Check for updates

Germline Genetic Variants in GATA3 and Breast Cancer Treatment Outcomes in SWOG S8897 Trial and the Pathways Study

Victoria Larsen,[1] William E. Barlow,[2] Jun J. Yang,[3] Qianqian Zhu,[4] Song Liu,[4]
Marilyn L. Kwan,[5] Isaac J. Ergas,[5] Janise M. Roh,[5] Laura F. Hutchins,[6]
Susan A. Kadlubar,[6] Kathy S. Albain,[7] James M. Rae,[8] I-Tien Yeh,[9]
Peter M. Ravdin,[9] Silvana Martino,[10] Alan P. Lyss,[11] C. Kent Osborne,[12]
Gabriel N. Hortobagyi,[13] Lawrence H. Kushi,[4] Daniel F. Hayes,[8]
Christine B. Ambrosone,[1] Song Yao[1]

Abstract

In an ancillary study to a completed clinical trial, we identified *GATA3* genotypes predictive of survival after adjuvant chemotherapy, particularly among those subsequently treated with tamoxifen. Although the replication effort in a second cohort proved to be futile owing to substantial differences between the discovery and the replication cohorts, mechanistic exploration of the identified variants supported their functional significance.

Introduction: GATA3 is a critical transcription factor in maintaining the differentiated state of luminal mammary epithelial cells. We sought to determine the prognostic and predictive roles of *GATA3* genotypes for breast cancer. **Patients and Methods:** Twelve single nucleotide polymorphisms (SNPs) were genotyped in 2 breast cancer cohorts, including the SWOG S8897 trial where patients were treated with adjuvant chemotherapy (CAF [cyclophosphamide, doxorubicin, 5-fluorouracil] vs. CMF [cyclophosphamide, methotrexate, 5-fluorouracil]) or untreated, and the observational Pathways Study. **Results:** In the S8897 trial, rs3802604 and rs568727 were associated with disease-free survival and overall survival in the treated group, regardless of chemotherapy regimen. The GG genotype of rs3802604 conferred poorer overall survival (adjusted hazard ratio, 2.45; 95% confidence interval, 1.48-4.05) and disease-free survival (adjusted hazard ratio, 1.95; 95% confidence interval, 1.27-2.99) compared with the AA genotype. Similar associations were found for rs568727. In contrast, no association with either SNP was found in the untreated group. Subgroup analyses indicated that these 2 SNPs more strongly influenced outcomes in the patients who also received tamoxifen. However, the associations in the subgroup with tamoxifen treatment were not replicated in the Pathways Study, possibly owing to substantial differences between the 2 patient cohorts, such as chemotherapy regimen and length of follow-up. Results from joint analyses across these 2 cohorts were marginally significant, driven by the results in S8897. Bioinformatic analyses support potential functional disruption of the *GATA3*

[1]Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY
[2]SWOG Statistical Center, Seattle, WA
[3]Pharmaceutical Science, St. Jude Children's Research Hospital, Memphis, TN
[4]Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY
[5]Division of Research, Kaiser Permanente Northern California, Oakland, CA
[6]University of Arkansas for Medical Sciences, Little Rock, AR
[7]Loyola University Chicago Stritch School of Medicine, Maywood, IL
[8]University of Michigan Comprehensive Cancer Center, Ann Arbor, MI
[9]University of Texas Health Science Center at San Antonio, San Antonio, TX

[10]The Angeles Clinic and Research Institute, Santa Monica, CA
[11]Heartland NCORP, Missouri Baptist Medical Center, St. Louis, MO
[12]Baylor College of Medicine, Houston, TX
[13]University of Texas M.D. Anderson Cancer Center, Houston, TX

Address for correspondence: Song Yao, PhD, Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263
E-mail contact: song.yao@roswellpark.org

SNPs in breast tissue. **Conclusions:** The present study provides some evidence for the predictive value of *GATA3* genotypes for breast cancer adjuvant therapies. Future replication studies in appropriate patient populations are warranted.

Introduction

GATA binding protein 3 (GATA3) is among the most important and well-studied transcription factors in breast cancer.[1-3] GATA3 and its downstream target FOXA1 are critical for maintaining the luminal differentiation status of mammary epithelial cells.[4,5] It has a strong and specific presence in estrogen receptor (ER)-positive luminal subtype tumors, which is gradually lost as tumor cells acquire a stem cell-like status. Moreover, *GATA3* is the third most commonly mutated gene in breast cancer (11%), following *TP53* and *PIK3CA*, with the mutations occurring mostly in luminal subtypes.[6]

Because GATA3 suppresses the epithelial to mesenchymal transition of tumor cells and thus inhibits metastatic dissemination, in part, via regulating microRNAs,[7] the expression of *GATA3* gene and protein has been investigated and shown to be prognostic in breast cancer, although the evidence has not been definitive. In an early genome-wide gene expression meta-analysis, *GATA3* was identified as one of the top genes wherein low expression was related to poor breast cancer outcomes.[8] Similar results were reported with immunohistochemistry expression of GATA3, with the effects dependent on ER status in some studies.[9-13] Several studies have also evaluated immunohistochemistry expression of GATA3 with treatment response, and found it predictive of favorable endocrine therapy outcomes.[14-16] In a neoadjuvant study, GATA3 negativity was an independent predictor for pathologic complete response,[17] whereas in the case of chemoresistant triple-negative breast cancer, a GATA3-containing luminal-like gene signature was associated with better prognosis.[18]

Despite mounting evidence of GATA3 as a critical transcription factor with potential prognostic and predictive value in breast cancer, few studies have investigated the inherited genetic variations with breast cancer outcomes. This becomes an even more warranted research question to address given the existing high-level evidence that 2 *GATA3* single nucleotide polymorphisms (SNPs), rs3824662 and rs3781093, have been identified as the top hits in genome-wide association studies (GWAS) of risk and relapse of childhood acute lymphoblastic leukemia (ALL)[19-21] and another SNP with risk of Hodgkin lymphoma.[22] It is also notable that GATA3 plays an important role in T-cell differentiation reminiscent of that in mammary epithelial cells.[23] Thus, using a cross-validating study design, we sought to investigate the prognostic and predictive roles of *GATA3* SNPs in 2 cohorts of patients with breast cancer.

Materials and Methods

Patient Populations

Data and DNA samples for this study were obtained from 2 populations of patients with breast cancer: a completed phase III cooperative group trial, SWOG S8897 (INT-0102), and a cohort of survivors of breast cancer in the Pathways Study. Study schemas of the 2 populations are shown in Supplemental Figure 1 (in the online version), and descriptive characteristics of the 2 patient populations are shown in Table 1.

As previously described, women diagnosed with T1-T3a node-negative invasive breast cancer were eligible for S8897.[24] Patients were initially assigned to high-risk (tumor size $\geq$ 2 cm, or hormonal receptor-negative), low-risk (tumor size $\leq$ 1 cm), or uncertain-risk (tumor size between 1 and 2 cm or hormonal receptor-positive) groups. The last group was subsequently classified as either high- or low-risk based on the S-phase fraction by flow cytometry. Patients in the high-risk group were randomly assigned to 1 of 2 adjuvant chemotherapy treatment arms to receive either an anthracycline-based regimen (CAF), which consisted of 6 cycles of oral cyclophosphamide, intravenous doxorubicin, and 5-fluorouracil, or the non-anthracycline regimen (CMF), which consisted of 6 cycles of oral cyclophosphamide, intravenous methotrexate, and 5-fluorouracil. After receiving chemotherapy, women in the treated group were randomly assigned to receive or to not receive tamoxifen endocrine therapy for 5 years. Women in the low-risk group were followed for observation only and did not receive either adjuvant chemotherapy or endocrine therapy. The trial enrolled a total of 3965 patients between 1989 and 1993. After a median of 10.8 years of follow-up, the study concluded with slightly better overall survival (OS) but not disease-free survival (DFS) in patients treated with CAF versus CMF.[24] For this ancillary study, genomic DNA was required. Because lymph node tissues were not collected from patients in the initial high-risk group in S8897, only those from the initially low-risk group (n = 490) and the initially uncertain-risk group (n = 1029) who were subsequently assigned to high- and low-risk groups for adjuvant therapy ("treated group," n = 516) or no adjuvant therapy ("untreated group," n = 513), respectively, were included.

The Pathways Study is an observational prospective cohort study of survivors of breast cancer established at Kaiser Permanente Northern California (KPNC) to investigate factors associated with recurrence and mortality.[25,26] Between 2006 and 2013, a total of 4505 eligible patients diagnosed with incident invasive breast cancer were enrolled typically within 2 months of diagnosis. Baseline in-person interviews were conducted to query demographic, health history, and extensive lifestyle and other epidemiologic information potentially related to breast cancer risk and prognosis; anthropometric measures were also taken. Blood and/or saliva samples were obtained shortly thereafter. Regular follow-ups on lifestyle and other factors were conducted via mailed or telephone questionnaires at 6, 24, and 72 months post-baseline, and health outcomes and comorbidities at 12, 24, 48, 72, and 96 months post-baseline. Complete clinical data including initial diagnosis and treatment

Table 1 Descriptive Characteristics of Patient Populations in S8897 Trial and Pathways Study

Characteristic	S8897 Treated Group n = 441 (%)	S8897 Untreated Group n = 799 (%)	Pathways Chemotherapy Group n = 515 (%)	Pathways No Chemotherapy Group n = 483 (%)
Age, y				
<40	66 (15)	70 (9)	93 (19)	21 (4)
40-49	165 (37)	225 (28)	211 (44)	125 (26)
50-59	105 (24)	211 (26)	118 (24)	127 (26)
60-69	76 (17)	199 (25)	74 (15)	121 (25)
≥70	29 (7)	94 (12)	19 (4)	89 (18)
Median (range)	49 (27-85)	54 (24-89)	47 (25-76)	56 (23-89)
Race				
White (non-Hispanic)	391 (89)	750 (93)	270 (52)	298 (62)
Black (non-Hispanic)	27 (6)	30 (4)	37 (7)	37 (8)
Hispanic	12 (3)	12 (2)	93 (18)	57 (12)
Other	11 (2)	7 (1)	115 (23)	91 (19)
Menopausal status				
Premenopausal	231 (52)	306 (38)	350 (68)	213 (44)
Postmenopausal	210 (48)	493 (62)	165 (32)	270 (56)
Overall survival event				
Yes	329 (75)	577 (72)	45 (9)	32 (7)
No	112 (25)	222 (28)	470 (91)	451 (93)
Disease-free survival event				
Yes	289 (66)	482 (60)	80 (16)	71 (15)
No	152 (34)	317 (40)	435 (84)	412 (85)

were obtained from the KPNC Cancer Registry and KPNC electronic clinical and administrative databases. For the replication cohort, we selected a sub-cohort of patients enrolled in Pathways who received adjuvant tamoxifen therapy as the replication cohort.

Written informed consent was obtained from all patients for the use of their tissues in future research studies. This study was approved by the Institutional Review Boards at Roswell Park Comprehensive Cancer Center and KPNC.

Biospecimens and Genotyping

In S8897, 3 5-micron formalin-fixed paraffin-embedded slides of normal lymph node tissue were used for extraction of genomic DNA. Adequate DNA for genotyping was obtained from 1253 eligible patients, including 446 in the treated group and 807 in the untreated group. An Illumina GoldenGate 384-plex custom panel was designed for genotyping, which included 12 tagSNPs selected from *GATA3* based on HapMap CEU data. Two SNPs shown to be significant in the previous GWAS for childhood ALL, rs3824662 and rs3781093, were forced in during the tagSNPs selection process. For genotyping quality control purposes, 2% duplicates were randomly selected and blinded in across genotyping plates, and an in-house trio set was also included in each plate. The overall successful genotyping call rate was 98.3%, with complete concordance between duplicate pairs. Thirteen samples with a call rate < 85% were removed from analysis, resulting in a final sample size of 441 in the treated group and 799 in the untreated group.

In the Pathways Study, genomic DNA was extracted from blood and/or saliva samples and genotyping of the same set of 12 GATA3

SNPs were conducted using the Agena Bioscience MassARRAY, with the same percent of blind duplicate samples and in-house trio samples included. The overall genotyping call rate was 98.4%, and 10 samples with a call rate < 85% were removed from analysis, resulting in a final sample size of 1000 patients.

None of the SNPs in either study had Mendelian error or violated Hardy Weinberg Equilibrium ($P > .001$). The DNA preparation work was performed at Roswell Park Data Bank and Biorepository resource, and genotyping was performed at Roswell Park Genomics Shared Resource.

Statistical Analysis

Analytical endpoints were OS, defined as the time from registration to death from any cause (censored at the last date of contact), and DFS, defined as the time from registration to disease recurrence, new second primary, or death. OS and DFS probabilities were plotted with Kaplan-Meier survival curves by genotypes of each SNP separately in the treated and untreated groups in S8897, with P-values derived from log-rank test. Bonferroni correction was used to adjust for multiple testing, giving a significance threshold of $P \leq .001$ after correcting for testing 12 SNPs, 2 endpoints in 2 subgroups (0.05/48). Using a co-dominant model of inheritance, Cox proportional hazards regression was used to estimate hazards ratios (HRs) and 95% confidence intervals (CIs) for each SNP in treated and untreated groups separately, before and after adjusting for covariates including age at diagnosis, race and ethnicity, and treatment arm in the treated group. An additive model of inheritance was used in Cox regression analysis to estimate HRs and accompanying 95% CIs per risk allele for each variant. As *GATA3* plays

an important role in regulating estrogen signaling in breast cancer, which may have an impact on treatment response to tamoxifen, the 12 SNPs were further analyzed among patients who received tamoxifen following adjuvant chemotherapy and those who did not. Similar approaches were used to analyze data from the Pathways Study cohort. All analyses were conducted in R program.

Bioinformatic Analysis

To explore the potential functional impact of GATA3 SNPs, bioinformatic analyses were conducted based on publicly available data resources using HaploReg v4.1.[27] The program comprehensively interrogates chromatin state (promoter histone marks, enhancer histone marks, DNAse sensitivity), protein binding,

Figure 1 Genotyped Single Nucleotide Polymorphisms in *GATA3* and Breast Cancer Outcomes in S8897 Trial. *P*-Values From Log-Rank Test of Single Nucleotide Polymorphisms With Breast Cancer Outcomes, Including DFS and OS, are log10-Transformed and Plotted Along With Linkage Disequilibrium Map. Analyses Were Stratified by the Treated and Untreated Group, and Results From Each Subgroup Analysis are Indicated by Colors and Shapes Shown Above. The Lower and Upper Dashed Lines Indicate Nominal Significance Threshold (.05) and That After Correction for Multiple Testing (.001), Respectively
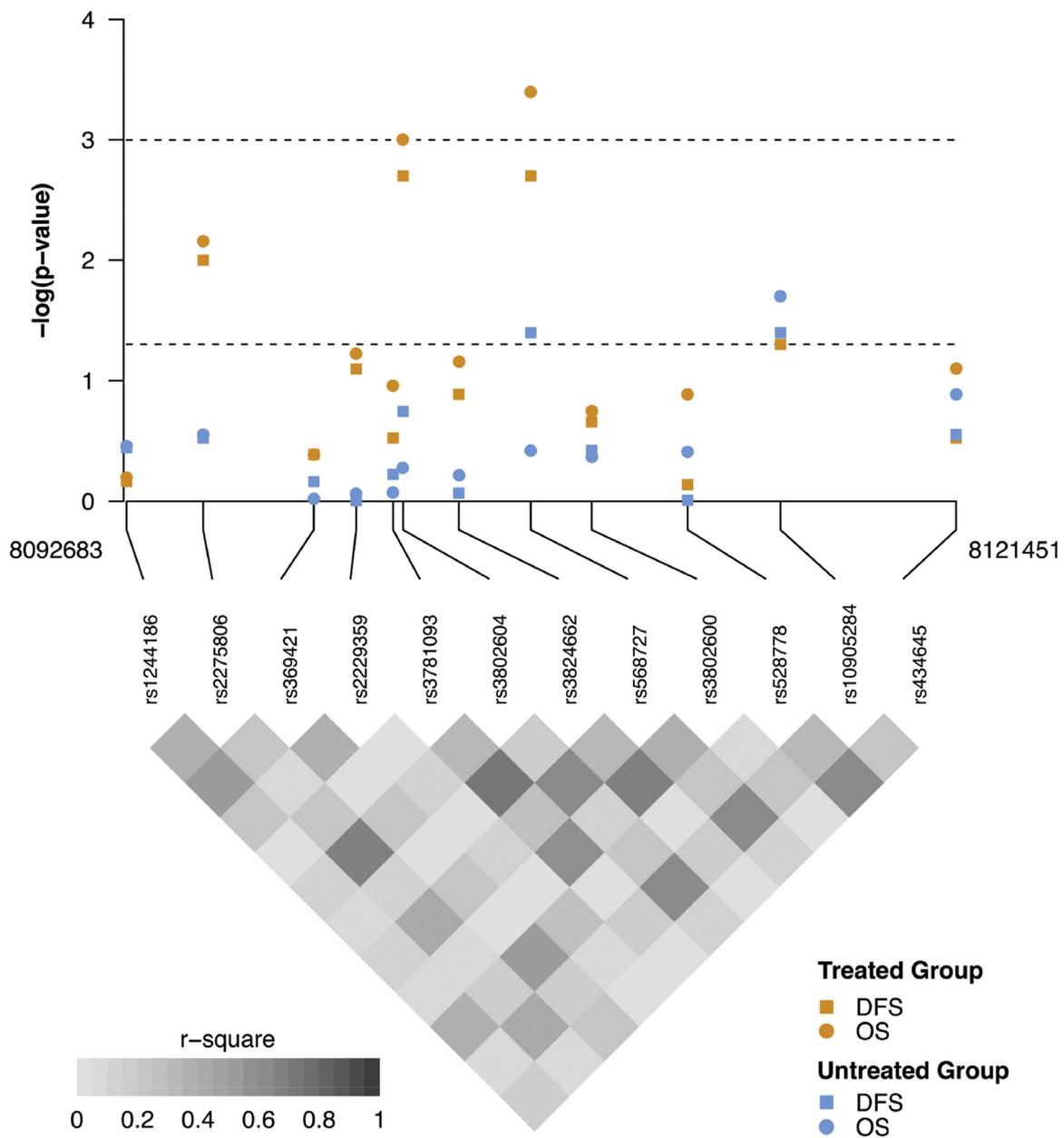
Abbreviations: DFS = Disease-free survival; OS = overall survival.

transcription factor motif, expression quantitative trait loci (eQTL), and evolutionary sequence conservation across mammals.

Results

GATA3 SNPs and Breast Cancer Outcomes in the Treated Versus Untreated Groups in S8897

Descriptive characteristics of S8897 patients with successful genotyping results are summarized in Table 1. Patients in the treated group were younger, with a median age of 49 years compared with patients in the untreated group with a median age of 54 years ($P < .05$) and were also more likely to be premenopausal ($P < .05$). A majority (92%) of patients self-reported as non-Hispanic Whites, 5% as non-Hispanic Blacks, 2% as Hispanics, and 1% as others.

Figure 1 depicts the results of log-rank tests for OS and DFS in the treated and untreated groups, along with the linkage disequilibrium (LD) map. Of the 12 *GATA3* SNPs tested, 4 were associated with both OS and DFS in the treated group with a nominal significance ($P < .05$). After correction for multiple testing, 2 SNPs, rs3802604 and rs568727 in relatively strong LD ($r^2 = 0.79$),

remained statistically significant (threshold $P = .001$) in the treated group, whereas the other 2 SNPs, rs568727 and rs10905284, became non-significant.

Kaplan-Meier curves of OS by the top 2 SNPs are shown in Figure 2, and curves of DFS are shown in Supplemental Figure 2 (in the online version). Among women treated with adjuvant therapy, those carrying the GG genotype of rs3802604 had poorer OS compared with the AA genotype (HR, 2.45; 95% CI, 1.48-4.05; $P = .0005$), in a dose-dependent manner (per G allele: HR, 1.57; 95% CI, 1.21-2.03; $P = .0007$) (Table 2). Similarly, those in the treated group carrying the AA genotype of rs568727 had poorer OS compared with the CC genotype (HR, 2.57; 95% CI, 1.43-4.30; $P = .0003$), also in a dose-dependent manner (per A allele: HR, 1.59; 95% CI, 1.23-2.06; $P = .0005$). Analyses of those 2 SNPs with DFS showed results consistent with OS, and the results were similar in subgroups defined by CAF versus CMF regimen. Neither of those 2 SNPs were associated with breast cancer outcomes in the untreated group of patients who did not receive adjuvant therapy (Table 2) (for complete results of all 12 SNPs, see Supplemental Tables 1 and 2 in the online version). Interaction tests of the 2

Figure 2 Kaplan-Meier Curves of Overall Survival by the Genotypes of rs3802604 and rs568727, the 2 Single Nucleotide Polymorphisms in GATA3 That Remained Significant After Correcting for Multiple Comparison, Are Plotted Separately for the Treated and Untreated Group in the S8897 Trial. *P*-Values are From Log-Rank Test
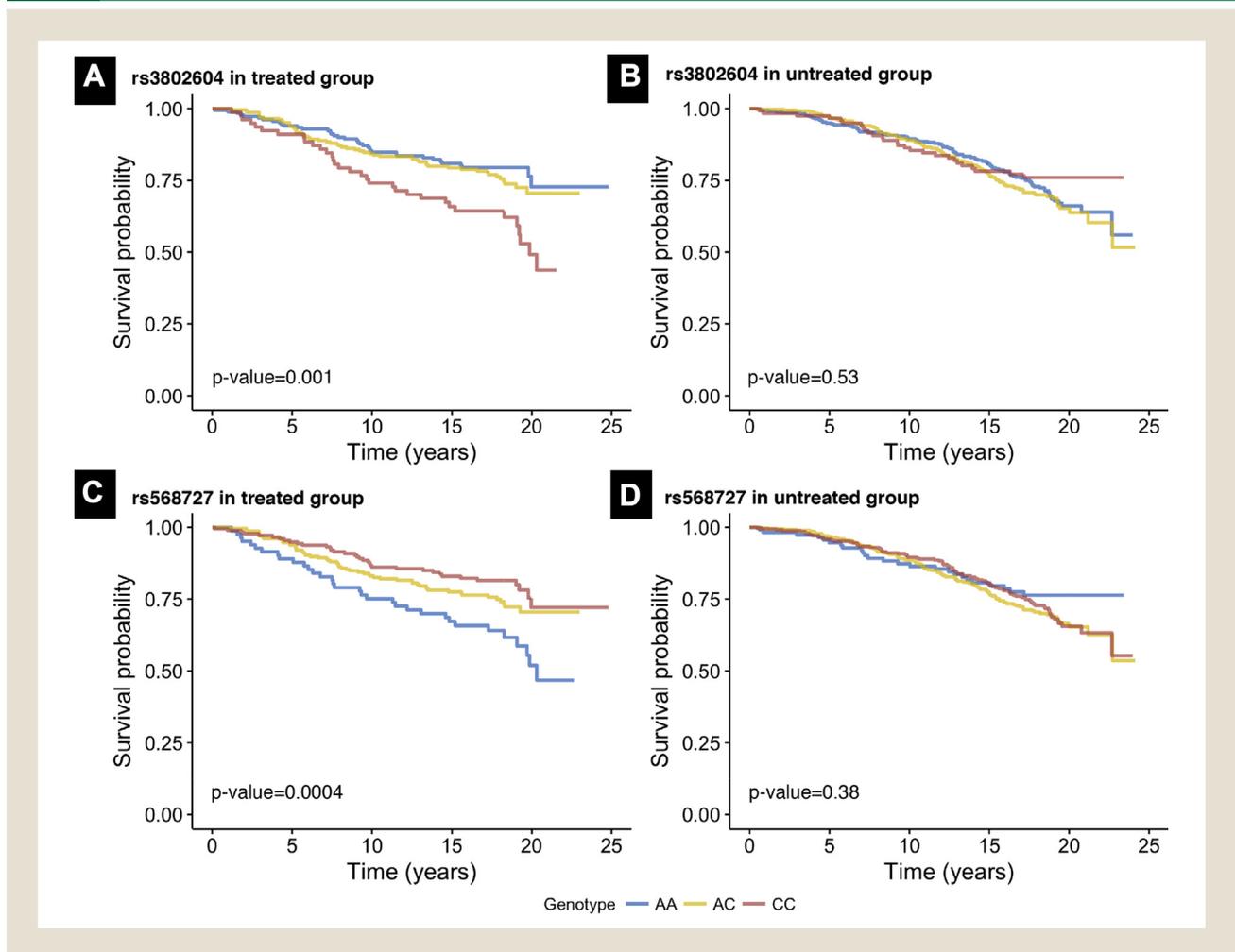
SNPs between the treated and untreated groups did not reach statistical significance ($P > .05$).

GATA3 SNPs and Breast Cancer Outcomes in the Treated Group by Tamoxifen Therapy in S8897

Because GATA3 is critical for the estrogen signaling pathways in breast cancer, the analyses in the treated group in S8897 were further stratified by whether or not patients received adjuvant tamoxifen therapy following chemotherapy. As shown in Table 3 for the top 2 SNPs, the associations of rs3802604 and rs568727 with both OS and DFS remained nominally significant among patients treated with tamoxifen, with the associations restricted to the homogenous variant genotypes, although the *P*-values did not meet the cutoff of significance level after correction for multiple comparison ($P > .001$). Among those not treated with tamoxifen, the associations, particularly with DFS, were weaker and non-significant after correction for multiple testing. Interaction tests of the 2 SNPs between the groups treated with and without tamoxifen did not reach statistical significance ($P > .05$).

In addition to the above 2 SNPs, 6 other GATA3 SNPs showed some evidence of association with OS and/or DFS among patients treated with tamoxifen, and one of them, rs369421, remained significant after correction for multiple testing ($P < .001$) (see Supplemental Tables 4 and 5 in the online version). Interestingly, the 2 GWAS SNPs previously associated with childhood ALL risk and relapse, rs3781093 and rs3824662, were associated with OS, and the latter also with DFS, at a nominal significance level among patients with no tamoxifen therapy. The directions of the associations were consistent with those from the ALL GWAS.

Replication of GATA3 SNPs With Breast Cancer Outcomes in the Pathways Study

Because the associations of GATA3 SNPs with breast cancer outcomes appeared to be stronger in patients treated with adjuvant chemotherapy followed by tamoxifen, consistent with its putative roles in estrogen signaling pathways in breast cancer, the replication analysis was conducted in the Pathways Study among 1000 patients with early-stage breast cancer who received adjuvant tamoxifen treatment. Descriptive characteristics of the replication cohort are summarized in Table 1. The median follow-up time was 6.3 years, with an OS rate of 7.7% and a DFS rate of 15.1%. In Cox regression models, none of the 12 SNPs, including rs3802604 and rs568727,

Table 2	GATA3 Single Nucleotide Polymorphisms and Breast Cancer Outcomes in S8897 Trial					
	Treated Group (n = 441)			**Untreated Group (n = 799)**		
Genotype	**No. Events/Patients**	**Adjusted HR (95% CI)**	**Adjusted P**	**No. Events/Patients**	**Adjusted HR (95% CI)**	**Adjusted P**
A. Overall survival						
rs3802604						
AA	31/164	1.00		96/337	1.00	
GA	51/206	1.36 (0.87-2.14)	.18	99/342	1.07 (0.81-1.42)	.64
GG	30/70	2.45 (1.48-4.05)	**.0005**	27/166	0.78 (0.51-1.20)	.27
Per risk allele		1.57 (1.21-2.03)	**.0007**		0.94 (0.78-1.13)	.48
rs568727						
CC	28/160	1.00		95/328	1.00	
AC	52/204	1.50 (0.94-2.39)	.09	102/357	1.03 (0.78-1.36)	.86
AA	32/77	2.57 (1.43-4.30)	**.0003**	24/111	0.74 (0.47-1.16)	.19
Per risk allele		1.59 (1.23-2.06)	**.0005**		0.91 (0.75-1.10)	.32
B. Disease-free survival						
rs3802604						
AA	49/164	1.00		134/337	1.00	
GA	66/206	1.08 (0.74-1.56)	.69	145/342	1.15 (0.91-1.45)	.25
GG	37/70	1.95 (1.27-2.99)	.002	38/116	0.80 (0.56-1.15)	.24
Per risk allele		1.38 (1.10-1.73)	.005		0.96 (0.82-1.12)	.62
rs568727						
CC	45/160	1.00		131/328	1.00	
AC	68/204	1.22 (0.83-1.79)	.31	153/357	1.14 (0.90-1.44)	.28
AA	39/77	2.02 (1.31-3.12)	**.001**	32/111	0.71 (0.48-1.04)	.08
Per risk allele		1.41 (1.13-1.76)	.003		0.92 (0.79-1.08)	.33

HRs, 95% CIs, and *P*-values were derived from Cox hazard regression models with adjustment for age at enrollment, race and ethnicity, and treatment arms. Estimates of per risk allele were derived from an additive genetic model by treating genotypes as 0, 1, and 2 according to the copy number of risk allele. *P*-values ≤ .001 (indicated in bold) were considered statistically significant to account for multiple testing for 12 genetic variants in *GATA3*.
Abbreviations: CI = Confidence interval; HR = hazard ratio.

	Tamoxifen-Treated Group (n = 220)			No Tamoxifen-Treated Group (n = 221)		
Genotype	**No. Events/Patients**	**Adjusted HR (95% CI)**	**Adjusted *P***	**No. Events/Patients**	**Adjusted HR (95% CI)**	**Adjusted *P***
A. Overall survival						
rs3802604						
AA	14/78	1.00		17/85	1.00	
GA	19/105	1.09 (0.54-2.20)	.82	31/100	1.56 (0.86-2.81)	.14
GG	17/35	2.66 (1.31-5.43)	.007	13/35	2.23 (1.08-4.60)	.03
Per risk allele		1.66 (1.13-2.44)	.009		1.50 (1.05-2.14)	.03
rs568727						
CC	13/76	1.00		14/82	1.00	
AC	19/105	1.02 (0.50-2.06)	.96	33/99	1.97 (1.05-3.71)	.04
AA	18/37	2.89 (1.40-5.96)	.004	14/40	2.28 (1.08-4.80)	.03
Per risk allele		1.71 (1.15-2.54)	.008		1.51 (1.07-2.14)	.02
B. Disease-free survival						
rs3802604						
AA	21/78	1.00		28/85	1.00	
GA	27/105	0.95 (0.53-1.70)	.87	38/100	1.16 (0.71-1.89)	.56
GG	22/35	2.59 (1.41-4.73)	.002	15/35	1.53 (0.81-2.87)	.19
Per risk allele		1.62 (1.16-2.26)	.005		1.22 (0.90-1.67)	.2
rs568727						
CC	20/76	1.00		24/82	1.00	
AC	28/105	0.98 (0.55-1.74)	.94	40/99	1.43 (0.86-2.39)	.17
AA	22/37	2.60 (1.41-4.82)	.002	17/40	1.65 (0.88-3.07)	.12
Per risk allele		1.60 (1.14-2.24)	.007		1.29 (0.96-1.75)	.09

Table 3 GATA3 Single Nucleotide Polymorphisms and Breast Cancer Outcomes Among Patients in the Treated Group by Tamoxifen Therapy in S8897 Trial

HRs, 95% CIs, and *P*-values were derived from Cox hazard regression models with adjustment for age at enrollment, race and ethnicity, and treatment arms. Estimates of per risk allele were derived from an additive genetic model by treating genotypes as 0, 1, and 2 according to the copy number of risk allele. *P*-values ≤ .001 (indicated in bold) were considered statistically significant to account for multiple testing for 12 genetic variants in *GATA3*.
Abbreviations: CI = Confidence interval; HR = hazard ratio.

were associated with either OS or DFS. The results remained unchanged when stratified by whether patients received adjuvant chemotherapy (see Supplemental Table 6 in the online version), aromatase inhibitors, menopausal status, or when the analyses were restricted to non-Hispanic White women (data not shown). In a joint analysis in which data were pooled from the SWOG and Pathways Study cohorts, the AA genotype of rs568727 was associated with both OS and DFS at a nominal significance level ($P < .05$), and the GG genotype of rs3802604 showed a marginal association with OS and DFS (Table 4). However, the test for study heterogeneity between the 2 cohorts was significant ($P < .05$).

Potential Functional Impact of GATA3 SNPs

As summarized in Table 5, there is ample bioinformatic evidence supporting that those SNPs may alter the functionality of GATA3. For rs3802604, the most significant SNP in our study, it was predicted to disrupt chromatin state in a variety of tissues, including breast; alter motifs for multiple transcription factors, including estrogen receptor alpha central in the ER signaling pathway; and perturb the binding of proteins in Chip-Seq experiments. Moreover,

this SNP has also been identified as a *cis*-eQTL associated with GATA3 expression in peripheral blood monocytes ($P = 1.7e-6$).[28]

Discussion

There are strong biological rationales for a potential predictive value of *GATA3* SNPs for outcomes after breast cancer adjuvant chemotherapy, particularly among patients who were subsequently treated with tamoxifen. Results from the S8897 cohort support this hypothesis, and in silico bioinformatic prediction of SNP functionality also support these findings.

However, the observations in the S8897 cohort were not replicated in the independent Pathway Study cohort, with null results. We considered several reasons for the lack of a cross-validation success between the 2 cohorts. First, the length of follow-up time differed, with a median 18 years in S8897 and only 6 years in the Pathways Study. Indeed, Kaplan-Meier curves of OS by rs3802604 genotypes did not start to separate until approximately 7 years after enrollment in S8897. Nevertheless, the separation of survival curves of DFS and by the other top SNP rs568727 did become apparent early during the follow-up. Second, because of the varying lengths of

Genotype	No. Events/ Patients	Adjusted HR (95% CI)	Adjusted P
Table 4	**Joint Analysis of GATA3 Single Nucleotide Polymorphisms and Breast Cancer Outcomes in the S8897 Trial and Pathways Study**		
A. Overall survival			
rs3802604			
AA	34/261	1.00	
GA	44/368	0.98 (0.62-1.53)	.92
GG	23/122	1.59 (0.94-2.71)	.09
Per risk allele		1.23 (0.93-1.63)	.15
rs568727			
CC	39/290	1.00	
AC	37/364	0.71 (0.45-1.11)	.13
AA	25/96	1.86 (1.11-3.12)	.02
Per risk allele		1.26 (0.94-1.68)	.13
B. Disease-free survival			
rs3802604			
AA	52/261	1.00	
GA	71/368	0.96 (0.67-1.38)	.84
GG	33/122	1.49 (0.96-2.30)	.08
Per risk allele		1.18 (0.94-1.49)	.15
rs568727			
CC	57/290	1.00	
AC	66/364	0.85 (0.59-1.21)	.37
AA	33/96	1.80 (1.16-2.79)	.009
Per risk allele		1.25 (0.98-1.59)	.07

HRs, 95% CIs, and P-values were derived from Cox hazard regression models with adjustment for study, age at enrollment, race and ethnicity, and treatment arms. Estimates of per risk allele were derived from an additive genetic model by treating genotypes as 0, 1, and 2 according to the copy number of risk allele.
Abbreviations: CI = Confidence interval; HR = hazard ratio.

chemotherapy efficacy. If the SNP effects of GATA3 were specific to 5-fluorouracil, which is no longer commonly used in the adjuvant settings for breast cancer, instead of cyclophosphamide or cytotoxic chemotherapy in general, we would not expect the replication of associations in the second cohort.

With the first cohort from a clinical trial conducted in the early 1990s where treatment regimens were closely prescribed, and the second cohort from a more contemporary observational cohort based in a large integrated healthcare system that reflects a community-based clinical setting, these cohort differences are logical reasons for the inconsistent findings. If the findings had been replicated, heterogeneity between the discovery and the validation cohorts would have underscored the robustness of the GATA3 SNPs as biomarkers of potential clinical utility outside the tightly controlled "experimental" setting of clinical trials. However, failure to replicate suggests that these SNPs do not play important roles in breast cancer under current practice.

Alternatively, the lack of replication in the second cohort in our study could reflect the true lack of an impact of GATA3 genetic variations on breast cancer outcomes. The 2 GWAS index SNPs with ALL susceptibility and relapse, rs3824662 and rs3781093, were not significant in the treated or untreated group in S8897, but only showed associations at a nominal significance level in the subgroup of patients who received adjuvant chemotherapy without tamoxifen. Although the direction of the associations in our study of breast cancer outcomes was consistent with the GWAS findings in ALL, it is in contrast to our hypothesis of a more important role of GATA3 for endocrine therapy. There was some evidence of rs3824662 being an eQTL and causal SNP responsible for the GWAS signal,[19] and rs3824602 as an eQTL associated with GATA3 expression in monocytes, yet definitive evidence from functional studies still await, and it remains to be determined whether the functional impact is specific to leukocytes or applicable to other cell types including mammary cells.

To date, genetic association studies for breast cancer prognosis and prediction of drug efficacy and toxicity, candidate-gene based or genome-wide design alike, have not produced many bona fide risk variants, which is in stark contrast with studies of breast cancer susceptibility, where more than 100 loci have been identified through GWAS based on ever increasing sample sizes.[29] It was hypothesized that the effect size of common genetic variants for toxicity and survival outcomes might be larger than that for cancer susceptibility owing to lack of evolutionary selection pressure, which presumably would facilitate the power of discovery.[30] However, this advantage may have been more or less dampened by the complexity of cancer outcomes as study phenotypes. Unlike binary cancer susceptibility (yes vs. no), cancer outcomes are time-varying and under influences such as cytotoxic cancer therapy, the effects of which may be overwhelming and mask any moderate genetic effects. Also, owing to complex outcome phenotypes and multifaceted heterogeneity in patient characteristics and treatment modalities, it is more challenging to pool and harmonize data collected under different settings and from various sources, despite the necessity and proven value of pooling across studies in cancer susceptibility GWAS. Our present study highlights those challenges and calls for large collaborative efforts to overcome them.

follow-up time, the event rates in the Pathways Study were much lower than those in S8897. However, HRs from regression models in the Pathways Study were closer to null than any suggestive trends consistent with those from S8897. Thus, it is unclear whether simply extending the follow-up time and accumulating more events in the second cohort would result in convergence of findings with that in the first cohort. Third, in the treated group in S8897, all patients received CAF versus CMF, whereas in the Pathways Study, chemotherapy regimens were more heterogeneous, with a majority of patients receiving CA (cyclophosphamide and doxorubicin) (39%), CT (cyclophosphamide and docetaxel) (23%), or C (cyclophosphamide) monotherapy (19%), and CAF and CMF only occasionally prescribed (<1% each). Because the findings related to GATA3 SNPs in S8897 were observed primarily in the treated group but not in the untreated group without adjuvant chemotherapy, we speculated the associations to be predictive for

Table 5	Bioinformatic Evidence of Potential Functionality of GATA3 Single Nucleotide Polymorphisms						
Variant	Promoter Histone Marks	Enhancer Histone Marks	DNAse Sensitivity	Proteins Bound in Chip-Seq Experiment	Motifs Changed	eQTL	Evolutionary Conservation
rs1244186	19 tissues (including BRST)	9 tissues (including BRST)	35 tissues (including BRST)	FOXA1, FOXA2, HNF4A, P300, TCF4, HAE2F1	NANOG, TATA	GATA3 in lymphoblastoid	Yes
rs2275806	15 tissues (including BRST)	14 tissues (including BRST)	36 tissues (including BRST)	CTCF, RAD21, SMC3, YY1, CTCFL, HAE2F1, POL2, ZN F143	GR, ZBRK1		
rs369421	15 tissues (including BRST)	5 tissues (including BRST)	26 tissues (including BRST)	BATF, NFKB, PU1, POL2, CTBP2	EBF, ETS, PU1, STAT, TATA	GATA3 in whole blood	Yes
rs2229359	6 tissues (including BRST)	9 tissues (including BRST)	11 tissues (including BRST)	CMYC, POL2	EGR1, LBP1, NRSF, PAX4		Yes
rs3781093	10 tissues (including BRST)	9 tissues (including BRST)	15 tissues (including BRST)	ZNF263		CEMP1 and FCGR2B in breast tumors	
rs3802604	14 tissues (including BRST)	10 tissues (including BRST)	34 tissues (including BRST)	BATF, EBF1, POL2, SIX5, NFKB, P300, RAD21, ZNF263	BP1, ESR1, EWSR1-FL1, GR, NRSF, PU1, RAD2 1, SMC3, SPZ1, TATA, WT1, ZNF263, ZNF143	GATA3 in blood monocyte	Yes
rs3824662	BRST	13 tissues (including BRST)	12 tissues (including BRST)	PU1, HAE2F1	NF1, NRSF, PAX6, ZBTB3	XAGE1B in blood mononcytes	
rs568727	ESC, KID	10 tissues (including BRST)	9 tissues (including BRST)	POL2	MAF, NR2F2		
rs3802600		4 tissues	BLD	GATA2	GR, HSF		
rs528778	BRN	6 tissues (including BRST)	BRST, BRN, SKIN	POL2	POU5F1		
rs10905284			ESDR, THYM, VAS		DMRT2, YY1	AOC2 in blood monocytes	
rs434645		12 tissues (including BRST)	ESDR, ESDR, IPSC		BCL, EGR1, ETS, GATA, OBOX6, PU1, PAX5, SPIB		

Abbreviations: BLD = blood; BRN = brain; BRST = breast tissue; eQTL = expression quantitative trait; ESC = ES-13 cells; ESDR = H1 BMP4 derived trophoblast cultured cells; IPSC = iPS DF 6.9 cells; KID = fetal kidney; SKIN = skin; THYM, fetal thymus; VAS = HUVEC umbilical vein endothelial primary cells.

In addition to the heterogeneities between the discovery cohort and the replication cohort as discussed above, another limitation should also be noted. S8897 was designed to randomize patients deemed at high risk to CMF versus CAF with or without tamoxifen, whereas those at low risk were not randomized and did not received chemotherapy. Because achieved normal lymph nodes were not available from patients at high risk based on large tumor size or hormonal receptor negativity, they were not included in this ancillary study. Thus, the patient population in this study represented those with smaller tumor or with hormonal receptor-positive cancer. This limits the generalizability of the study findings to low-risk patients.

In conclusion, consistent with the established role of GATA3 as one of the most important transcription factors in breast tissue, our present study provides some indication for potential predictive values of *GATA3* germline variants for breast cancer chemotherapy and endocrine therapy, and bioinformatic evidence supports these findings. However, these observations in one cohort were not replicated in another cohort. This may have been a result of substantial heterogeneity between the discovery and the replication cohorts, including differences in chemotherapy regimens and length of follow-up. Our findings call for future studies to continue investigating *GATA3* SNPs in breast cancer treatment outcomes and whether they are relevant to current and evolving practice in breast cancer treatment.

Disclosure

The authors have stated that they have no conflicts of interest.

Supplemental Data

Supplemental tables accompanying this article can be found in the online version at https://doi.org/10.1016/j.clbc.2019.02.010.

References

1. Chou J, Provot S, Werb Z. GATA3 in development and cancer differentiation: cells GATA have it! *J Cell Physiol* 2010; 222:42-9.
2. Takaku M, Grimm SA, Wade PA. GATA3 in breast cancer: tumor suppressor or oncogene? *Gene Expr* 2015; 16:163-8.
3. Asch-Kendrick R, Cimino-Mathews A. The role of GATA3 in breast carcinomas: a review. *Hum Pathol* 2016; 48:37-47.
4. Theodorou V, Stark R, Menon S, Carroll JS. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res* 2013; 23:12-22.
5. Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* 2006; 127:1041-55.
6. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012; 490:61-70.
7. Chou J, Lin JH, Brenot A, Kim JW, Provot S, Werb Z. GATA3 suppresses metastasis and modulates the tumour microenvironment by regulating microRNA-29b expression. *Nat Cell Biol* 2013; 15:201-13.
8. Mehra R, Varambally S, Ding L, et al. Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res* 2005; 65: 11259-64.
9. Yoon NK, Maresh EL, Shen D, et al. Higher levels of GATA3 predict better survival in women with breast cancer. *Hum Pathol* 2010; 41:1794-801.
10. Voduc D, Cheang M, Nielsen T. GATA-3 expression in breast cancer has a strong association with estrogen receptor but lacks independent prognostic value. *Cancer Epidemiol Biomarkers Prev* 2008; 17:365-73.
11. Albergaria A, Paredes J, Sousa B, et al. Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. *Breast Cancer Res* 2009; 11:R40.
12. Ciocca V, Daskalakis C, Ciocca RM, Ruiz-Orrico A, Palazzo JP. The significance of GATA3 expression in breast cancer: a 10-year follow-up study. *Hum Pathol* 2009; 40:489-95.
13. Jacquemier J, Charafe-Jauffret E, Monville F, et al. Association of GATA3, P53, Ki67 status and vascular peritumoral invasion are strongly prognostic in luminal breast cancer. *Breast Cancer Res* 2009; 11:R23.
14. Parikh P, Palazzo JP, Rose LJ, Daskalakis C, Weigel RJ. GATA-3 expression as a predictor of hormone response in breast cancer. *J Am Coll Surg* 2005; 200:705-10.
15. Gulbahce HE, Sweeney C, Surowiecka M, Knapp D, Varghese L, Blair CK. Significance of GATA-3 expression in outcomes of patients with breast cancer who received systemic chemotherapy and/or hormonal therapy and clinicopathologic features of GATA-3-positive tumors. *Hum Pathol* 2013; 44:2427-31.
16. Fang SH, Chen Y, Weigel RJ. GATA-3 as a marker of hormone response in breast cancer. *J Surg Res* 2009; 157:290-5.
17. Tominaga N, Naoi Y, Shimazu K, et al. Clinicopathological analysis of GATA3-positive breast cancers with special reference to response to neoadjuvant chemotherapy. *Ann Oncol* 2012; 23:3051-7.
18. Yu KD, Zhu R, Zhan M, et al. Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer. *Clin Cancer Res* 2013; 19:2723-33.
19. Perez-Andreu V, Roberts KG, Harvey RC, et al. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nat Genet* 2013; 45:1494-8.
20. Perez-Andreu V, Roberts KG, Xu H, et al. A genome-wide association study of susceptibility to acute lymphoblastic leukemia in adolescents and young adults. *Blood* 2015; 125:680-6.
21. Migliorini G, Fiege B, Hosking FJ, et al. Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype. *Blood* 2013; 122:3298-307.
22. Enciso-Mora V, Broderick P, Ma Y, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). *Nat Genet* 2010; 42:1126-30.
23. Ho IC, Tai TS, Pai SY. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. *Nat Rev Immunol* 2009; 9:125-35.

24. Hutchins LF, Green SJ, Ravdin PM, et al. Randomized, controlled trial of cyclophosphamide, methotrexate, and fluorouracil versus cyclophosphamide, doxorubicin, and fluorouracil with and without tamoxifen for high-risk, node-negative breast cancer: treatment results of Intergroup Protocol INT-0102. *J Clin Oncol* 2005; 23:8313-21.

25. Kwan ML, Ambrosone CB, Lee MM, et al. The Pathways Study: a prospective study of breast cancer survivorship within Kaiser Permanente Northern California. *Cancer Causes Control* 2008; 19:1065-76.

26. Yao S, Kwan ML, Ergas IJ, et al. Association of serum level of vitamin d at diagnosis with breast cancer survival: a case-cohort analysis in the Pathways Study. *JAMA Oncol* 2017; 3:351-7.

27. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; 40(Database issue):D930-4.

28. Heinzen EL, Ge D, Cronin KD, et al. Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol* 2008; 6:e1.

29. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 2015; 47:373-80.

30. Wheeler HE, Maitland ML, Dolan ME, Cox NJ, Ratain MJ. Cancer pharmacogenomics: strategies and challenges. *Nat Rev Genet* 2013; 14:23-34.

Supplemental Data

Supplemental Figure 1	Schemas of the S8897 Trial and the Pathways Study
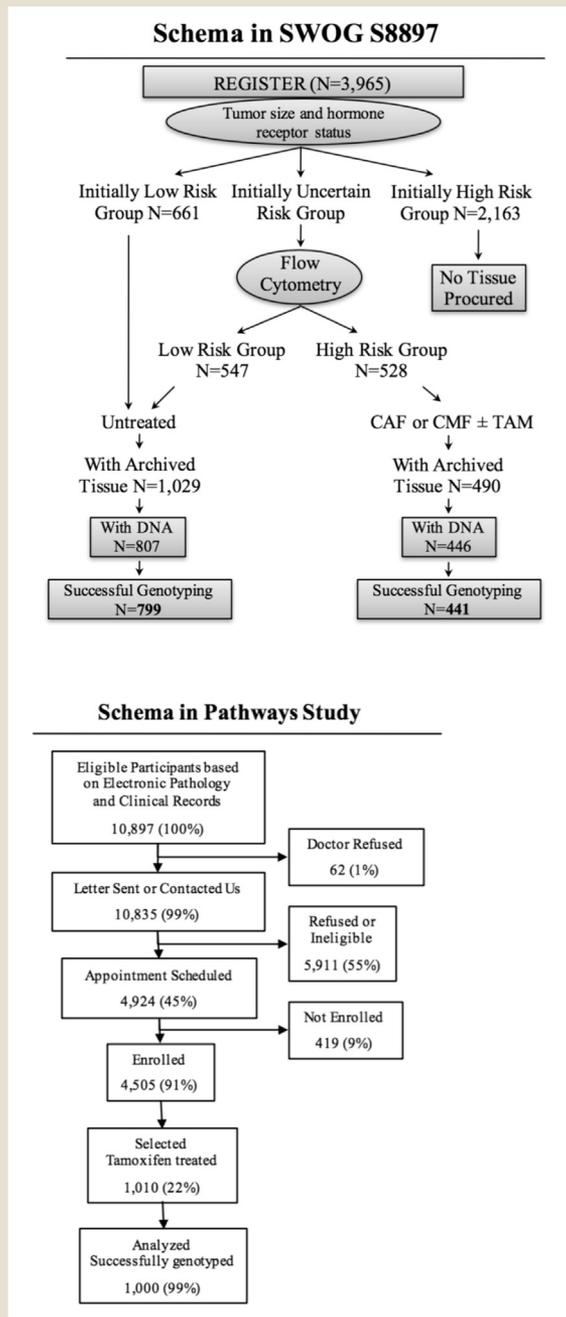
Abbreviations: CAF = Cyclophosphamide, doxorubicin, 5-fluorouracil; CMF = cyclophosphamide, methotrexate, 5-fluorouracil; TAM = tamoxifen.

Supplemental Figure 2 Kaplan-Meier Curves of Disease-free Survival by the Genotypes of rs3802604 and rs568727, the Two Single Nucleotide Polymorphisms in GATA3 Remained Significant after Correcting for Multiple Comparison, are Plotted Separately for the Treated and Untreated Group in S8897 Trial. *P*-Values are From Log-Rank test. A, rs3802604 in Treated Group; B, rs3802604 in Untreated Group; C, rs568727 in Treated Group; D, rs568727 in Untreated group