Research Article

Cancer Epidemiology, Biomarkers & Prevention

Rare Coding Variants and Breast Cancer Risk: Evaluation of Susceptibility Loci Identified in Genome-Wide Association Studies

Yanfeng Zhang[1], Jirong Long[1], Wei Lu[4], Xiao-Ou Shu[1], Qiuyin Cai[1], Ying Zheng[4], Chun Li[2], Bingshan Li[3], Yu-Tang Gao[5], and Wei Zheng[1]

Abstract

Background: To date, common genetic variants in approximately 70 loci have been identified for breast cancer via genome-wide association studies (GWAS). It is unknown whether rare variants in these loci are also associated with breast cancer risk.

Methods: We investigated rare missense/nonsense variants with minor allele frequency (MAF) $\leq$5% located in flanking 500 kb of each of the index single-nucleotide polymorphism (SNP) in 67 GWAS loci. Included in the study were 3,472 cases and 3,595 controls from the Shanghai Breast Cancer Study. Both single marker and gene-based analyses were conducted to investigate the associations.

Results: Single marker analyses identified 38 missense variants being associated with breast cancer risk at $P <$ 0.05 after adjusting for the index SNP. SNP rs146217902 in the *EDEM1* gene and rs200340088 in the *EFEMP2* gene were only observed in 8 cases ($P = 0.004$ for both). SNP rs200995432 in the *EFEMP2* gene was associated with increased risk with an OR of 6.2 [95% confidence interval (CI), 1.4–27.6; $P = 6.2 \times 10^{-3}$]. SNP rs80358978 in the *BRCA2* gene was associated with 16.5-fold elevated risk (95% CI, 2.2–124.5; $P = 2.2 \times 10^{-4}$). Gene-based analyses suggested eight genes associated with breast cancer risk at $P < 0.05$, including the *EFEMP2* gene ($P = 0.002$) and the *FBXO18* gene ($P = 0.008$).

Conclusion: Our results identified associations of several rare coding variants neighboring common GWAS loci with breast cancer risk. Further investigation of these rare variants and genes would help to understand the biologic mechanisms underlying the associations.

Impact: Independent studies with larger sample size are warranted to clarify the relationship between these rare variants and breast cancer risk. *Cancer Epidemiol Biomarkers Prev; 23(4); 622–8. ©2014 AACR.*

Introduction

Breast cancer is one of the most commonly diagnosed malignancies of women in the world (1). It is well established that genetic factors play an important role in breast cancer risk (2). Over the past several years, common variants, usually with minor allele frequency (MAF) > 5%, in approximately 70 loci have been identified as breast cancer risk factors via genome-wide association studies (GWAS; ref. 3). However, these common variants together explained only a small portion of the heritability for breast cancer.

It has been increasingly recognized that the missing heritability for breast cancer and other complex diseases may be partially explained by low-frequency variants. There are a large number of low-frequency variants in the human genome, and these rare coding variants are enriched for functional importance (4). Rare coding variants have been associated with multiple diseases, such as the *MTNR1B* gene for type 2 diabetes (5), *IFIH1* gene for type 1 diabetes (6), *APOA5*, *GCKR*, *LPL*, and *APOB* genes for hypertriglyceridemia (7) and *CHEK2*, *ATM*, *BRIP1*, *PALB2*, *RAD51C*, *RAD51D*, and *PPM1D* genes for breast cancer (8–13). Herein, we investigated low MAF coding variants in GWAS-identified loci regions for their association with breast cancer risk. Focusing on the flanking 500-kb regions of 67 GWAS-identified loci, we investigated low MAF nonsense/missense variants and their corresponding genes in a total of 3,472 cases and 3,595 controls from the Shanghai Breast Cancer Genetics Study.

Authors' Affiliations: [1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center; [2]Department of Biostatistics; [3]Department of Molecular Physiology and Biophysics, Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, Tennessee; [4]Shanghai Center for Disease Control and Prevention; and [5]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Corresponding Author: Jirong Long, Vanderbilt University School of Medicine, 2525 West End Avenue, 8th Floor, Nashville, TN 37203. Phone: 615-343-6741; Fax: 615-936-8241; E-mail: Jirong.long@vanderbilt.edu

doi: 10.1158/1055-9965.EPI-13-1043

Materials and Methods

Study populations

Study participants in the present study were drawn from four population-based studies conducted in Shanghai, the Shanghai Breast Cancer Study (SBCS), Shanghai Women's Health Study (SWHS), Shanghai Breast Cancer Survival Study (SBCSS), and the Shanghai Endometrial Cancer Study (SECS, contributed control data only). Detailed descriptions of participating studies have been published elsewhere (14–16). In brief, the SBCS is a two-stage (SBCS-I and SBCS-II), population-based, case–control study. SBCS-I recruitment occurred between August 1996 and March 1998; SBCS-II recruitment occurred between April 2002 and February 2005. Both studies identified patients with incident primary breast cancer through the population-based Shanghai Cancer Registry and randomly selected community controls from the general population in Shanghai. The SBCSS included newly diagnosed breast cancer cases ascertained via the Shanghai Cancer Registry between April 2002 and December 2006. The SECS is a population-based, case–control study of endometrial cancer conducted between January 1997 and December 2003 using a protocol similar to the SBCS; only community controls from the SECS were included in the present study. The SWHS is a population-based prospective cohort study of women from urban communities in Shanghai who were recruited between 1996 and 2000. The cohort has been followed by a combination of record linkage and active follow-ups to identify cause-specific mortality and cancer incidence by sites. All these studies are conducted among Chinese women in Shanghai, a genetically homogenous population, using very similar protocols in data and sample collection. Genomic DNA for all included participants was extracted using commercial DNA purification kits. Study protocols were approved by the institutional review boards of all institutions involved in the study, and informed consents were obtained from all study participants.

Genotyping array

Genotype assays were done by the Asian Exomechip, an expanded Illumina HumanExome-12v1_A Beadchip. The original exome array includes 247,870 markers focused on protein-coding regions selected from >12,000 samples with exome and genome sequencing data. The vast majority of these samples were from European ancestry populations, and approximately 600 Asian samples were included. Details about single-nucleotide polymorphism (SNP) contents and characteristics are described at Exomechip design (17). In brief, nonsynonymous variants observed three or more times in at least two studies, and splicing and stop-altering variants observed two or more times in at least two studies were selected. Additional array content includes variants associated with complex traits in previous GWAS, human leukocyte antigen tags, ancestry-informative markers, markers for identity-by-descent estimation, and random synonymous SNPs.

To improve the coverage for the low frequency variants in Asian population, we designed the Asian Exomechip by adding approximately 60K customer content variants onto the Illumina HumanExome-12v1_A Beadchip based on additional sequencing data. Included on the chip are also top SNPs selected from GWAS for follow-up. Three sequencing datasets were used to add additional nonsense/missense variants: exome sequencing in 581 Chinese women from SBCS, exome sequencing in 496 Singapore Chinese, and Asian data in the 1000 Genomes Project. Nonsynonymous, splicing, and stop-altering variants observed two or more times in any of these datasets or once in any two of the three datasets, were added ($N = 33{,}342$). Additional common variants ($N = 28{,}637$) were added to the chip for various GWAS follow-up and GWAS loci fine-mapping projects.

Genotyping and quality control

All samples were genotyped at the Genome Quebec Innovation Centre (Montreal, Quebec, Canada) following Illumina's protocol. On each 96-well plate, blind duplicate samples and two HapMap samples were included as quality control (QC). Genotype calling was carried out using Illumina's GenTrain version 2.0 clustering algorithm in GenomeStudio version 2011.1. Cluster boundaries were determined using study samples. After clustering, approximately 80,000 variants were manually reviewed and clusters were edited for 27,506 variants.

Further QC procedures were conducted using plink (18). We evaluated concordance rates for HapMap samples genotyped in our study and sequenced by the 1,000 Genomes Project (4). Principal components analyses (PCA) were conducted based on 3,200 ancestry informative markers on the Exomechip using EIGENSTRAT (19) to identify population outliers with the 1,000 Genomes Project data as reference. We also estimated pair-wise proportion of identify-by-descent to identify potentially genetically identical, unexpected duplicated samples or close relatives. The samples were excluded if (i) call rate < 98%, or (ii) consistence rates between the HapMap samples with 1000 Genomes data <99%, or (iii) heterozygosity outlier, or (iv) ethnic outliers, or (v) samples with close relationship, or (vi) consistence rates among duplicated samples < 99%, or (vii) samples with wrong sex. The SNPs were excluded if (i) MAF = 0, or (ii) call rate <98%, or (iii) genotyping concordance rate <98% in QC samples, or (iv) Hardy-Weinberg equilibrium test $P < 10^{-5}$, or (v) redundant SNPs, or (vi) cautions SNPs discovered by the Exomechip design group (17). A total of 8,200 samples plus 192 QC samples were genotyped. The final analysis dataset included 127,267 SNPs genotyped on 3,472 breast cancer cases and 3,595 controls.

Statistical Analyses

We used ANNOVAR program (20) to annotate all SNPs. We included all missense/nonsense variants located flanking 500 kb of the indexed SNP of 67 GWAS loci. If a protein-coding gene was partially covered within the

flanking 500-kb region, all missense/nonsense variants in the whole gene were included for analyses. For single-variant analysis, we used logistic score test adjusted for age implemented in Efficient and Parallelizable Association Container Toolbox (EPACTS) package (21). Further conditional analyses were conducted by adjusting the corresponding index SNP in each locus.

For gene-based analysis, we used the SKAT-O test with default parameters implemented in the EPACTS package. SKAT-O (22) encompasses burden tests and SKAT (23). Low-frequency variants of MAF $\leq$5% or MAF $\leq$1% within each gene were aggregated.

Results

Characteristics of the study population are shown in Table 1. All the known risk factors were associated with breast cancer risk in this study setting. Cases had higher educational attainment and were more likely to have a first-degree relative with breast cancer, a history of benign breast disease, be postmenopausal, and report early menarche than controls.

Single marker analyses

In the flanking regions of those 67 GWAS loci, a total of 1,272 missense/nonsense variants were included on the chip; 1,080 were rare variants with 0 < MAF $\leq$5% (Supplementary Table S1). A total of 38 rare variants (0 < MAF $\leq$5%) showed an association with breast cancer risk at $P <$ 0.05 after adjusted for the corresponding index SNP (Table 2). Notably, five rare variants were associated with breast cancer risk at $P < 0.01$. SNP rs146217902 in the *EDEM1* at 3p26.1 and rs200340088 in the *EFEMP2* at 11q13.1 were observed in 8 cases but not in any controls ($P = 0.004$ for both). Another SNP rs200995432 in the *EFEMP2* gene was associated with increased breast cancer risk with an OR of 6.2 [95% confidence interval (CI), 1.4–27.6; $P = 6.2 \times 10^{-3}$]. SNP rs80358978 in the *BRCA2* gene, 42 kb upstream from the GWAS SNP rs11571833, was associated with 16.5-fold elevated risk (95% CI, 2.2–124.5; $P = 2.2 \times 10^{-4}$). A rare variant rs143563006 in the *FBXO18* gene was associated with decreased risk of breast cancer with an OR being 0.60 (95% CI, 0.41–0.88) and a P value of 8.2×10^{-3}.

Gene-based analyses

Collapsing variants with MAF $\leq$5% within each gene suggested eight genes associated with breast cancer at $P <$ 0.05 (Table 3 and Supplementary Table S2). As the MAF of the majority of rare variants was $\leq$1%, similar results were found when MAF was set to $\leq$1%. These associations did not change materially after adjusting for corresponding GWAS index SNPs. At the locus *11q13.1*, two genes, *EFEMP2* and *RNASEH2C*, showed an association with breast cancer risk with $P = 0.002$ and $P = 0.04$, respectively. The *EFEMP2* gene was approximately 61.3 kb downstream from the GWAS SNP rs12575663, and the *RNASEH2C* gene was 87 kb upstream from the index SNP. At the 10p15.1, the *FBXO18* (consisting of 5 variants with MAF < 0.05) was strongly associated with breast cancer risk ($P = 8.0 \times 10^{-3}$). The other five genes showing associations were *KLHL26*, *OR2A12*, *TGFBR2*, *TRIP13*, and *VTI1A*.

Discussion

In the present study, we investigated associations of 1,080 missense/nonsense variants with an MAF of $\leq$5% in 337 genes at 67 GWAS loci among 3,472 Chinese breast cancer cases and 3,595 controls. Single marker analyses showed an association for 38 variants at $P < 0.05$. In particular, five variants were associated with breast cancer risk at $P < 0.01$, including rs200340088 and rs200995432 in

Table 1. Distribution of demographic characteristics and known breast cancer risk factors for cases and controls included in the study

Category	Cases (N = 3,472)	Controls (N = 3,595)	P
Demographic factors[a]			
Age, y (±SD)	53.2 ± 10.0	53.0 ± 9.3	0.38
Education ≥ high school (%)	55.9	40	<0.01
Reproductive risk factors			
Age at menarche (y)	14.4 ± 1.7	14.9 ± 1.8	<0.01
Postmenopausal (%)[b]	49.8	53.1	<0.01
Age at menopause[b]	49.0 ± 4.2	48.8 ± 3.9	0.29
Age at first live birth (y)[c]	26.9 ± 3.8	25.6 ± 4.1	<0.01
Other risk factors			
First-degree relative with breast cancer (%)	5.3	2.2	<0.01
Body mass index	24.1 ± 3.5	23.9 ± 3.4	0.05
Body mass index[b]	24.7 ± 3.7	24.4 ± 3.5	0.01

[a]Unless otherwise specified, mean ± SD are presented.
[b]Among postmenopausal women.
[c]Among parous women.

Table 2. SNPs associated with breast cancer risk at $P < 0.05$ in single marker analyses

Chromosome	Position	rs#	Alleles	Polyphen-2 score	SIFT score	Amino acid change	Gene	AF.case	AF.control	*P*	$P_{conditional}$	OR	Low 95% CI	Up 95% CI
3	5257572	rs146217902	A/G	0.096	0.12	Arg->Gln	*EDEM1*	0.0012	0.0000	4.09E−03	4.1E−03	1.67E+09	0.00	inf
5	917213	rs200263887	G/A	0	0	Ile->Val	*TRIP13*	0.0053	0.0028	0.02	1.7E−02	1.92	1.11	3.32
5	56155651	rs201579608	A/G	0.997	0	Arg->Gln	*MAP3K1*	0.0001	0.0013	0.01	1.3E−02	0.11	0.01	0.91
5	57750547	NA	G/A	0.088	0.1	Tyr->His	*PLK2*	0.0013	0.0029	0.04	3.4E−02	0.44	0.20	0.97
6	151914357	rs34563373	A/G	0.073	0.1	Arg->His	*C6orf97*	0.0006	0.0000	0.04	4.1E−02	1.67E+09	0.00	inf
6	152560708	rs190673256	T/C	0.001	0.78	Arg->Gln	*SYNE1*	0.0023	0.0008	0.03	3.7E−02	2.76	1.08	7.07
6	152603091	NA	T/C	0.006	0.11	Glu->Lys	*SYNE1*	0.0000	0.0006	0.05	4.9E−02	0.00	0.00	inf
10	5937039	rs143563006	C/G	0.262714	0	Val->Leu	*FBXO18*	0.0061	0.0100	0.01	8.1E−03	0.60	0.41	0.88
10	64005772	rs3765004	G/T	0.133	0.08	Lys->Thr	*RTKN2*	0.0207	0.0163	0.05	4.6E−02	1.28	1.00	1.64
10	114286892	rs184549091	C/T	0.973	0.01	Leu->Pro	*VTI1A*	0.0049	0.0076	0.04	3.6E−02	0.63	0.41	0.98
10	123843210	rs141547215	A/G	0.995	0	Gly->Arg	*TACC2*	0.0260	0.0324	0.02	1.9E−02	0.80	0.65	0.97
11	1491556	NA	T/C	NA	0.2	Arg->Gln	*AC091196*	0.0010	0.0026	0.02	2.4E−02	0.38	0.16	0.90
11	65408708	rs3741379	T/G	0.001	0.88	Ala->Ser	*SIPA1*	0.0266	0.0214	0.05	4.2E−02	1.25	1.00	1.54
11	65487550	NA	A/C	0.991	0.14	Arg->Leu	*RNASEH2C*	0.0001	0.0010	0.04	4.0E−02	0.15	0.02	1.21
11	65629960	rs2298447	T/C	0.156	0.03	Leu->Phe	*MUS81*	0.0013	0.0029	0.04	3.8E−02	0.45	0.20	0.97
11	65630970	rs143145862	A/G	0.092	0.04	Arg->Gln	*MUS81*	0.0004	0.0018	0.02	1.5E−02	0.24	0.07	0.84
11	65636047	rs200340088	T/C	0.993	0.02	Glu->Lys	*EFEMP2*	0.0012	0.0000	4.07E−03	4.0E−03	1.67E+09	0.00	inf
11	65639801	rs200995432	T/G	0.966	0	Pro->Thr	*EFEMP2*	0.0017	0.0003	0.01	6.2E−03	6.18	1.38	27.62
12	95603070	NA	A/T	0	0.19	Ser->Cys	*FGD6*	0.0000	0.0006	0.05	4.9E−02	0.00	0.00	inf
12	96288809	rs140348782	G/A	0.924	0.15	Ser->Pro	*CCDC38*	0.0016	0.0032	0.05	4.8E−02	0.49	0.24	1.01
12	96379914	rs181887143	T/C	0.346	0.08	Arg->His	*HAL*	0.0009	0.0024	0.03	2.5E−02	0.36	0.14	0.92
13	32907359	rs80358457	C/A	0.561	0.06	Thr->Pro	*BRCA2*	0.0006	0.0019	0.02	2.3E−02	0.30	0.10	0.90
13	32930651	rs80358978	A/G	0.999	0	Gly->Ser	*BRCA2*	0.0023	0.0001	2.18E−04	2.2E−04	16.51	2.19	124.50
14	37132495	NA	G/A	0.985	0.28	Asn->Ser	*PAX9*	0.0000	0.0006	0.05	4.6E−02	0.00	0.00	inf
14	91739003	rs182423495	A/G	NA	0.06	Pro->Leu	*CCDC88C*	0.0006	0.0018	0.03	3.5E−02	0.32	0.10	0.97
14	91763720	rs142539336	A/G	NA	0	Arg->Cys	*CCDC88C*	0.0062	0.0036	0.03	3.4E−02	1.70	1.05	2.76
14	91780306	NA	G/C	NA	0	Arg->Ser	*CCDC88C*	0.0035	0.0014	0.02	1.4E−02	2.38	1.15	4.92
16	53358202	NA	A/G	NA	0	Gly->Ser	*CHD9*	0.0003	0.0013	0.04	4.4E−02	0.23	0.05	1.07
16	53682940	rs142349647	T/C	0.242	0.03	Arg->Gln	*RPGRIP1L*	0.0001	0.0013	0.01	1.4E−02	0.12	0.01	0.91
16	81075049	rs192089732	T/C	0.992	0	Pro->Leu	*ATMIN*	0.0006	0.0000	0.04	4.1E−02	1.68E+09	0.00	inf
19	17317529	rs186514880	A/G	NA	0.13	Val->Ile	*MYO9B*	0.0014	0.0031	0.04	4.3E−02	0.47	0.22	0.99
19	17373389	NA	T/C	0.006	1	Arg->Gln	*USHBP1*	0.0007	0.0000	0.02	2.2E−02	1.69E+09	0.00	inf
19	17831801	NA	A/G	0.003	0.59	Val->Ile	*MAP1S*	0.0017	0.0004	0.02	1.6E−02	4.17	1.17	14.77
19	18546203	rs145899718	G/C	0.003	0.07	Glu->Gln	*ISYNA1*	0.0012	0.0026	0.04	3.7E−02	0.43	0.19	0.99
19	18779799	rs117020142	T/A	0.988	0.57	Tyr->Phe	*KLHL26*	0.0006	0.0018	0.03	2.7E−02	0.31	0.10	0.96
19	44118014	NA	T/C	NA	0.07	Arg->Cys	*SRRM5*	0.0017	0.0006	0.04	3.8E−02	3.12	1.01	9.69
19	44470165	rs141927408	T/C	0.416	0.01	Arg->Cys	*ZNF221*	0.0000	0.0007	0.03	2.8E−02	0.00	0.00	inf
22	40417776	rs141861984	A/G	0.058	0.37	Arg->Gln	*FAM83F*	0.0000	0.0006	0.05	4.8E−02	0.00	0.00	inf

Abbreviation: inf, infinite.

Table 3. Genes associated with breast cancer risk at $P < 0.05$ in gene-based analyses

Genes	Number of variants	Variants	Distance to index SNP (bp)	*P*	$P_{conditional}$[a]
EFEMP2	5	p.Pro9Thr, p.Thr312Ala, p.His292Tyr, p.Glu261Lys, p.Ile259Val	61,271	0.0021	0.0020
FBXO18	5	p.Val15Leu, p.His33Arg, p.Pro88Leu, p.Val552Ile, p.Ile981Phe	50,305	0.0082	0.0082
KLHL26	3	p.Val109Leu, p.Arg115Gln, p.Tyr531Phe	207,391	0.0303	0.0260
OR2A12	6	p.Tyr34His, p.Thr76Ala, p.Met181Thr, p.Val183Ile, p.Ala223Thr, p.Ser264Asn	−282,629	0.0498	NA
RNASEH2C	1	p.Arg145Leu	−86,985	0.0389	0.0398
GFBR2	8	p.Ser46Arg, p.Val191Ile, p.Arg193Trp, p.Thr206Met, p.Asp247Val, p.Arg313Gln, p.Thr315Met, p.Arg403His	−18,225	0.0447	NA
TRIP13	4	p.Val82Ile, p.Ile99Val, p.Gly173Glu, p.Ile196Val	−384,737	0.0238	0.0237
VTI1A	2	p.Asn40Asp, p.Leu104Pro	−475,838	0.0429	0.0424

[a]Conditional to the corresponding GWAS index SNP.

EFEMP2, rs146217902 in *EDEM1*, rs143563006 in *FBXO18*, and rs80358978 in *BRCA2*. Gene-based analyses showed an association at $P < 0.01$ for *EFEMP2* and *FBXO18* genes and at $P < 0.05$ for six genes, including *RNASEH2C, KLHL26*, *OR2A12*, *TGFBR2*, *TRIP13*, and *VTI1A*.

The most significant association was observed for a missense variant, rs80358978 (Gly2508Ser), in the *BRCA2* gene. It was 42 kb upstream from the GWAS SNP rs11571833. This variant was observed in 16 heterozygous breast cancer cases and only one control participant. This variant was not present among the 1,092 individuals included in the 1,000 Genomes Project or the 6,400 individuals of European or African ancestry included in the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (24). This variant was found in four Asian breast cancer women in the Breast Cancer Information Core (25). Though the clinical importance of this variant was unknown, it may be potentially functional and it is predicted to be "probably damaging" based on its Polyphen-2 score (0.999) and "deleterious" based on its sorting intolerant from tolerant (SIFT) score (0).

In addition, of the 38 rare variants causing missense mutations, the predominantly single amino acid change is from a basic or acidic amino acid to a neutral amino acid, such is the case for *CCDC88C*, *MAP3K1*, and *SRRM5* genes are predicted to be deleterious based on the SIFT score (Table 2). It further suggests, to some extent, that these rare missense mutations would affect the protein's topologic structure and physicochemical properties.

For gene-based analysis, our results indicated that the significant association with breast cancer risk is driven by one single variant in each gene. The reason is greatly related to our focuses on the missense/nonsense variants with low frequency. Generally speaking, the consequence of missense mutations has direct impacts on protein structure and function. Thus, it is more likely to undergo purifying selection (26, 27), making the probability of two or more rare missense mutations happening in the same gene quite low.

The most significant result from gene-based analyses is for the association observed with the *EFEMP2* gene, encoding a protein containing four EGF2 domains and six calcium-binding EGF2 domains. This gene is necessary for elastic fiber formation and connective tissue development (28). Several studies indicated that the expression level of the *EFEMP2* gene, even at an early cancer stage, was increased in cancer tissues of the patients with colorectal and endometrial cancer (29–31). *RNASEH2C*, another gene located at the 11q13.1 locus, also showed a significant association in this study. This gene encodes one of Ribonuclease H2 (RNase H2) subunits, a major nuclear enzyme involved in the degradation of RNA/DNA hybrids and removal of ribonucleotides misincorporated in genomic DNA to maintain genomic integrity. Mutations in each of the three RNase H2 genes have been implicated in a human autoinflammatory disorder, Aicardi–Goutières syndrome (32, 33). Crystal structure of the RNase H2 complex indicated that residues in the C-terminal kinked helix (RNASEH2C:143–160) contact both RNASEH2A and RNASEH2B (34), suggesting the detected variant (R145L) in the *RNASEH2C* gene may influence the complex formation of RNase H2.

FBXO18 (also called FBH1 or FBX18) is a member of the UvrD family of DNA helicases (35, 36). Its helicase activity induces DNA double-strand breakage and activation of ATM and DNA-PK and phosphorylation of RPA2 and p53 (37). The *ATM* and *p53* genes are two of the most well-established breast cancer susceptibility genes. A previous study has revealed a connection between rare missense variants in the *ATM* gene and breast cancer risk (11). Here,

we provide evidence that rare variants in the *FBXO18* gene may also contribute to the risk of breast cancer.

It has been well established that the TGF-β pathway plays a critical role in the development and progression of a large number of human cancers, including breast cancer (38–40). TGF-β1 is the most abundant form of TGF-β and regulates cellular processes by binding to TGFBR2. Therefore, defective expression of *TGFBR2* may play a significant role in carcinogenesis. Our previous evaluation of the associations of genetic variants in the TGF-β signaling pathway with breast cancer risk found that one common SNP (rs1078985) in the *TGFBR2* was associated with breast cancer risk (41). The gene-based results in this study provide further evidence that the *TGFBR2* gene is significantly associated with breast cancer risk.

In the present study, we identified multiple rare coding variants associated with breast cancer in GWA-identified loci. However, after adjusting for multiple comparisons, some of them became insignificant. The statistical power in the present study is limited for rare variants, even though more than 6,000 cases and controls were included. Independent studies with larger sample size are warranted to clarify the relationship between this rare variants and breast cancer risk.

In conclusion, we identified associations of additional genes/variants flanking the known susceptibility loci with breast cancer risk. These findings may provide new insights into the etiology of breast cancer as well as future potential therapeutic targets.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agents.

Authors' Contributions

Conception and design: W. Lu, X.-O. Shu, C. Li, B. Li, W. Zheng
Development of methodology: Y. Zhang, C. Li, B. Li
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): X.-O. Shu, Q. Cai, Y. Zheng, Y.-T. Gao, W. Zheng
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y. Zhang, C. Li, B. Li, W. Zheng
Writing, review, and/or revision of the manuscript: Y. Zhang, J. Long, X.-O. Shu, Q. Cai, C. Li, B. Li, Y.-T. Gao, W. Zheng
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Y. Zheng, Y.-T. Gao, W. Zheng
Study supervision: J. Long, W. Lu, X.-O. Shu, Y.-T. Gao, W. Zheng

Acknowledgments

The authors thank the study participants and research staff for their contributions and support to this project, Regina Courtney and Jie Wu for DNA preparation, Jing He for data processing and analyses, and Samantha Stansel for assistance in the preparation of this article. Sample preparation was conducted at the Survey and Biospecimen Shared Resources, which are supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485).

Grant Support

This work was supported in part by U.S. NIH grants R01CA158473 (to W. Zheng), R01CA148667 (to W. Zheng and J. Long), R01CA137013 (to J. Long), R37CA70867 (to W. Zheng), and R01CA124558 (to W. Zheng).

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.

Received October 8, 2013; revised December 5, 2013; accepted January 7, 2014; published OnlineFirst January 27, 2014.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin 2012;62:10–29.
2. Cariaso M, Lennon G. SNPedia: a Wiki supporting personal genome annotation, interpretation and analysis. Nucleic Acids Res 2012;40:D1308–D1312.
3. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013;45:353–61.
4. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.
5. Bonnefond A, Clement N, Fawcett K, Yengo L, Vaillant E, Guillaume JL, et al. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet 2012;44:297–301.
6. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 2009;324:387–9.
7. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, et al. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet 2010;42:684–7.
8. McInerney NM, Miller N, Rowan A, Colleran G, Barclay E, Curran C, et al. Evaluation of variants in the CHEK2, BRIP1 and PALB2 genes in an Irish breast cancer cohort. Breast Cancer Res Treat 2010;121:203–10.
9. Ruark E, Snape K, Humburg P, Loveday C, Bajrami I, Brough R, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. Nature 2013;493:406–10.
10. Lu W, Wang X, Lin H, Lindor N, Couch F. Mutation screening of RAD51C in high-risk breast and ovarian cancer families. Familial Cancer 2012;11:381–5.
11. Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, et al. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. Am J Hum Genet 2009;85:427–46.
12. Le Calvez-Kelm F, Lesueur F, Damiola F, Vallee M, Voegele C, Babikyan D, et al. Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. Breast Cancer Res 2011;13:R6.
13. Dowty J, Lose F, Jenkins M, Chang JH, Chen X, Beesley J, et al. The RAD51D E233G variant and breast cancer risk: population-based and clinic-based family studies of Australian women. Breast Cancer Res Treat 2008;112:35–9.
14. Gao YT, Shu XO, Dai Q, Potter JD, Brinton LA, Wen W, et al. Association of menstrual and reproductive factors with breast cancer risk: results from the Shanghai breast cancer study. Int J Cancer 2000;87:295–300.
15. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 2009;41:324–8.
16. Long J, Cai Q, Sung H, Shi J, Zhang B, Choi JY, et al. Genome-wide association study in East Asians identifies novel susceptibility loci for breast cancer. PLoS Genet 2012;8:e1002532.

17. Exomechip design [Internet]. Michigan: University of Michigan; 2013 [cited 2013 Mar 19]. Available from: http://genome.sph.umich.edu/wiki/Exome_Chip_Design.
18. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
19. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.
20. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
21. EPACTS [Internet]. Michigan: University of Michigan; 2013 [cited 2013 Mar 19]. Available from: http://genome.sph.umich.edu/wiki/EPACTS.
22. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 2012;91:224–37.
23. Wu M, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 2011;89:82–93.
24. Exome Variant Server [Internet]. Washington: University of Washington; 2013 [cited 2013 Jul 11]. Available from: http://evs.gs.washington.edu/EVS.
25. Breast Cancer Information Core [Internet]. Bethesda: National Institutes of Health; 2013 [cited 2013 Aug 18]. Available from: http://research.nhgri.nih.gov/bic/.
26. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science 2012;335:823–8.
27. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 2012;337:64–9.
28. Dasouki M, Markova D, Garola R, Sasaki T, Charbonneau NL, Sakai LY, et al. Compound heterozygous mutations in fibulin-4 causing neonatal lethal pulmonary artery occlusion, aortic aneurysm, arachnodactyly, and mild cutis laxa. Am J Med Genet 2007;143A:2635–41.
29. Yao L, Lao W, Zhang Y, Tang X, Hu X, He C, et al. Identification of EFEMP2 as a serum biomarker for the early detection of colorectal cancer with lectin affinity capture assisted secretome analysis of cultured fresh tissues. J Proteome Res 2012;11:3281–94.
30. Colas E, Perez C, Cabrera S, Pedrola N, Monge M, Castellvi J, et al. Molecular markers of endometrial carcinoma detected in uterine aspirates. Int J Cancer 2011;129:2435–44.
31. Yuen HF, McCrudden CM, Huang YH, Tham JM, Zhang X, Zeng Q, et al. TAZ expression as a prognostic indicator in colorectal cancer. PLoS ONE 2013;8:e54211.
32. Vogt J, Agrawal S, Ibrahim Z, Southwood TR, Philip S, MacPherson L, et al. Striking intrafamilial phenotypic variability in Aicardi–Goutieres syndrome associated with the recurrent Asian founder mutation in RNASEH2C. Am J Med Genet 2013;161:338–42.
33. Rice G, Patrick T, Parmar R, Taylor CF, Aeby A, Aicardi J, et al. Clinical and molecular phenotype of Aicardi-Goutieres syndrome. Am J Hum Genet 2007;81:713–25.
34. Reijns MAM, Bubeck D, Gibson LCD, Graham SC, Baillie GS, Jones EY, et al. The structure of the human rnase h2 complex defines key interaction interfaces relevant to enzyme function and human disease. J Biol Chem 2011;286:10530–9.
35. Kim J-H, Kim J, Kim D-H, Ryu G-H, Bae S-H, Seo Y-S. SCFhFBH1 can act as helicase and E3 ubiquitin ligase. Nucleic Acids Res 2004; 32:2287–97.
36. Kim J, Kim JH, Lee SH, Kim DH, Kang HY, Bae SH, et al. The novel human DNA helicase hFBH1 Is an F-box protein. J Biol Chem 2002; 277:24530–7.
37. Jeong YT, Rossi M, Cermak L, Saraf A, Florens L, Washburn MP, et al. FBH1 promotes DNA double-strand breakage and apoptosis in response to DNA replication stress. J Cell Biol 2013;200:141–9.
38. Benson JR. Role of transforming growth factor beta in breast carcinogenesis. Lancet Oncol 2004;5:229–39.
39. Elliott RL, Blobe GC. Role of transforming growth factor beta in human cancer. J Clin Oncol 2005;23:2078–93.
40. Bierie B, Moses HL. Tumour microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer. Nat Rev Cancer 2006;6: 506–20.
41. Ma X, Beeghly-Fadiel A, Lu W, Shi J, Xiang YB, Cai Q, et al. Pathway analyses identify TGFBR2 as potential breast cancer susceptibility gene: results from a consortium study among Asians. Cancer Epidemiol Biomarkers Prev. 2012;21:1176–84.