Common genetic variants associated with breast cancer in Korean women and differential susceptibility according to intrinsic subtype

Wonshik Han¹ ², Jung Hoon Woo³, Jong-Han Yu⁴, Min-Ju Lee¹, Hyeong-Gon Moon², Dahee Kang⁵, Dong-Young Noh¹ ²

¹Cancer Research Institute, ²Department of Surgery, and ³Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea; ⁴Genference Inc, Seoul, Korea; ⁵Department of Surgery, Asan Medical Center, University of Ulsan College of Medicine, Seoul

Running Title: Common genetic variants and breast cancer subtypes

Correspondence to:
Dong-Young Noh
Department of Surgery, Seoul National University College of Medicine,
28 Yongon-dong, Chongno-gu, Seoul 110-744, Korea
Tel: 82-2-2072-2921, Fax: 82-2-766-3975, Email: dynoh@plaza.snu.ac.kr
ABSTRACT

Background: Recently identified genetic variants from genome-wide association studies on breast cancer have not been validated in Asian populations, except in China. In this study, we sought to confirm the association between known variants and breast cancer in Korean women and further evaluate the associations of individual SNPs with different intrinsic subtypes of breast cancer.

Methods: In total, 3321 invasive breast cancer patients and 3500 healthy controls were genotyped for 5 SNPs using the TaqMan assay. The SNPs genotyped included rs2046210 (6q25.1), rs2981582 (FGFR2), rs889312 (MAP3K1), rs3803662 (TOX3/TNRC9), and rs4973768 (SLC4A7). Tumors were classified into four intrinsic subtypes based on estrogen receptor, progesterone receptor, HER2, and Ki67 expression.

Results: All five SNPs were significantly associated with risk of breast cancer in dominant, recessive, and additive models. Odds ratios (ORs) per risk allele (95% confidence interval) were 1.29 (1.16-1.43), 1.40 (1.18-1.68), 1.22 (1.06-1.41), 1.52 (1.30-1.77), and 1.20 (1.08-1.33) for rs2046210, rs2981582, rs889312, rs3803662, and rs4973768, respectively. A multi-gene logistic regression risk model was generated with the SNPs. In subtype analysis, all 5 SNPs were associated with the Luminal A subtype. Two SNPs (rs2046210 and rs3803662) were linked to the ER-HER2+ subtype, and only rs2046210 SNP was associated with the Triple-negative subtype.

Conclusions: The 5 SNPs from genome-wide association studies were significantly associated with breast cancer risk in Korean women. Associations were heterogeneous according to the intrinsic subtype of breast cancer.

Impact: Our result is an important contribution to the literature about genetic susceptibility for breast cancer in non-white populations.

Keywords: Breast cancer, genetic variant, polymorphism, intrinsic subtype, Korean
Introduction

Breast cancer is one of the most common malignancies affecting women worldwide. The recent increase in incidence has made breast cancer one of the most frequently recorded diseases among Korean women since 2001 (1). The age-adjusted death rate due to breast cancer in Korea is also rising, with the most rapid increase in the world from 1985 to 1995 (2).

Genetic factors play an important role in breast cancer development. After completion of the human genome project, single nucleotide polymorphisms (SNP) were highlighted as the key variations leading to genetic differences in breast cancer susceptibility between individuals. However, candidate gene approaches have not been successful in reproducibly identifying the significant SNPs associated with breast cancer. Recent genome-wide association studies (GWAS) (3-8) have led to the detection of multiple and robust genetic susceptibility loci for breast cancer. These are common in the general population, but the relative risk for breast cancer conferred by each locus is low. All documented GWAS to date have been conducted on women of European ancestry, except for one Chinese study. Since minor allele frequencies of SNPs are highly variable and the linkage disequilibrium (LD) patterns differ between ethnicities, it is important to confirm the effects of the variants identified in GWAS in different ethnic populations.

Breast cancer is a heterogeneous disease in which risk factors are suggested to be differentially associated with the development of distinct tumor subtypes with variable biological and clinical features. Consistent with this theory, there is growing evidence that known breast cancer risk factors vary depending on hormone receptor status (9-11). Recent studies have shown that common genetic variants are differentially associated with ER+ or ER- breast cancer, providing further support for the hypothesis that ER+ and ER- diseases result from different etiologic pathways (6, 7, 12). However, limited information has been obtained on additional heterogeneity of genetic risk factors within ER+ or ER- tumors according to the intrinsic breast cancer subtype (13). Genomic studies have established that breast cancer can be divided into four major intrinsic subtypes (Luminal A, Luminal B,
HER2-enriched, basal-like) that differ significantly in terms of incidence, survival and response to therapy (14-16). The origins and developmental pathways of these subtypes may be subtype-dependent (17).

In the present study, we evaluated Korean breast cancer patients for the five SNPs recently identified from four European and one Chinese GWAS. The associations of all GWAS-identified SNPs with breast cancer were further assessed in relation to the four intrinsic subtypes defined using surrogate immunohistochemistry (IHC) markers.
Materials and Methods

Subjects and Samples
Consecutive patients with histologically confirmed primary breast cancer subjected to operative procedures between 2002 and 2009 in Seoul National University Hospital were included for analysis. Patients diagnosed with non-invasive breast cancer (ductal carcinoma in situ and lobular carcinoma in situ) were excluded. Peripheral venous blood samples were obtained and stored at the time of operation. DNA samples from controls were donated from the genome bank of Korea Centers for Disease Control & Prevention. Control subjects were randomly selected from a population-based cohort of 12,000 health examinees. The Korean Health Examinees Study (KHES), which was established in 2004, aims to carry out a follow-up every two years for each member. The subjects recruited in 14 hospitals and institutions were from a community adult population under 70 years of age, who had no experience of participating in cohort studies and consented to genetic analysis.

In total, 3321 breast cancer cases and 3500 healthy control women were genotyped. The mean ages of cases and controls were 48.5±9.8 (range: 21-82) and 52.6±8.1 (range: 23-79), respectively.

This study was approved by the Institutional Review Board for human research at Seoul National University Hospital. Informed consent was obtained from all patients.

SNP selection
Among the common breast cancer susceptibility alleles established from recent GWAS, we selected 5 SNPs for genotyping, specifically, rs2046210 (6q25.1)(8), rs2981582 (FGFR2)(3, 5), rs889312 (MAP3K1)(3), rs3803662 (TOX3/TNRC9)(3), and rs4973768 (SLC4A7)(18) (Table 1). SNPs that were previously reported as not significantly associated with breast cancer in Korean and other Asian subpopulations were excluded from analysis.
Genotyping method

Genomic DNA was extracted from peripheral blood leukocytes using a genomic DNA kit (Qiagen, Germany), according to the manufacturer's instructions. SNP genotyping was performed on an Applied Biosystems 7900HT real-time PCR System (Applied Biosystems, Foster City, CA). Assay reagents for each SNP were additionally obtained from Applied Biosystems. DNA was genotyped following the manufacturer's protocol. Briefly, the components for the genotyping reactions included: 2 \( \mu \)L of 10 ng/nL genomic DNA, 2.5 \( \mu \)L of TaqMan Genotyping Master Mix (Applied Biosystems, Foster City, CA), 0.125 \( \mu \)L of assay mix (40×), and water up to a total volume of 5 \( \mu \)L. The thermocycler conditions were as follows: 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 sec and 60°C for 60 sec. Reactions were analyzed using Applied Biosystems Sequence Detection Software (Version 2.3).

Intrinsic subtypes

Immunohistochemical (IHC) tests were performed to determine the tumor expression levels of estrogen receptor (ER), progesterone receptor (PR), HER2, and Ki67. The primary antibodies and dilution factors employed in this study have been described in a previous report (19). IHC tests were performed on paraffin blocks from tumor tissues fixed in 10% neutral-buffered formalin. All IHC tests were carried out in a single laboratory (Seoul National University Hospital) immediately after surgery. Slides were reviewed by a pathologist with no knowledge of the genotype results. We classified ER and PR expression status as positive at a cut-off value of 10% or more positively stained cells per 10 high-power fields. HER2 immunostaining was considered positive when strong (3+) membranous staining was observed in at least 10% of tumor cells, whereas cases classified as 0 to 1+ staining were regarded as negative. Cases displaying moderate membranous staining (2+) were confirmed using HER2 fluorescence in situ hybridization (FISH). At a Ki67 cut-off point of 10% or higher, established by Jung et al. (19), tumors were designated 'high proliferation'. The study population was subgrouped into four breast cancer subtypes defined by Hugh et al.(20),...
specifically, triple negative (ER-negative, PR-negative, HER2-negative), ER-HER2+ (ER-
negative, PR-negative, HER2-positive), Luminal B (ER-positive and/or PR-positive and
either HER2-positive and/or Ki67\textsuperscript{high}), and Luminal A (ER-positive and/or PR-positive and
not HER2-positive or Ki67\textsuperscript{high}).

**Statistical analysis**

The chi-square test for genotype distribution was conducted to evaluate deviation from the
Hardy-Weinberg equilibrium in each case and control group. Breast cancer risk was estimated
as odds ratios (OR) and 95% confidence intervals (CI), based on unconditional logistic
regression adjusted for age (years). Analyses were performed assuming a dominant, recessive,
and additive allelic effect for each polymorphism. The likelihood ratio test was used to
examine the effect of each SNP at the 5% significance level.

Stratified analysis according to the 4 breast cancer subtypes was additionally conducted.
We calculated ORs and 95% CI of SNPs, based on the additive model that assumes a linear
increase in disease risk with increasing number of risk alleles.

For the multi-gene logistic regression risk model, we used the ‘glm’ function in statistical
language R (version 2.5.1).

**Results**

The distribution of all genetic polymorphisms did not deviate from Hardy-Weinberg
equilibrium ($P>0.30$). For rs889312 and rs3803662, the major allele was the risk allele,
consistent with the findings of the Chinese study (21) (Table 1).

All five breast cancer-associated SNPs identified in previous GWAS (rs2046210,
rs4973768, rs2981582, rs3803662, and rs889312) were significantly associated with breast
cancer risk in dominant, recessive, and additive models, except rs4973768 in the recessive
model (Table 2). OR values ranged from 1.13 (rs889312 in the dominant model) to 1.52 (rs3803662 in the additive model).

Next, we generated a multi-gene logistic regression risk model with the five SNPs. The variables and coefficients are presented in Table 3. The odds of breast cancer determined from this model varied from 0.43 for subjects homozygous (2 copies) for protective variants at all 5 markers \([\exp(-0.8334307) = 0.43]\) to 2.36 for subjects homozygous for risk variants at all markers \([\exp(-0.8334307 + 0.3771801 + 0.3679599 + 0.2210641 + 0.4376367 + 0.2914944) = 2.36]\).

According to the intrinsic subtype classification system specified in Materials and Methods, 1685 breast cancer cases were subgrouped as Luminal A, 650 as Luminal B, 310 as ER-HER2+, and 574 as triple-negative subtype. In 112 cases, the subtype could not be determined due to the absence of one or more individual marker data. As shown in Figure 1, all five SNPs were significantly associated with the Luminal A subtype, and four out of five SNPs with the Luminal B subtype. Three SNPs, rs2981582 (FGFR2), rs889312 (MAP3K1), and rs4973768 (SLC4A7) showed stronger associations with ER+ than ER- tumors. The most remarkable pattern of subtype association was observed with the SNP in 6q25.1 (rs2046210). ORs of this SNP were higher in the Luminal B, ER-HER2+, and triple-negative subtypes, compared to the Luminal A subtype. Notably, this was the only significant SNP associated with the triple-negative subtype. On the other hand, the SNP in TOX3/TNRC9 (rs3803662) was significantly linked to the ER-HER2+ but not triple-negative subtype of breast cancer.

**Discussion**

In this large case-control study involving 6821 participants, we showed that the five SNPs identified in GWAS conducted on women of mainly European ancestry were also significantly associated with breast cancer susceptibility in Korean women. This is the second Asian study (21, 22) to validate earlier GWAS results in a large population. The information obtained would be useful for classifying Korean women into the relevant genetic risk groups.
This type of ethnicity-specific study is important, because allele frequencies of SNPs vary among populations and a positive SNP association with breast cancer in individuals of one ethnicity may be negative in another ethnic population. For example, CASP8 D302H, which was found to be significantly associated with breast cancer in a large population of Western women (23), was not polymorphic in Korean women (24). Moreover, SNPs in 2q35 (rs13387042)(6), 5p12 (rs10941679)(7), 6q22.33 (rs2180341)(4), and 8q24.21 (rs13281615)(3) identified from GWAS were not associated with breast cancer in Chinese women (21, 22).

We created a multi-gene logistic regression risk model with the five SNPs. The odds ratios according to percentile (data not shown) were similar to those reported by Pharoah et al. (25) calculated using data from 7 SNPs. Based on their results, Pharoah et al. recommended that women with higher genetic risk should be subjected to annual screening mammography at a younger age (25). Zheng and colleagues (22) proposed the use of a risk assessment model including both genetic markers and clinical predictors. However, Wacholder and co-workers showed that genetic factors only modestly improved the performance of risk models for breast cancer (26). Thus, there is an urgent need to identify more SNPs to generate a more useful breast cancer risk model composed of common genetic variants (27, 28). Other SNPs from published GWAS should be validated, and large-scale GWAS studies on Korean women are critical.

In our study, rs2981582 (FGFR2) and rs889312 (MAP3K1) showed stronger associations with ER+ than ER- tumors, consistent with data from European studies on susceptibility loci and ER status (6, 7, 12). The SLC4A7 (rs4973768) variant was also more strongly associated with ER+ than ER- tumors, consistent with the Chinese study (21). A large-scale study from the Breast Cancer Association Consortium (BCAC) reported that the TNRC9/TOX3 (rs3803662) variant was associated with both ER+ and ER- subtypes, although the association was slightly weaker for ER- tumors (12). We observed an association of this SNP with ER+ as well as the ER-HER2+ subtypes, but not the triple-negative subtype. This further heterogeneity of genetic association within ER- or ER+ tumors is a novel finding, and
requires further validation in another cohort. Earlier Chinese studies reported a closer association between the variant in rs2046210 (6q25.1) with ER- than ER+ tumors (8, 21), which was re-confirmed in our investigation. Furthermore, among the ER+ tumors, this SNP was associated with the Luminal B subtype as strongly as with ER- tumors. These findings collectively imply that the associations between SNP and breast cancer are not simply determined on the basis of ER status, but vary according to the intrinsic subtype. Our results are significant, since they support the theory that tumors with different intrinsic subtypes result from different etiologic pathways that have evolved from cancer stem cells (17).

In summary, we have shown that the five selected variants identified from earlier GWAS performed on women of European or Chinese origin are similarly associated with breast cancer in the Korean population. Although the currently identified loci provide low discriminatory accuracy to distinguish between the low- and high-risk groups, combinations of loci established in future studies may be useful in population screening strategies. The finding that SNPs are differentially associated with individual intrinsic subtypes of breast cancer is consistent with the hypothesis that the genetic mechanism of etiology differs between subtypes.

**Acknowledgments**

This work was supported by a grant from the Seoul National University Hospital Research Fund (09-2006-006-0) and the Korea Health 21 R&D Project, Ministry of Health & Welfare, R.O.K (01-PJ3-PG6-01GN07-0004 and A050558).
References

Table 1. Genotype frequencies of each SNP among control and case

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome and Genes</th>
<th>Genotype</th>
<th>Control N (%)</th>
<th>Case N (%)</th>
<th>HWEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2046210</td>
<td>6q25.1 (upstream of ESR1)</td>
<td>GG</td>
<td>1586 (45.4)</td>
<td>1260 (38.8)</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AG</td>
<td>1531 (43.8)</td>
<td>1565 (48.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>376 (10.8)</td>
<td>426 (13.1)</td>
<td></td>
</tr>
<tr>
<td>rs2981582</td>
<td>10q26.13 (FGFR2)</td>
<td>GG</td>
<td>1751 (50.2)</td>
<td>1497 (46.3)</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AG</td>
<td>1457 (41.8)</td>
<td>1393 (43.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>281 (8.1)</td>
<td>342 (10.6)</td>
<td></td>
</tr>
<tr>
<td>rs889312</td>
<td>5q11.2 (MAP3K1)</td>
<td>CC</td>
<td>1011 (28.9)</td>
<td>1048 (31.8)</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA</td>
<td>1743 (49.8)</td>
<td>1628 (49.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>743 (21.2)</td>
<td>620 (18.8)</td>
<td></td>
</tr>
<tr>
<td>rs3803662</td>
<td>16q12.1 (TOX3/TNRC9)</td>
<td>AA</td>
<td>1361 (39.0)</td>
<td>1481 (45.1)</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AG</td>
<td>1617 (46.3)</td>
<td>1435 (43.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG</td>
<td>516 (14.8)</td>
<td>369 (11.2)</td>
<td></td>
</tr>
<tr>
<td>rs4973768</td>
<td>3p24 (SLC4A7, NEK10)</td>
<td>CC</td>
<td>2195 (62.8)</td>
<td>1913 (57.9)</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CT</td>
<td>1143 (32.7)</td>
<td>1213 (36.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TT</td>
<td>159 (4.5)</td>
<td>177 (5.4)</td>
<td></td>
</tr>
</tbody>
</table>

HWEP: Hardy-Weinberg equilibrium p value
Table 2. Associations between the five polymorphisms and breast cancer risk

<table>
<thead>
<tr>
<th></th>
<th>Dominant model</th>
<th>Recessive model</th>
<th>Additive model</th>
<th>Homozygote wild-type vs. Homozygote variant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>p</td>
<td>OR (95% CI)</td>
<td>p</td>
</tr>
<tr>
<td>rs2046210</td>
<td>1.31 (1.19-1.45)</td>
<td>7.91 x 10^{-8}</td>
<td>1.25 (1.07-1.45)</td>
<td>4.57 x 10^{-5}</td>
</tr>
<tr>
<td>rs2981582</td>
<td>1.15 (1.04-1.27)</td>
<td>4.70 x 10^{-3}</td>
<td>1.34 (1.13-1.59)</td>
<td>7.88 x 10^{-4}</td>
</tr>
<tr>
<td>rs889312</td>
<td>1.13 (1.02-1.26)</td>
<td>2.25 x 10^{-2}</td>
<td>1.15 (1.02-1.31)</td>
<td>2.14 x 10^{-2}</td>
</tr>
<tr>
<td>rs3803662</td>
<td>1.27 (1.15-1.41)</td>
<td>1.98 x 10^{-6}</td>
<td>1.37 (1.18-1.59)</td>
<td>2.44 x 10^{-5}</td>
</tr>
<tr>
<td>rs4973768</td>
<td>1.20 (1.09-1.33)</td>
<td>2.76 x 10^{-4}</td>
<td>1.19 (0.94-1.49)</td>
<td>1.33 x 10^{-4}</td>
</tr>
</tbody>
</table>

*Adjusted for age
Table 3. Variables and coefficients of the multi-gene logistic regression risk model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.8334307</td>
<td>0.09603210</td>
</tr>
<tr>
<td>rs2046210 (AG)</td>
<td>0.2688592</td>
<td>0.05282297</td>
</tr>
<tr>
<td>rs2046210 (AA)</td>
<td>0.3771801</td>
<td>0.08115160</td>
</tr>
<tr>
<td>rs2981582 (AG)</td>
<td>0.1247290</td>
<td>0.05203057</td>
</tr>
<tr>
<td>rs2981582 (AA)</td>
<td>0.3679599</td>
<td>0.08905035</td>
</tr>
<tr>
<td>rs889312 (CA)</td>
<td>0.1178304</td>
<td>0.06562798</td>
</tr>
<tr>
<td>rs889312 (CC)</td>
<td>0.2210641</td>
<td>0.07132804</td>
</tr>
<tr>
<td>rs3803662 (AG)</td>
<td>0.2194539</td>
<td>0.07866000</td>
</tr>
<tr>
<td>rs3803662 (AA)</td>
<td>0.4376367</td>
<td>0.07915529</td>
</tr>
<tr>
<td>rs4973768 (CT)</td>
<td>0.1971477</td>
<td>0.05266590</td>
</tr>
<tr>
<td>rs4973768 (TT)</td>
<td>0.2914944</td>
<td>0.11566619</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1. SNP associations are breast cancer subtype-specific.

X-axis: Odds Ratio adjusted for age.

Horizontal bars showing 95% confidence interval.
Common genetic variants associated with breast cancer in Korean women and differential susceptibility according to intrinsic subtype

Wonshik Han, Jung Hoon Woo, Jong-Han Yu, et al.

Cancer Epidemiol Biomarkers Prev  Published OnlineFirst March 17, 2011.

Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-10-1282

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.