.16	0.11	1.64	1.33	2.02	3.76E-06	7.20E-06
rs44256192	4	157016504	CTSO	156845270–156875063		0.39	0.31	1.44	1.23	1.68	4.27E-06	3.10E-06
rs11925220	3	133569777	RAB6B	133543083–133614680		0.54	0.46	1.39	1.21	1.61	6.17E-06	9.63E-06
rs7853211	9	11830974	N/A	N/A		0.19	0.13	1.55	1.28	1.88	6.94E-06	8.98E-06
rs10960317	9	11848457	N/A	N/A		0.16	0.11	1.61	1.31	1.99	7.01E-06	1.34E-05
rs6928160	9	11830879	N/A	N/A		0.17	0.12	1.59	1.3	1.94	7.28E-06	1.10E-05
rs7832497	8	104126852	AK001351	104133260–104147519		0.2	0.27	0.68	0.57	0.8	9.98E-06	1.84E-05

Abbreviations: SNP: single nucleotide polymorphism; NSABP: National Surgical Adjuvant Breast and Bowel Project; Chrom: chromosome; bp: base position; CI: confidence interval.

ZNF423 and CTSO SNP Joint Analysis on Breast Cancer Risk: NSABP P-1 and P-2

Table 2

CTSO (rs10030044)				
ZNF423 (rs8060157)				
mm, OR (95% CI) [cases:controls]	MM, OR (95% CI) [cases:controls]	Mm, OR (95% CI) [cases:controls]	mm, OR (95% CI) [cases:controls]	
Mm, OR (95% CI) [cases:controls]				
MM, OR (95% CI) [cases:controls]				

SNP: single nucleotide polymorphism; NSABP: National Surgical Adjuvant Breast and Bowel Project

OR: odds ratio

M = major allele, m, is the risk allele for ZNF423 with an OR of 0.70.

m = minor allele, m, is the risk allele for CTSO with an OR of 1.40.