

Common Genetic Variation in *GATA-Binding Protein 3* and Differential Susceptibility to Breast Cancer by Estrogen Receptor α Tumor Status

Montserrat Garcia-Closas,¹ Melissa A. Troester,² Ying Qi,^{3,4} Anita Langerød,⁴ Meredith Yeager,³ Jolanta Lissowska,⁶ Louise Brinton,¹ Robert Welch,³ Beata Peplonska,⁷ Daniela S. Gerhard,⁸ Inger Torhild Gram,¹⁰ Vessela Kristensen,^{4,5} Anne-Lise Børresen-Dale,^{4,5} Stephen Chanock,^{1,3,9} and Charles M. Perou²

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland; ²Departments of Genetics and Pathology, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina; ³Core Genotyping Facility Advanced Technology Center, National Cancer Institute, Gaithersburg, Maryland; ⁴Department of Genetics, Institute for Cancer Research, Rikshospitalet-Radiumhospitalet Medical Centre and ⁵Faculty of Medicine, University of Oslo, Oslo, Norway; ⁶Cancer Center and M. Skłodowska Curie Institute of Oncology, Warsaw, Poland; ⁷Nofer Institute of Occupational Medicine, Lodz, Poland; ⁸Office of Cancer Genomics and ⁹Pediatrics Oncology Branch, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland; and ¹⁰Institute of Community Medicine, University of Tromsø, Tromsø, Norway

Abstract

GATA-binding protein 3 (*GATA3*) is a transcription factor and a putative tumor suppressor that is highly expressed in normal breast luminal epithelium and estrogen receptor α (ER)-positive breast tumors. We hypothesized that common genetic variation in *GATA3* could influence breast carcinogenesis. Four tag single-nucleotide polymorphisms (SNP) in *GATA3* and its 3' flanking gene *FLJ4598* were genotyped in two case control studies in Norway and Poland (2,726 cases and 3,420 controls). Analyses of pooled data suggested a reduced risk of breast cancer associated with two intronic variants in *GATA3* in linkage disequilibrium (rs3802604 in intron 3 and rs570613 in intron 4). Odds ratio (95% confidence interval) for rs3802604 heterozygous and rare homozygous versus common homozygous were 0.85 (0.75-1.95) and 0.82 (0.62-0.96), respectively ($P_{\text{trend}} = 0.004$). Stronger associations were observed for subjects

with ER-negative, than ER-positive, tumors ($P_{\text{heterogeneity}} = 0.01$ for rs3802604; $P_{\text{heterogeneity}} = 0.09$ for rs570613). Although no individual SNPs were associated with ER-positive tumors, two haplotypes (GGTC in 2% of controls and AATT in 7% of controls) showed significant and consistent associations with increased risk for these tumors when compared with the common haplotype (GATT in 46% of controls): 1.71 (1.27-2.32) and 1.26 (1.03-1.54), respectively. In summary, data from two independent study populations showed two intronic variants in *GATA3* associated with overall decreases in breast cancer risk and suggested heterogeneity of these associations by ER status. These differential associations are consistent with markedly different levels of *GATA3* protein by ER status. Additional epidemiologic studies are needed to clarify these intriguing relationships. (Cancer Epidemiol Biomarkers Prev 2007;16(11):2269-75)

Received 5/16/07; revised 8/3/07; accepted 8/16/07.

Grant support: NIH National Cancer Institute Breast Cancer Specialized Program of Research Excellence grants P50-CA58223 and R01-CA-101227-01 (C.M. Perou) and UNC Lineberger Cancer Control Education Program grant R25 CA57726 (M.A. Troester). The Polish Breast Cancer Study was supported by the Intramural Research Program of NIH National Cancer Institute Division of Cancer Epidemiology and Genetics and the Center for Cancer Research. The Tromsø Mammography and Breast Cancer Study was conducted in collaboration with Department of Clinical Research and the Department of Radiology, Center for Breast Imaging, University Hospital of North Norway; Norwegian Women and Cancer Study, University of Tromsø; and Cancer Registry of Norway. The study was supported by the Norwegian Cancer Society, Aakre Foundation, and Norwegian Women's Public Health Association. This work was also supported by the European Union's EU FP6 grant 502983, Research Council of Norway grant 155218/300, and Norwegian Cancer Society grant D 99061. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Cancer Epidemiology Biomarkers and Prevention Online (<http://cebp.aacrjournals.org/>). M. Garcia-Closas and M.A. Troester contributed equally to the drafting of the manuscript.

Requests for reprints: Montserrat Garcia-Closas, Hormonal and Reproductive Epidemiology Branch Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, 6120 Executive Boulevard, Room 7076, MSC 7234, Rockville, MD 20852-7234. Phone: 301-435-3981; Fax: 301-402-0916. E-mail: montse@nih.gov
Copyright © 2007 American Association for Cancer Research.
doi:10.1158/1055-9965.EPI-07-0449

Introduction

GATA-binding protein 3 (*GATA3*) is a transcription factor that is highly expressed in normal breast luminal epithelium and the luminal A tumor subtype that has been defined in studies of gene expression patterns (1-5). *GATA3* expression highly correlates with expression of estrogen receptor α (ER; refs. 6, 7), as well as other genes thought to be important in breast luminal epithelial cell biology, including *LIV1*, *REGG*, and *TFF3* (8). Luminal A tumors are associated with better prognosis than other tumor subtypes (luminal B, basal-like, and HER2+ subtypes; refs. 1, 2), the latter of which show little or no expression of *GATA3*. Building upon previous reports, including metaanalysis of microarray studies of *GATA3* transcripts (7), Mehra et al. (9) have shown that *GATA3* protein expression in tumors predicted prognosis. These authors also confirmed that *GATA3* levels are lower in tumors that are ER-negative and in those with high histologic grade. Parikh et al. have also shown that among ER-positive tumors, *GATA3*

protein expression seems to predict tumor estrogen responsiveness (10).

The mechanisms underlying the role of *GATA3* in estrogen response and breast cancer prognosis are not clear. For example, *GATA3* mRNA is down-regulated in normal breast xenograft samples after estradiol treatment (11), but estradiol-treated MCF-7 cells do not show altered expression of *GATA3* mRNA (6). The discovery of somatic mutations in *GATA3* in some ER-positive breast cancers, coupled with the observation that *GATA3*-transduced cells have greatly decreased proliferation rates *in vitro*, suggest that *GATA3* act as a tumor suppressor (12). Thus, we hypothesized that common genetic variation could alter expression and/or function of *GATA3* and thus play an etiologic role in breast cancer. To address this hypothesis, we carried out a comprehensive evaluation of common variation in *GATA3* and evaluated its relation with breast cancer risk in two independent study populations from Norway and Poland.

Materials and Methods

Study Populations

Norwegian Breast Cancer Study. Breast cancer patients ($n = 731$) were enrolled in accordance with local institutional review board guidelines and included four series of previously described case series.

(a) Breast cancer patients sequentially enrolled at Ullevål University Hospital from 1990 to 1994, representing the breast cancer population during this period (13, 14). The mean age was 64 years (range, 28-92 years). Blood samples were collected in 1994 to 1996 from 119 patients who were still alive (~80%), living in the Oslo area, and consented to give blood (~70% of eligible women). Time from diagnosis to blood collection was 0 to 6 years.

(b) Breast cancer patients admitted to the Norwegian Radium Hospital in 1972 to 1991 (15). The mean age at diagnosis was 57 years (range, 27-94 years). Blood was drawn between 1987 and 1991 ($n = 224$), either during follow-up visits, relapse, or at diagnosis.

(c) Breast cancer patients diagnosed at the Norwegian Radium Hospital and treated with radiotherapy in 1975 to 1986. The mean age at diagnosis was 59 years (range, 26-75 years). Blood samples were collected in 1996 from a subset of patients alive at that time (21%) who were part of a treatment evaluation and agreed (82%) to provide a blood sample ($n = 263$; ref. 16).

(d) Breast cancer patients with stage I and stage II disease enrolled in the Oslo micrometastases study between 1995 and 1998 who had blood samples collected at the time of diagnoses ($n = 125$; refs. 17, 18). The average age at diagnosis was 56 years (range, 29-82 years). Blood sampling was done just before primary surgery for their breast cancer from all patients included in the study.

The majority of control subjects ($n = 1,015$) were women with a negative mammogram from the Tromsø Mammography and Breast Cancer Study conducted in 2001 and 2002. About 70% of the women, 55 to 71 years of age, residing in the municipality of Tromsø in Norway and attending the Norwegian Breast Cancer

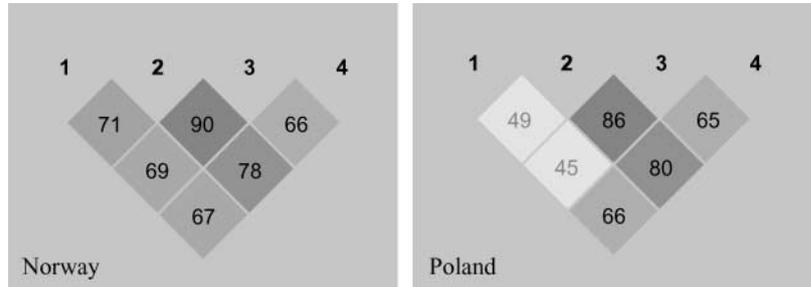
Screening Program at the University Hospital of North Norway agreed to participate (19). The study was approved by the National Data Inspection Board and the Regional Committee for Medical Research Ethics. In addition, we included healthy women ($n = 109$), 55 to 72 years of age, participating in the Norwegian Breast Cancer Screening Program in Bergen in 1999, with two negative mammogram during a 2-year period (20). From this latter group, women with p.o. hormone replacement therapy or history of diabetes or other endocrine disorders were excluded. The mean (range) ages for cases ($n = 731$) and controls ($n = 1,124$) in the Norwegian study were 56 (26-93) and 62 (55-72) years, respectively.

Polish Breast Cancer Study. A population-based case control study was conducted among women residing in two Polish cities, Warsaw and Lodz (21). Eligible cases were women ages 20 to 74 years who were newly diagnosed with either histologically or cytologically confirmed *in situ* or invasive breast cancer in 2000 to 2003. About 90% of cases were identified through a rapid identification system in participating hospitals, and the remainder through cancer registries, to ensure complete case ascertainment. Controls with no history of breast cancer were randomly selected from population lists during the case ascertainment period and were frequency matched to cases by city and age in 5-year categories. Institutional Review Board approval was obtained from all participating institutions, and signed informed consent was obtained for all respondents.

A total of 2,386 cases (79% of eligible cases) and 2,502 controls (69% of eligible controls) provided a personal interview on known and suspected risk factors. Blood samples for DNA extraction were obtained from 1,995 cases (84% of participating cases) and 2,296 controls (94% of participating controls). Most cases (94%) were diagnosed with invasive tumors. The mean (range) age was 56 (27-74) for cases and 56 (24-75) years for controls.

Genotyping. A resequence analysis of all exons, including the 5' and 3' untranslated regions and evolutionarily conserved intronic regions in *GATA3* was done in 94 healthy Norwegian women and 102 individuals in the SNP500Cancer panel (23). Single-nucleotide polymorphisms (SNP) were selected using the haplotype-tagging program of Stram et al. 22 with r^2 of >0.80 and minor allele frequency of >0.05 and genotyped in the Norwegian and Polish studies. SNP selection included the closely positioned upstream neighboring gene *FLJ45983* because of linkage disequilibrium observed with SNPs in *GATA3*, suggesting that variation in this gene might be involved in regulatory functions affecting *GATA3*. The function of this gene is unknown. Genotype analyses were done on blood DNA at the Core Genotyping Facility of the Division of Cancer Epidemiology and Genetics, National Cancer Institute for three SNPs in *GATA3* (rs3802604, IVS4+1468G>A; rs570613, IVS4+401T>C; rs422628, IVS4-27C>T) and one SNP in the 3' flanking gene *FLJ45983* (rs1149901, Ex1-425G>A). Description and methods for each genotype assay can be found at <http://snp500cancer.nci.nih.gov> (23). Duplicated DNA pairs from 95 subjects in the Polish study showed >99% concordance for all but one assay in intron 3 of *GATA3* (rs3802604) with 98% concordance.

Figure 1. Patterns of linkage disequilibrium across the *GATA3* and neighboring gene in the Norwegian and Polish control populations. Gray scale scheme is based on D' and LOD score values: white, $D' < 1$ and $\text{LOD} < 2$; shades of gray, $D' < 1$ and $\text{LOD} \geq 2$. Numbers in squares, D' values the order of SNPs from left to right is *FLJ45983* (rs1149901), *GATA3* (rs3802604, rs570613, rs422628).



Completion was $\geq 95\%$ for all assays in the Polish study (5% of samples fell into the NTC cluster, whereas $\leq 1\%$ were undetermined calls), and the percentage of completion was similar for cases and controls. Completion was $\geq 98\%$ for all assays in the Norwegian study, except for rs570613 which had a completion of 92%. We did not observe significant departures from Hardy-Weinberg equilibrium for any of the SNPs evaluated in the Norwegian or Polish control populations.

Statistical Analyses. Odds ratios (OR) and their 95% confidence intervals (95% CI) were derived from unconditional logistic regression models adjusting for age in 5-year categories. The association between genotypes and breast cancer risk was tested using a trend test. Differences in mean age at diagnosis in cases across different genotypes were tested using a t test. Age-specific estimates of ORs (95% CI) for genotype-disease associations were obtained from logistic regression analyses with interaction terms between genotypes (assuming a linear trend) and age categories. The test for interaction between genotypes and age was done by including an interaction term in a logistic regression model, considering both variables as continuous. We evaluated heterogeneity in the genotype ORs by hormone receptor status in logistic regression models among cases with receptor status as the outcome variable and genotypes as explanatory variables adjusting for age and study. Case control analyses were also done to estimate associations between genotypes and different tumor types. Heterogeneity of estimated ORs by study was tested by introducing an interaction term for genotype and study. Estimates obtained using pooled data from both studies were adjusted by study.

Pairwise linkage disequilibrium was estimated between SNPs based on D' and r^2 values using Haploview.¹¹ Block structure was determined using genotype data from the control population, and the solid spline of linkage disequilibrium option ($D'_{\text{threshold}} > 0.80$). Haplotype frequencies within each block, ORs, and their 95% CIs were estimated using HaploStats¹² (version 1.2.1; ref. 24). A global score statistic, adjusted for the matching factors age (in 5-year categories) and study site (Lodz or Warsaw), was used to evaluate the overall difference in haplotype frequencies between cases and controls. Phylogenetic trees [neighbor-joining (ref. 25),

nucleotide p distance] were constructed using MEGA 3.1¹³ (26) to assess nucleotide similarity of different haplotypes.

Results

We observed similar linkage disequilibrium patterns between SNPs in the control populations of the two study populations, although linkage disequilibrium between the SNP in the 3' neighboring gene *FLJ45983* and SNPs in *GATA3* was less in the Polish, than in the Norwegian, study (Fig. 1). There was low correlation between SNPs, with the exception of intron 3 (rs3802604) and intron 4 (rs570613) of *GATA3* ($r^2 = 0.67$ and 0.65 in the Norwegian and Polish populations, respectively). The allele frequencies among controls in the two study populations were similar. rs1149901 allele frequency was 0.26 and 0.28 for the Norwegian and Polish populations, respectively, 0.40 and 0.37 for rs3802604, 0.45 and 0.40 for rs570613, and 0.26 and 0.24 for rs422628. None of these differences were statistically significant.

An SNP in intron 4 of *GATA3* (rs570613) was associated with a significant decreased risk of breast cancer in the Norwegian study (Table 1). Data from the Polish study were consistent with a reduction in risk; however, the association was weaker and not statistically significant. Pooled analyses adjusting for study and age showed a significant reduction in risk ($P_{\text{trend}} = 0.004$), with no significant evidence for study heterogeneity (Table 1). Similar associations were observed for a linked SNP in intron 3 (rs3802604; $D'/r^2 = 0.86/0.65$ in Polish controls), although they were not statistically significant. Neither the SNPs in exon 1 of *FLJ45983* nor a SNP in intron 4 of *GATA3* (rs422628) were significantly associated with breast cancer risk in the Norwegian or the Polish studies. We observed no significant differences in genotype frequencies among the different types of controls or cases in the Norwegian study (data not shown).

We obtained information on ER and progesterone receptor status from diagnostic hospitals for 448 cases (58% of total) in Norway and 1,464 cases (73% of total) in Poland. The numbers of ER-negative cases were 178 in Norway and 505 in Poland. Case-only analyses showed that genotypes with the variant allele for the intron 3 SNP (rs3802604) in *GATA3* were less common for cases with ER-negative tumors than ER-positive tumors in both populations (Supplementary Table S1). Although the

¹¹ <http://www.broad.mit.edu/mpg/haploview/index.php>

¹² <http://mayoresearch.mayo.edu/mayo/research/biostat/schaid.cfm>

¹³ <http://www.megasoftware.net/>

association was only significant in the Polish study, there was no evidence of study heterogeneity, and the pooled ORs (95% CIs) for the association between genotypes and ER status among cases were 0.86 (0.70-1.06) and 0.65 (0.47-0.90) for heterozygous and homozygous variants compared with common homozygous ($P_{\text{trend}} = 0.01$).

Pooled case control analyses stratified by ER status showed a reduced risk for ER-negative tumors associated with the intron 3 polymorphism [pooled OR (95% CI), 0.86 (0.72-1.04) for heterozygotes and 0.72 (0.54-0.96) for homozygote variants; $P_{\text{trend}} = 0.02$] and no association with ER-positive tumors ($P_{\text{trend}} = 0.80$), which was consistent between the two study populations (Table 2). Similar associations were observed for the intron 4 SNP (rs570613) in *GATA3*; however, differences by ER status were not statistically significant.

Genotypes were not significantly associated with progesterone receptor status or progesterone receptor in combination with ER in analyses restricted to cases (Supplementary Tables S2 and S3). These analyses indicate that the genotype associations with breast cancer risk are not modified by progesterone receptor status. Data from the Polish study suggested the presence of interactions between the SNPs evaluated and age (Supplementary Table S4). Age-specific estimates showed the associations with reduced breast cancer risk to be limited to older women. Similarly, analyses by menopausal status in this study showed the inverse associations with risk for the three *GATA3* SNPs to be limited to postmenopausal women (67.2% of controls). Specifically, the per allele OR (95% CI) for premenopausal and postmenopausal women were, respectively, 1.14 (0.95-1.36) and 0.91 (0.81-1.01) for rs3802604 ($P_{\text{interaction}} = 0.034$), 1.11 (0.93-1.33) and 0.88 (0.79-0.98) for rs570613 ($P_{\text{interaction}} = 0.026$), and 1.24 (1.02-1.51) and 0.91 (0.81-1.04) for rs422628 ($P_{\text{interaction}} = 0.010$).

These analyses could not be done in the Norwegian study because controls were 55 years or older. Case-only analyses in both study populations showed that the SNPs

evaluated were significantly associated with age at diagnosis in the Polish study; however, associations were weaker and not statistically significant in the Norwegian population (data not shown). The associations with age at diagnosis in the Polish study remained significant after adjusting for ER status of the tumors (data not shown).

We observed nine haplotypes with a frequency >1% in the control populations (Table 3; Supplementary Table S5 for analyses by study). The most common haplotype (GATT) carried common alleles for all SNPs and was present in 46% of the pooled control population. Analyses of pooled data showed four haplotypes carrying both variants in introns 3 and 4 (rs3802604 and rs570613) individually associated with decreased breast cancer risk (GGCT in 14% of controls, GGCC in 1% of controls, AGCC in 18% of controls, and AGCT in 2% of controls). Compared with the common haplotype GATT, only two of these haplotypes (GGCT and AGCC) were associated with reductions in the risk for ER-negative tumors [0.71 (0.58-0.88) and 0.85 (0.72-1.01), respectively; Table 3]. We observed two haplotypes carrying only one of the two variants (GACT in 6% of controls and GGTC in 2% of controls), and none of them were associated with significant reductions in risk of ER-negative tumors. Interestingly, the haplotype carrying the variant in intron 3 (GGTC) and a haplotype with a variant in exon 1 of *FLJ45983* (AATT in 7% of controls) were associated with significant increases in risk for ER-positive tumors [1.71 (1.27-2.32) and 1.26 (1.03-1.54), respectively; Table 3]. This increase in risk was consistently found in the Norwegian and Polish populations (Supplementary Table S5).

Discussion

This comprehensive evaluation of common genetic variation of *GATA3* in two independent breast cancer studies in Norway and Poland showed consistent evidence for a differential association between common variation in *GATA3* and specific tumor types defined by

Table 1. Breast cancer risk and polymorphism in *GATA3* and its 3' flanking gene *FLJ45983* in two study populations in Norway and Poland

Gene	Norwegian study				Polish study			$P_{\text{heterogeneity}}$	Pooled estimates OR* (95% CI)
	SNP	Genotype	Case	Control	OR* (95% CI)	Case	Control		
<i>FLJ45983</i>	rs1149901	GG	385	609	1.00	960	1,113	1.00	1.00
	(Ex1-425G>A)	AG	273	410	1.14 (0.86-1.50)	785	903	1.01 (0.88-1.15)	1.03 (0.92-1.15)
		AA	35	85	0.55 (0.29-1.04)	151	158	1.11 (0.88-1.42)	0.99 (0.80-1.22)
		P_{trend}			0.51			0.52	0.81
<i>GATA3</i>	rs3802604	AA	276	387	1.00	777	859	1.00	1.00
	(IVS4+1468G>A)	AG	320	536	0.82 (0.61-1.11)	870	1,029	0.93 (0.82-1.06)	0.91 (0.81-1.03)
		GG	85	175	0.78 (0.51-1.17)	242	277	0.96 (0.79-1.17)	0.90 (0.76-1.06)
		P_{trend}			0.16			0.46	0.11
	rs570613	TT	256	320	1.00	724	776	1.00	1.00
	(IVS4+401T>C)	CT	335	568	0.72 (0.53-0.97)	893	1,075	0.89 (0.78-1.02)	0.85 (0.75-0.95)
		CC	104	211	0.62 (0.41-0.93)	278	325	0.91 (0.75-1.10)	0.82 (0.69-0.96)
		P_{trend}			0.01			0.19	0.004
	rs422628	TT	364	576	1.00	1,104	1,261	1.00	1.00
	(IVS4-27C>T)	CT	239	396	1.09 (0.81-1.46)	692	809	0.97 (0.85-1.11)	0.98 (0.88-1.10)
CC		30	77	0.71 (0.38-1.33)	105	109	1.10 (0.83-1.45)	0.97 (0.77-1.24)	
	P for trend			0.72			0.91	0.10	

*Adjusted by age, in addition to study in pooled analyses.

Table 2. Association between polymorphisms in *GATA3* and its 3' flanking gene and breast cancer risk stratified by ER status in two study populations in Norway and Poland

Gene SNP	Genotype	Controls	ER positive		P_{trend}	ER negative		P_{trend}	$P_{\text{heterogeneity}}^{\dagger}$
			Cases	OR* (95% CI)		Cases	OR* (95% CI)		
Norwegian study									
<i>FLJ45983</i>									
rs1149901 (Ex1-425G>A)	GG	609	141	1.00	0.42	101	1.00	0.43	0.42
	AG	410	108	1.04 (0.71-1.54)		68	1.18 (0.72-1.93)		
	AA	85	11	0.54 (0.22-1.3)		7	0.18 (0.02-1.36)		
<i>GATA3</i>									
rs3802604 (IVS4+1468G>A)	AA	387	98	1.00	0.26	73	1.00	0.14	0.21
	AG	536	126	0.86 (0.57-1.29)		85	0.77 (0.46-1.29)		
	AA	175	34	0.73 (0.4-1.31)		18	0.58 (0.26-1.29)		
rs570613 (IVS4+401T>C)	TT	320	96	1.00	0.02	68	1.00	0.21	0.44
	CT	568	128	0.69 (0.46-1.04)		85	0.77 (0.45-1.32)		
	CC	211	39	0.52 (0.29-0.92)		23	0.64 (0.31-1.34)		
rs422628 (IVS4-27C>T)	TT	576	137	1.00	0.99	92	1.00	0.43	0.75
	CT	396	89	0.91 (0.59-1.4)		60	1.12 (0.68-1.87)		
	CC	77	13	1.13 (0.54-2.36)		8	0.21 (0.03-1.55)		
Polish study									
<i>FLJ45983</i>									
rs1149901 (Ex1-425G>A)	GG	1,113	445	1.00	0.20	255	1.00	0.81	0.24
	AG	903	405	1.12 (0.96-1.32)		183	0.88 (0.72-1.09)		
	AA	158	70	1.11 (0.82-1.50)		41	1.13 (0.78-1.64)		
<i>GATA3</i>									
rs3802604 (IVS4+1468G>A)	AA	859	360	1.00	0.50	212	1.00	0.05	0.02
	AG	1,029	420	0.97 (0.82-1.15)		217	0.85 (0.69-1.05)		
	AA	277	131	1.13 (0.89-1.44)		52	0.76 (0.55-1.06)		
rs570613 (IVS4+401T>C)	TT	776	340	1.00	0.70	192	1.00	0.08	0.15
	CT	1,075	442	0.94 (0.79-1.11)		219	0.82 (0.66-1.02)		
	CC	325	139	0.98 (0.77-1.24)		65	0.81 (0.59-1.10)		
rs422628 (IVS4-27C>T)	TT	1,261	525	1.00	0.44	288	1.00	0.76	0.75
	CT	809	346	1.03 (0.87-1.21)		160	0.87 (0.70-1.07)		
	CC	109	53	1.17 (0.83-1.65)		31	1.25 (0.82-1.89)		
Pooled data									
<i>FLJ45983</i>									
rs1149901 (Ex1-425G>A)	GG	1,722	586	1.00	0.28	356	1.00	0.71	0.13
	AG	1,313	513	1.13 (0.98-1.31)		251	0.93 (0.78-1.12)		
	AA	243	81	1.01 (0.77-1.33)		48	1.02 (0.73-1.43)		
<i>GATA3</i>									
rs3802604 (IVS4+1468G>A)	AA	1,246	458	1.00	0.80	285	1.00	0.02	0.01
	AG	1,565	546	0.97 (0.84-1.13)		302	0.86 (0.72-1.04)		
	AA	452	165	1.05 (0.85-1.31)		70	0.72 (0.54-0.96)		
rs570613 (IVS4+401T>C)	TT	1,096	436	1.00	0.20	260	1.00	0.006	0.09
	CT	1,643	570	0.90 (0.78-1.05)		304	0.80 (0.66-0.97)		
	CC	536	178	0.89 (0.72-1.1)		88	0.71 (0.54-0.94)		
rs422628 (IVS4-27C>T)	TT	1,837	662	1.00	0.56	380	1.00	0.67	0.30
	CT	1,205	435	1.01 (0.88-1.17)		220	0.90 (0.75-1.09)		
	CC	186	66	1.12 (0.83-1.51)		39	1.08 (0.75-1.58)		

*Adjusted by age, in addition to study in pooled analyses.

[†] Based on case-only comparisons between genotype frequencies among ER-positive and ER-negative tumors (see Supplementary Table S1 for details).

ER status. In particular, two SNPs in strong linkage disequilibrium located in intron 3 (rs3802604) and intron 4 (rs570613) of *GATA3* and the two most common haplotypes carrying both of these variant alleles were associated with a decreased risk of ER-breast cancer. Two other haplotypes were associated with an increased risk of ER-positive tumors.

No previously published studies have evaluated the association between *GATA3* polymorphisms and breast cancer risk. However, two tag SNPs in this report (rs570613 and rs3802604) were included in a genome wide-scan done in a total of 1,200 breast cancer cases and 1,200 controls from the Nurse's Health Study (NHS) under the Cancer Genetic Markers of Susceptibility project.¹⁴

The NHS data were consistent with a reduced risk of breast cancer associated with the homozygous variant genotype for rs570613 [OR (95% CI), 1.01 (0.86-1.22) for CT versus TT and 0.78 (0.60-1.00) for CC versus TT; calculated from genotype frequencies found at <http://cgems.cancer.gov/data/>], as found in our two study populations. However, the rs3802604 SNP was not associated with risk in the Nurse's Health Study ($P_{2df \text{ adjusted}} = 0.72$). None of the other tag SNPs in *GATA3* genotyped in the Nurse's Health Study/Cancer Genetic Markers of Susceptibility project showed significant associations with breast cancer risk ($P_{2df \text{ adjusted}} > 0.18$).

The biology of *GATA3* is complex and encompasses functions in stem cells during differentiation and in terminally differentiated cells (27). Kaufman et al. have shown by targeted disruption of *GATA3* in mice that *GATA3* is required for proper hair follicle stem cell

¹⁴ <http://cgems.cancer.gov/>

Table 3. Association between *GATA3* haplotypes and breast cancer risk in pooled data from the Norwegian and Polish breast cancer studies, by ER status

Haplotypes*	Overall association				Association with ER-positive tumors			Association with ER-negative tumors			$P_{\text{heterogeneity}}^{\dagger}$
	Controls [‡]	All cases [‡]	OR [‡] (95% CI)	<i>P</i>	ER-positive cases [‡]	OR [§] (95% CI)	<i>P</i>	ER-negative cases [‡]	OR [§] (95% CI)	<i>P</i>	
G A T T	0.46	0.47	1.00		0.45	1.00		0.50	1.00		
. . . C	0.02	0.02	0.96 (0.72-1.30)	0.81	0.02	0.88 (0.59-1.32)	0.53	0.03	1.07 (0.69-1.67)	0.76	0.59
. . C .	0.06	0.06	0.89 (0.75-1.05)	0.17	0.06	0.95 (0.76-1.19)	0.67	0.06	0.89 (0.68-1.17)	0.42	0.64
. G . C	0.02	0.03	1.32 (1.01-1.71)	0.04	0.03	1.71 (1.27-2.32)	0.0005	0.02	1.07 (0.69-1.67)	0.76	0.02
. G C .	0.14	0.13	0.90 (0.79-1.01)	0.08	0.14	1.01 (0.86-1.17)	0.93	0.11	0.71 (0.58-0.88)	0.002	0.002
. G C C	0.01	0.01	0.90 (0.61-1.33)	0.61	0.01	0.85 (0.50-1.43)	0.54	0.02	1.06 (0.58-1.91)	0.86	0.58
A G C C	0.18	0.17	0.91 (0.82-1.01)	0.07	0.17	0.97 (0.84-1.11)	0.66	0.16	0.85 (0.72-1.01)	0.07	0.10
A . . .	0.07	0.08	1.13 (0.97-1.33)	0.13	0.09	1.26 (1.03-1.54)	0.02	0.08	1.00 (0.77-1.30)	0.99	0.07
A G C .	0.02	0.02	0.88 (0.61-1.28)	0.51	0.01	0.90 (0.56-1.45)	0.66	0.02	0.87 (0.49-1.56)	0.65	0.94
Rare haplotypes		1.04	(0.75-1.46)	0.80		1.29 (0.87-1.93)	0.21		0.59 (0.30-1.17)	0.13	0.04
<i>P</i> (global test)		0.04				0.008			0.09		

*SNPs are in the same order as in Table 1. Nucleotide changes individually associated with reduced breast cancer risk are bolded. Haplotypes are sorted by similarity of nucleotide sequences according to phylogenetic trees (26).

[†] *P* value for a test of heterogeneity of haplotype ORs by ER status.

[‡] Haplotype frequencies.

[§] Adjusted by age, in addition to study in pooled analyses.

function (28). In the immune system, *GATA3* is required for multiple developmental decisions (29). Thus, if *GATA3* signaling differs by developmental stage and cell type, then *GATA3* SNPs might be expected to have different effects on the risk of different tumor subtypes that arise from different cell lineages. Heterogeneity in *GATA3* signaling across different developmental lineages may underlie the etiologic heterogeneity with respect to ER status observed in this study.

A differential association of *GATA3* variants by ER status is consistent with previous observations about different roles of *GATA3* in ER-positive and ER-negative breast tumor subtypes. *GATA3* protein expression varies according to ER status (1, 2), predicts hormone therapy responsiveness (10), and predicts outcomes in ER-positive patients (7, 9). In human breast tissue, *GATA3* and ER expression are highly correlated with proteins expressed highly in cells that line the ducts (luminal epithelial cells). In addition, *GATA3* and ER are known to transcriptionally regulate a number of genes in common (6, 30, 31), so ER may influence *GATA3* and vice versa.

This study represents the first report of an association between *GATA3* polymorphisms and breast cancer risk. The SNP in intron 3 (rs3802604) is found in a highly conserved region of *GATA3*. Conservation is often high in regions of DNA with important functional consequences, but functional data has not been reported for any of the *GATA3* SNPs we investigated. The two intronic SNPs associated with risk were in linkage disequilibrium ($D' \sim 0.90$, $r^2 \sim 0.65$), and thus, it is unclear whether one or both SNPs are responsible for the observed associations. Neither of the two haplotypes with only one of these SNPs was significantly associated with ER-negative tumors, suggesting that both variants might be needed; however, this could also be due to low power to detect significant association for the less common haplotypes. It is also possible that one or both are in linkage disequilibrium with a protective allele not measured in our study. The observed associations between two haplotypes [one carrying the variant for

the intron 3 (rs3802604) SNP] and increased risk of ER-positive tumors were unexpected based on individual SNP analyses and require confirmation in other study populations.

The agreement between two large and independent study populations in this report and in NHS data online suggests validity of findings because potential biases that could explain associations are unlikely to be the same in both populations. The rate of participation in the Polish study is among the highest for population-based studies with collection of biological specimens. Although we cannot rule out selection bias, associations with most established risk factors for breast cancer were of expected direction and magnitude (21), indicating that selection bias is unlikely to be important. We had less knowledge regarding participation rates and distribution of breast cancer risk factors among the breast cancer cases in the Norwegian study; however, the findings were generally consistent between the study populations, particularly after stratification by ER tumor status. Both study populations were of homogeneous ethnic background, thus reducing the possibility of bias due to population stratification. Blood samples from most cases in the Norwegian were collected sometime after diagnosis, which provides the possibility of survival biases if the genotypes were to be related to survival.

Results from this report, together with the evidence that *GATA3* plays a role as a tumor suppressor (12), suggest that common variation in *GATA3* differentially affects the risk for developing breast cancer depending on the ER status of the tumors. Further epidemiologic studies aimed at clarifying this relation are warranted.

Acknowledgments

We thank Drs. Neonila Szeszenia-Dabrowska (Nofer Institute of Occupational Medicine), Witold Zatonski (M. Sklodowska-Curie Institute of Oncology and Cancer Center), and Aljcia Bardin-Mikolajczak (M. Sklodowska-Curie Institute of Oncology and Cancer Center) for their contribution to the design of the study and field work of the Polish Breast Cancer Study, Douglas

Richesson (DCEG, National Cancer Institute) for his assistance on statistical analyses, Anita Soni (Westat) for her work on study management for the Polish Breast Cancer Study, Pei Chao (IMS) for her work on data and sample management, and physicians, nurses, interviewers, and study participants for their efforts during field work.

References

- Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52.
- Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
- Gruvberger S, Ringner M, Chen Y, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* 2001;61:5979–84.
- West M, Blanchette C, Dressman H, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A* 2001;98:11462–7.
- 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.
- Hoch RV, Thompson DA, Baker RJ, Weigel RJ. GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int J Cancer* 1999;84:122–8.
- van de, RM, Perou CM, Tibshirani R, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *Am J Pathol* 2002;161:1991–6.
- Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003;100:8418–23.
- 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.
- Pariikh 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.
- Wilson CL, Sims AH, Howell A, Miller CJ, Clarke RB. Effects of oestrogen on gene expression in epithelium and stroma of normal human breast tissue. *Endocr Relat Cancer* 2006;13:617–28.
- Usary J, Llaca V, Karaca G, et al. Mutation of GATA3 in human breast tumors. *Oncogene* 2004;23:7669–78.
- Bukholm IK, Nesland JM, Karesen R, Jacobsen U, Borresen AL. Relationship between abnormal p53 protein and failure to express p21 protein in human breast carcinomas. *J Pathol* 1997;181:140–5.
- Langerød AZHBØ, Nesland JM, Bukholm IK, et al. TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. *Breast Cancer Res* 2007.
- Andersen TI, Holm R, Nesland JM, Heimdal KR, Ottestad L, Borresen AL. Prognostic significance of TP53 alterations in breast carcinoma. *Br J Cancer* 1993;68:540–8.
- Edvardsen H, Kristensen VN, Alnæs GIG, et al. Germline glutathione S-transferase variants in breast cancer: Relation to diagnosis and cutaneous long-term adverse effects after two fractionation patterns of radiotherapy. *Int J Radiation Oncology Biol Phys* 2007.
- Wiedswang G, Borgen E, Karesen R, et al. Detection of isolated tumor cells in bone marrow is an independent prognostic factor in breast cancer. *J Clin Oncol* 2003;21:3469–78.
- Naume B, Zhao H, Synnestevedt M, et al. Presence of micrometastasis in bone marrow is associated with different recurrence risk within molecular subtypes of breast cancer. *Molecular Oncology* 2007.
- Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E. Percentage density, Wolfe's and Tabar's mammographic patterns: agreement and association with risk factors for breast cancer. *Breast Cancer Res* 2005;7:R854–61.
- Helle SI, Ekse D, Holly JM, Lonning PE. The IGF-system in healthy pre- and postmenopausal women: relations to demographic variables and sex-steroids. *J Steroid Biochem Mol Biol* 2002;81:95–102.
- Garcia-Closas M, Brinton LA, Lissowska J, et al. Established breast cancer risk factors by clinically important tumour characteristics. *Br J Cancer* 2006;95:123–9.
- Stram DO, Leigh PC, Bretsky P, et al. Modeling and E-M estimation of haplotype-specific relative risks from genotype data for a case-control study of unrelated individuals. *Hum Hered* 2003;55:179–90.
- Packer BR, Yeager M, Burdett L, et al. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Res* 2006;34:D617–21.
- Schaid DJ. Evaluating associations of haplotypes with traits. *Genet Epidemiol* 2004;27:348–64.
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–25.
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 2004;5:150–63.
- Zhong JF, Zhao Y, Sutton S, et al. Gene expression profile of murine long-term reconstituting vs. short-term reconstituting hematopoietic stem cells. *Proc Natl Acad Sci U S A* 2005;102:2448–53.
- Kaufman CK, Zhou P, Pasolli HA, et al. GATA-3: an unexpected regulator of cell lineage determination in skin. *Genes Dev* 2003;17:2108–22.
- Pai SY, Truitt ML, Ting CN, Leiden JM, Glimcher LH, Ho IC. Critical roles for transcription factor GATA-3 in thymocyte development. *Immunity* 2003;19:863–75.
- Finlin BS, Gau CL, Murphy GA, et al. RERG is a novel ras-related, estrogen-regulated and growth-inhibitory gene in breast cancer. *J Biol Chem* 2001;276:42259–67.
- Oh DS, Troester MA, Usary J, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol* 2006;24:1656–64.