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

INTRODUCTION

Worldwide, breast cancer is the most frequently diagnosed cancer and most frequent cause of cancer-related death in women (1). In high-risk women, breast cancer can be prevented by treatment with selective estrogen receptor modulators (SERM), 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 more than 33,000 women (2–4) and were the basis for U.S. Food and Drug Administration approval of both drugs for breast cancer prevention. However, despite 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 5 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 conducted in an attempt to apply genome-wide techniques to identify single-nucleotide polymorphisms (SNP) 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 worldwide 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) SNP genotypes, 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 (LCL) for which we had generated dense SNP and mRNA expression data—a model system that has already shown its power both for 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 ERs with the BRCA1 promoter, and the mechanisms underlying the estrogen-de- pendent 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 vari- ant 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 the performance of a discovery GWAS that involved 592 cases (i.e., participants who developed breast cancer while on SERM therapy) and 1,171 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, the 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, λ, was 1.021, indicating little influence of population stratification (18). To avoid local genomic linkage disequilibrium (LD; refs. 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 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 the 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 3 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 the criteria for genome-wide significance, because we had used the largest sample set available, as well as the majority of samples available worldwide, to conduct 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, whereas 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 C, 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 significa- ntly 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 (OR 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 sub- sequent genotyping of imputed SNPs in this region of chromo- some 16 identified 8 additional SNPs (rs6500258, rs9940645, rs7499405, rs60841334, seq5958, rs1861343, rs5816658, and rs12446233) with low P values (P = 2.12E-06–4.22E-06), all of which mapped to the second intron of ZNF423 (Supplementary Table S4D). All of these SNPs were in tight LD (Pearson correlation r2 > 0.9), with minor allele frequencies (MAF) that varied from 0.39 to 0.41.

As 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 nmol/L E2 and observed that ZNF423 mRNA expression was induced 75- and 13-fold after 32 and 36 hours in the 2 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.
ZNF423 SNPs and BRCA1 Expression in SERM Prevention

Figure 1. A, Manhattan plot of $P$ values for conditional logistic regression adjusted for 9 eigenvectors. Black: $P \geq 1E^{-4}$, blue: $P < 1E^{-4}$ to $1E^{-5}$, red: $P < 1E^{-5}$. B, chromosome 16 region of interest: Manhattan plot of $-\log_{10}(P$ values) from conditional logistic regression adjusted for 9 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 1kG imputed SNPs (imputed using 1000 Genomes database; black circles). C, chromosome 4 region of interest: Manhattan plot of $-\log_{10}(P$ values) from conditional logistic regression adjusted for 9 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 1kG imputed SNPs (imputed using 1000 Genomes database; black circles).
Table 1. SNPs with lowest p values identified during GWAS genotyping: NSABP P-1 and P-2

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>SNP position (bp)</th>
<th>Gene name (±50 kb)</th>
<th>Gene position (bp)</th>
<th>MAF</th>
<th>Cases</th>
<th>Controls</th>
<th>OR (95% CI)</th>
<th>Adjusted for 9 eigenvectors</th>
<th>Unadjusted</th>
<th>Adjusted for 9 eigenvectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs8060157</td>
<td>16</td>
<td>49830204</td>
<td>ZNF423</td>
<td>49521435-49891830</td>
<td>0.39</td>
<td>0.47</td>
<td></td>
<td>0.7 (0.6–0.81)</td>
<td>2.12E–06</td>
<td>2.86E–06</td>
<td>1.11E–06</td>
</tr>
<tr>
<td>rs9510351</td>
<td>13</td>
<td>23384332</td>
<td>N/A</td>
<td>N/A</td>
<td>0.12</td>
<td>0.19</td>
<td></td>
<td>0.61 (0.5–0.75)</td>
<td>2.76E–06</td>
<td>5.11E–06</td>
<td>4.52E–06</td>
</tr>
<tr>
<td>rs10030044</td>
<td>4</td>
<td>15701923</td>
<td>CTSO</td>
<td>156845270-156875063</td>
<td>0.45</td>
<td>0.36</td>
<td></td>
<td>1.42 (1.23–1.65)</td>
<td>3.63E–06</td>
<td>3.16E–06</td>
<td>8.49E–07</td>
</tr>
<tr>
<td>rs11076499</td>
<td>16</td>
<td>49828405</td>
<td>ZNF423</td>
<td>49521435-49891830</td>
<td>0.39</td>
<td>0.47</td>
<td></td>
<td>0.71 (0.61–0.82)</td>
<td>3.70E–06</td>
<td>4.46E–06</td>
<td>1.87E–06</td>
</tr>
<tr>
<td>rs11999029</td>
<td>9</td>
<td>11860446</td>
<td>N/A</td>
<td>N/A</td>
<td>0.16</td>
<td>0.11</td>
<td></td>
<td>1.64 (1.33–2.02)</td>
<td>3.76E–06</td>
<td>7.20E–06</td>
<td>5.34E–06</td>
</tr>
<tr>
<td>rs256192</td>
<td>4</td>
<td>157016504</td>
<td>CTSO</td>
<td>156845270-156875063</td>
<td>0.39</td>
<td>0.31</td>
<td></td>
<td>1.44 (1.23–1.68)</td>
<td>4.27E–06</td>
<td>3.10E–06</td>
<td>2.46E–06</td>
</tr>
<tr>
<td>rs11925220</td>
<td>3</td>
<td>133569777</td>
<td>RAB6B</td>
<td>133543083-133614680</td>
<td>0.54</td>
<td>0.46</td>
<td></td>
<td>1.39 (1.21–1.61)</td>
<td>6.17E–06</td>
<td>9.63E–06</td>
<td>9.64E–06</td>
</tr>
<tr>
<td>rs7853211</td>
<td>9</td>
<td>11830974</td>
<td>N/A</td>
<td>N/A</td>
<td>0.19</td>
<td>0.13</td>
<td></td>
<td>1.55 (1.28–1.88)</td>
<td>6.94E–06</td>
<td>8.98E–06</td>
<td>6.49E–06</td>
</tr>
<tr>
<td>rs10960317</td>
<td>9</td>
<td>11848457</td>
<td>N/A</td>
<td>N/A</td>
<td>0.16</td>
<td>0.11</td>
<td></td>
<td>1.61 (1.31–1.99)</td>
<td>7.01E–06</td>
<td>1.34E–05</td>
<td>9.69E–06</td>
</tr>
<tr>
<td>rs16928160</td>
<td>9</td>
<td>11830879</td>
<td>N/A</td>
<td>N/A</td>
<td>0.17</td>
<td>0.12</td>
<td></td>
<td>1.59 (1.3–1.94)</td>
<td>7.28E–06</td>
<td>1.10E–05</td>
<td>7.33E–06</td>
</tr>
<tr>
<td>rs7832497</td>
<td>8</td>
<td>104126852</td>
<td>AK001351</td>
<td>104133260-104147519</td>
<td>0.2</td>
<td>0.27</td>
<td></td>
<td>0.68 (0.57–0.8)</td>
<td>9.98E–06</td>
<td>1.84E–05</td>
<td>2.52E–05</td>
</tr>
</tbody>
</table>

Abbreviations: bp, base position; CI, confidence interval.
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 carry out those experiments, we stably transfected LCLs homozygous for either WT or variant ZNF423 SNP genotypes with ERα, 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, whereas 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.

**Table 2. ZNF423 and CTSO SNP joint analysis on breast cancer risk: NSABP P-1 and P-2**

<table>
<thead>
<tr>
<th>ZNF423 (rs8060157)</th>
<th>CTSO (rs10030044)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM, OR (95% CI)</td>
<td>MM, OR (95% CI)</td>
</tr>
<tr>
<td>[cases:controls]</td>
<td>[cases:controls]</td>
</tr>
<tr>
<td>mm, OR (95% CI)</td>
<td>mm, OR (95% CI)</td>
</tr>
<tr>
<td>[cases:controls]</td>
<td>[cases:controls]</td>
</tr>
<tr>
<td>Mm, OR (95% CI)</td>
<td>Mm, OR (95% CI)</td>
</tr>
<tr>
<td>[cases:controls]</td>
<td>[cases:controls]</td>
</tr>
<tr>
<td>MM, OR (95% CI)</td>
<td>MM, OR (95% CI)</td>
</tr>
<tr>
<td>[cases:controls]</td>
<td>[cases:controls]</td>
</tr>
</tbody>
</table>

NOTE: 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. Abbreviation: CI, confidence interval.

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 carry out those experiments, we stably transfected LCLs homozygous for either WT or variant ZNF423 SNP genotypes with ERα, 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, whereas 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 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 carry out those experiments, we stably transfected LCLs homozygous for either WT or variant ZNF423 SNP genotypes with ERα, 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, whereas 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.

**Figure 2.** A and B, mRNA expression relative to ERα (mRNA) for ZNF423 (A) and BRCA1 (B) for LCLs of known genotypes for variant (V) and WT chromosome 16 ZNF423 SNPs (3 cell lines each), incubated for 24 hours with E2 concentrations that varied from 0.00001 to 0.01 nmol/L. Error bars represent SEM. C, ZNF423 mRNA expression relative to baseline in Hs578T cells after ZNF423 overexpression (OE; top left) or ZNF423 knockdown (KD; top right) together with corresponding BRCA1 transcriptional activity measured by luciferase activity at 48 hours (bottom). Error bars represent SEM. D, ChIP assay using human Hs578T cells stably transfected with ERα showing binding of ZNF423 to a region containing a ZNF423 consensus binding sequence located 1,735 to 1,518 bp from the ATG in BRCA1.
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 more than a decade that BRCA1 is estrogen inducible, although the mechanism has remained unclear (17). Because the phenotype for our study was breast cancer risk during 5 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 conducted quantitative real-time PCR (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 2,000 bp of the BRCA1 5′-flanking region (5′-FR) into a pGL3-Basic reporter plasmid (Promega). This area of BRCA1 has been reported to include the promoter for the gene (30). After knockdown of ZNF423 to 24% of baseline in Hs578T-ERα breast cancer cells (Fig. 2C, top 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, top right), BRCA1 5′-FR reporter gene transcriptional activity increased 3.6-fold (Fig. 2C, bottom 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 conducting 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 conducted ChIP assays after 24-hour exposure of the cells to 0.1 nmol/L E2 to induce the expression of ZNF423. We found that a 217 bp PCR ampli
con containing one of the putative ZNF423 binding motifs located 1,735 to 1,518 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 showed that estrogen exposure could induce ZNF423 expression and that ZNF423, in turn, could induce the expression of BRCA1.

However, as the estrogen-dependent induction of ZNF423 occurred in cells with WT but not variant SNP genotypes (Fig. 2A), these observations seemed to be at odds with the clinical GWAS data that 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 induc-

tion, 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 nmol/L 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, whereas, in contrast, expression in cells with WT genotypes decreased to levels at or below baseline (Fig. 3A and B). Of importance, similar experiments carried out 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 observation that variant ZNF423 SNP genotypes were associated with decreased breast cancer risk during 5 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 ERα Binding

To determine whether the SNPs in intron 2 of ZNF423 might alter ERα binding to estrogen response elements (ERE) present in intron 2 of the ZNF423 gene, we conducted ChIP assays with ERα antibody for all 5 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
Figure 3. A and B, mRNA expression for ZNF423 (A) and BRCA1 (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 E2 alone or E2 with increasing concentrations of 4-hydroxytamoxifen (4-OH-TAM). Error bars represent SEM. C, schematic depiction of the ZNF423 intron 2 area near the rs9940645 SNP. The locations of EREs are shown as boxes, and arrows show the locations of primers used to conduct the ChIP amplifications. D, ChIP assay for the area of ZNF423 that contains the rs9940645 SNP. V, variant. E, dual luciferase reporter-gene assays showing the effect of 4-OH-TAM (left) and raloxifene (right). A 500 bp DNA sequence that included either a WT or variant SNP sequence for rs9940645 and all of the EREs shown in Fig. 4C was cloned into the pGL3 basic reporter plasmid. A 1,500 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 0.01 nmol/L E2, with or without 10^{-7} mol/L 4-OH-TAM (left) or 10^{-7} mol/L raloxifene (right). Error bars represent SEM.
conducted 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 1,500 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 0.01 nmol/L 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 0.01 nmol/L E2 and 10⁻⁷ mol/L 4-hydroxytamoxifen was added, the activity of the variant SNP construct increased while the WT SNP sequence 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 mol/L raloxifene rather than 4-hydroxytamoxifen (Fig. 3E, right). 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 E) and ERα 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 conduct comet assays after the knockdown of ZNF423 and/or BRCA1 (Fig. 4A). As anticipated, there was evidence of DNA double-strand breaks after the knockdown of BRCA1, but the same was true after the knockdown of endogenous ZNF423. However, BRCA1 overexpression could overcome the effect of ZNF423 knockdown 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 CTOS 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 and Table 1). The most significant chromosome 4 SNP genotyped on the Illumina GWAS platform was rs10030044 (P = 2.76E-06) but, in this case, variant SNP alleles (MAF = 0.45) were associated with increased risk for developing breast cancer, with an OR of 1.42. Imputation revealed 2 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 CCCTC-binding factor (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 carrying out a series of experiments similar to those that we conducted with ZNF423, starting with a determination of whether CTSO might, like ZNF423, be estrogen-inducible and, if so, whether it might also be related to BRCA1 expression.

CTSO mRNA expression was induced by exposure to increasing concentrations of E2 in LCLs stably transfected with ERα (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 no induction of BRCA1 expression in these cell lines (Fig. 4B). However, cells containing the WT SNP alleles showed 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 (Fig. 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 (Supplementary 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 ERα 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 knockdown 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

As 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 ORs for these 2 sets of SNPs are listed in Table 2, which shows a broad range of relative ORs for the development of breast cancer while on SERM therapy for 5 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 conducted in an attempt to identify genetic factors that might contribute to risk for the development of breast cancer during 5 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., \( P < 0.05 \)), they were near that threshold and, as a result, we felt it important to investigate these signals functionally and mechanistically, especially as 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 as 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); that is, the SNPs in \( ZNF423 \) were related to the E2-dependent induction of \( BRCA1 \). We also showed that \( ZNF423 \) overexpression and knockdown 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 shown by ChIP assay (Fig. 2C and D). 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 an SNP-dependent manner (Fig. 3A and B). 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 an SNP-dependent mechanism for variation in SERM effect, a mechanism involving SNPs near, but not within, an ERE. The fact that 2 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, predominately breast cancer cell lines, to carry out 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 shown its value in both the generation of pharmacogenomic hypotheses and the testing of hypotheses 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: (i) trial and treatment arm (P-1 tamoxifen, P-2 tamoxifen, P-2 raloxifene); (ii) age at 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); 5-year predicted breast cancer risk based on the Gail model (200.0, 2.01–3.00, 3.01–5.00, >5.01); (ii) history of lobular carcinoma in situ (yes vs. no); (iv) history of atypical hyperplasia in breast (yes vs. no); (v) 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 1,176 matched controls were identified by NSABP for inclusion in the GWAS. Six samples had insufficient DNA (2 cases and 4 controls), and an additional 3 patients (2 cases and 1 control) were excluded after a series of quality-control steps, as noted below.

**Genotype Quality Control**

Two cases and 2 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 4 of our study subjects who had duplicate samples and 38 replicate CEPH samples (child–parent trio). This resulted in a total of 1,808 samples.
genotyped for this study (592 cases, 1,174 controls, 4 duplicates, and 38 CEPH).

For the 4 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/ethnicity ambiguity for 3 participants. 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 race/ethnicity ambiguity. We excluded these 3 participants from the analyses, resulting in a final total of 592 cases and 1,171 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 2 controls, and 51 cases each matched to one control. SNP genotypes were coded as additive effects on the log OR 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 control for the matched design were implicitly controlled in the conditional logistic regression.

Although none of the 9 eigenvectors identified, as noted in the Results, differed significantly between cases and controls, we conducted 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 to include the 42 controls who had no matched cases and the 2 cases who 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 subgenomes that were imputed, 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 (Camden, NJ), and all other cell lines came from the American Type Culture Collection (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 passed.

**BRC1 Reporter Plasmids and Transcriptional Activity Assays.** U2OS-ERα genomic DNA was used as a template for PCR reactions conducted with the following 2 pairs of primers: BRC1 forward, 5′-CTCGAGATGCTTGGATACTCACCTGCGGACA-3′; BRC1 reverse, 5′-AGATCTGAGTGAATCTAACATGGCGGACA-3′. The amplicons were cloned directly into the luciferase reporter plasmid pGL3 basic (Promega) after digestion with Nhe I and Bgl II to create a pGL3-BRCA1 construct. The sequences of the inserted amplicons were verified by DNA sequencing.

Hs5787-ERα cells were then seeded in triplicate in 12-well cell culture plates at a concentration of 10^5 cells per well. After 24 hours, the cells were transfected using Lipofectamine 2000 (Invitrogen) with 2 μg 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 conducted 48 hours 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, 7 cell lines with heterogeneous genotypes, and 8 cell lines homozygous for WT sequences at those SNPs, all stably transfected with ERα as described previously, were used in these studies. These stably transfected LCLs were cultured in RPMI media containing 15% (v/v) FBS with 200 μg/ml Zeocin. Before E2 treatment, 2 × 10^6 cells from each cell line were cultured for 24 hours in RPMI media containing 5% (v/v) charcoal-stripped FBS with 200 μg/ml Zeocin, followed by culture in the same medium without FBS for another 24 hours. All cells were then cultured for 24 hours in 6-well plates, with RPMI-1640 media that contained 0, 0.00001, 0.0001, 0.001, 0.01, 0.1, and 10 nmol/L E2. 4-Hydroxytamoxifen (Sigma-Aldrich) or raloxifene (Sigma-Aldrich) were then added to the same medium containing 0.01 nmol/L E2 at final 4-hydroxytamoxifen concentrations of 10⁻⁸, 10⁻⁷, 10⁻⁶, and 10⁻⁵ mol/L, or raloxifene concentrations 10⁻⁶ to 10⁻⁴ mol/L and the cells were cultured for an additional 24 hours. Total RNA was isolated from the cells with the RNeasy mini kit (Qiagen). Two hundred ng of total RNA was then used to conduct qRT-PCR with ZNF423, BRC1, and ERα primers. ZNF423 and BRC1 expression levels were normalized on the basis of ERα expression in each cell line.

**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.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Authors’ Contributions**

**Conception and design:** J.N. Ingle, D.L. Wickerham, D.J. Schaid, M. Kube, J.P. Costantino, S. Paik, M.P. Goetz, D.A. Flockhart, N. Wolmark, R.M. Weinshilboum

**Development of methodology:** M. Liu, J.P. Costantino, R.M. Weinshilboum

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** J.N. Ingle, D.L. Wickerham,
T. Mushiroda, M. Kubo, J.P. Costantino, V.G. Vogel, S. Paik, M.M. Ames, N. Wolmark, Y. Nakamura


Writing, review, and/or revision of the manuscript: J.N. Ingle, M. Liu, D.L. Wickerham, D.J. Schaid, L. Wang, J.P. Costantino, V.G. Vogel, S. Paik, M.P. Goetz, G.D. Jenkins, D.A. Flockhart, N. Wolmark, R.M. Weinsilboum

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J.P. Costantino, N. Wolmark, R.M. Weinsilboum

Study supervision: J.N. Ingle, N. Wolmark, Y. Nakamura, R.M. Weinsilboum

Acknowledgments

The authors thank 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.

Grant Support

These studies were supported in part by NIH grants U19 GM61388 (to J.N. Ingle, D.J. Schaid, L. Wang, G.D. Jenkins, A. Batzler, E.E. Carlson, and R.M. Weinsilboum; The Pharmacogenomics Research Network), P50CA16201 (Mayo Clinic Breast Cancer Specialized Program of Research Excellence; to J.N. Ingle and M.P. Goetz), U10CA37377 (to N. Wolmark) and U10CA69974 (to J.P. Costantino), U24CA114732 (to N. Wolmark), U10GM63173 (to D.A. Flockhart), and the RIKEN Center for Genomic Medicine and the Biobank Japan Project funded by the Ministry of Education, Culture, Sports, Science, and Technology, Japan (to T. Mushiroda, M. Kubo, and Y. Nakamura).

Received January 30, 2013; revised April 2, 2013; accepted April 10, 2013; published OnlineFirst June 13, 2013.

REFERENCES


