Polymorphisms and haplotypes of the NBS1 gene are associated with risk of sporadic breast cancer in non-Hispanic white women ≤55 years

Jiachun Lu, Qingyi Wei, Melissa L. Bondy, Donghui Li1, Abenaa Brewster2, Sanjay Shete, Tse-Kuan Yu1, Aysegul Sahin1, Funda Meric-Bernstam1, Kelly K. Hunt5, S. Eva Singletary5, Merrick I. Ross5 and Li-E Wang6

Department of Epidemiology, 1Department of Gastrointestinal Medical Oncology, 2Department of Clinical Cancer Prevention, 3Department of Radiation Oncology, 4Department of Pathology and 5Department of Surgical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA

To whom requests for reprints should be addressed: Department of Epidemiology, Unit 1365, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA.
Tel: +1 713 792 3020; Fax: +1 713 563 0999; Email: lwang@mdanderson.org

DNA double-strand breaks (DSBs) may lead to genomic instability and cancer if unrepaired. Nijmegen breakage syndrome 1 (NBS1) protein is one of the key proteins that participates in recognition and repair of DSBs in humans. We hypothesized that polymorphisms of NBS1 are associated with breast cancer risk. We selected three NBS1 haplotype-tagging polymorphisms (i.e. 924T>C, 8360G>C and 30537G>C) to represent all common (>5%) haplotypes reported in the National Institute of Environmental Health Sciences database and to reconstruct haplotypes. In a hospital-based case–control study of 421 non-Hispanic white patients with sporadic breast cancer (≤55 years) and 423 cancer-free controls who were frequency-matched with the cases by age (≥5 years and ≤55), we tested our hypothesis and found that compared with 924TT homozygotes the variant homozygote 924CC carriers had a 4.55-fold increased risk of breast cancer [95% confidence interval (CI) = 1.51–13.7] and that compared with the 8360GG genotype the variant genotypes were also associated with a significantly increased risk [adjusted odds ratio (OR) = 1.33, 95% CI = 1.00–1.78 for 8360CG; adjusted OR = 1.83, 95% CI = 1.14–2.94 for 8360CC]. However, these effects were not observed for the 30537G>C polymorphism. Furthermore, the derived haplotypes were associated with risk in a dose–response manner as the number of variant (risk) alleles (i.e. 8360C, 924C or 30537C) increased (adjusted OR = 1.07, 95% CI = 0.78–1.46 for 1–2 variant alleles; adjusted OR = 2.47, 95% CI = 1.48–4.14 for 3–6 variant alleles; *P* value = 0.006). These findings suggest that NBS1 polymorphisms and haplotypes may contribute to the etiology of sporadic breast cancer in young non-Hispanic white women. Large studies are warranted to confirm these findings.

Abbreviations: BPDE, benzo[a]pyrene diol epoxide; CI, confidence interval; DSB, double-strand break; hSNP, haplotype-tagging SNP; LD, linkage disequilibrium; NBS1, Nijmegen breakage syndrome 1; OR, odds ratio; PCR, polymerase chain reaction; SNP, single nucleotide polymorphism.

© The Author 2006. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oupjournals.org

Introduction

Breast cancer is the leading incident cancer among women in the United States, and the incidence rate has increased since the 1980s, with an estimated 212 920 new cases and 40 970 deaths in 2006 (1). Epidemiological studies have established many etiologic factors, including some environmental factors, such as ionizing radiation, high-fat diets, alcohol consumption and use of hormone replacement or oral contraceptives (2). These environmental factors can potentially cause DNA damage and thus lead to a higher risk of breast cancer. However, only a small fraction of exposed individuals develops breast cancer, suggesting that genetic susceptibility plays a role in an individual’s risk of developing breast cancer. Therefore, it is important to identify the at-risk populations so that they may be targeted for prevention and early detection.

One of the genetic factors for breast cancer development is a family history of the disease. Germline mutations in high-penetrance breast cancer susceptibility genes, such as BRCA1, BRCA2, p53, ATM and PTEN, have been found to account for the majority of familial hereditary breast cancers (2,3). These genes are all more or less involved in DNA repair; however, because of their low allele frequencies, they only account for 5–10% of all breast cancer cases (2,3). For most sporadic breast cancer cases, genetic susceptibility is still poorly understood. Low-penetration but highly prevalent susceptibility genes, acting together with exogenous and endogenous risk factors, are likely to account for many sporadic breast cancers (4).

It has been suggested that DNA repair capacity phenotype may be markers for breast cancer susceptibility (5,6). We have published two pilot studies, in which we found that breast cancer risks in young women (<55 years) were significantly associated with a reduced capacity for repairing DNA damage induced by the tobacco carcinogen benzo[a]pyrene diol epoxide (BPDE) (7) and a higher level of BPDE-induced chromatid breaks in peripheral blood lymphocytes (8). Ionizing radiation, one of the etiologic contributors to breast cancer, can cause DNA damage that is characteristic of DNA double-strand breaks (DSBs). If not repaired, DSBs may cause chromosomal breaks and genomic instability, thus increasing one’s probability of developing cancer.

In humans, there are two pathways that repair DSBs, the homologous recombination (HR) and non-homologous end-joining (NHEJ), in which the initial step is to recognize DSBs by a protein complex (MRE11/RAD50/NBS1 complex, MRN). Nijmegen breakage syndrome 1 (NBS1) is the key regulator of this protein complex (9–11). If DSBs persist, NBS1 promotes the enzymatic activity of RAD50 and MRE11 and directs the MRN complexes to the site of DNA damage. NBS1 also interacts with γ-H2ax to form nuclear foci and activates the ATM kinase that may activate...
cell-cycle checkpoint response to initiate the repair of DNA damage and maintain genomic stability (12–15). A marked impaired DSBR repair has been observed in cells from NBS patients (16). An NBS1 null mutant murine model showed that the mutants had increased chromosome damage, radiomimetic sensitivity and premature cell death (17). Human cell lines with NBS1 knockdown by small interfering RNA showed increased ionizing radiation-induced mutagenesis and telomere instability (18).

The NBS1 gene has been mapped to chromosome 8q21 (19–21), and it spans over 50 kb, contains 16 exons and encodes the 754-aa protein (Nibrin) (GenBank accession no. AF069291). Mutations in NBS1 cause NBS, a chromosome instability syndrome characterized by microcephaly, growth retardation, immunodeficiency, hypersensitivity to ionizing radiation and a predisposition to malignancy (22,23). The primary cause of deaths among NBS patients is cancer; 40% of NBS patients develop cancer before the age of 21 (23). A homozygous mutation, a 5 bp deletion in exon 6 (657del5), has been reported to be associated with an elevated risk of breast cancer in Polish and Russian populations (24,25). However, the 657del5 mutation appears to be of Slavic origin with a low frequency of ~0.5% in Eastern European populations and even rarer in other ethnic groups (26). Furthermore, several mutations and polymorphisms in NBS1 were identified in breast cancer and ovarian cancer cell lines (27). On the basis of these studies, it is likely that common polymorphisms of NBS1 may contribute to the risk of sporadic breast cancer.

Several studies have investigated the Glu185Gln polymorphism in exon 5 of the NBS1 gene (28–30). One study investigated four single nucleotide polymorphisms (SNPs) (i.e. Leu34Leu in exon 2, Asp399Asp in exon 10, Pro672Pro in exon 13 and Glu185Gln) simultaneously (31) in breast cancer, and the results suggested that these polymorphisms were not associated with risk of breast cancer. However, numerous polymorphic loci that exist in the exons of NBS1 are in complete linkage disequilibrium (LD), although not necessarily functional, and a few SNPs may not be representative of genetic variation in the gene. A more holistic approach using multiple, unlinked genetic variants is needed for such association studies. The use of haplotype-tagging SNPs (htSNPs) in association studies for identification of common disease-associated genes is not only hypothesis-free in selecting the SNPs and greatly reduces genotyping efforts but also appears to have a greater power over single-allele studies (32–34). Therefore, we hypothesized that polymorphisms and haplotypes of the NBS1 gene are associated with risk of sporadic breast cancer, particularly in young women, and tested this hypothesis in a case-control study that was built on our previous studies of young women (≤55 years old) (7,8) with three representative htSNPs based on prior pair-wise LD analysis.

Materials and methods

Study population

From August 1998 to August 2005, histopathologically confirmed, untreated primary female breast cancer patients who registered at the Nellig B. Connally Breast Center at The University of Texas M. D. Anderson Cancer Center were recruited. To facilitate the study of genetic susceptibility in a homogeneous population, the case subjects were limited to young women ≤55 years old and non-Hispanic whites only. We recruited 421 breast cancer cases, consisting of 75 cases of ductal or lobular carcinomas in situ and 346 cases of invasive carcinomas. According to the American Joint Committee on Cancer (AJCC) staging classifications (35), there were 75 cases of stage 0, 159 cases of stage I, 98 cases of stage II A, 43 cases of stage II B, 22 cases of stage III A, 8 cases of stage III B, 12 cases of stage III C and 4 cases of stage IV. During the same time period, we randomly recruited 423 cancer-free, female controls that were frequency-matched to the cases by age (±5 years) and smoking status (never versus ever) among hospital visitors who were either accompanying the patients but genetically unrelated with the cases included in this study or came to the hospital for our free skin cancer prevention screening and were found cancer-free. A simple questionnaire was used to collect information on demographic data, smoking and alcohol use for all cases and controls. Participants who had smoked at least 100 cigarettes in their lifetime were categorized as ever smokers and the rest were categorized as never smokers. Participants who had drunk alcoholic beverages at least once a week for 1 year or more in their lifetime were categorized as ever drinkers, and the rest were categorized as never drinkers. Additional information on the analysis of the NIEHS EGP-SNP database, we calculated the 5% (39) haplotypes that account for 44.4% of the 54 haplotypes. Three htSNPs were selected and they are located in the high LD block (Figure 1) that has the mostly studied nsSNP, with the FORCE IN command of Stram's tagSNPs program we replaced 626G/C, 8360G/A with 0.80 to select a minimal set of representative tagging SNPs (htSNPs) from the 84 SNPs included in the common [i.e. the frequency ≥5%] (39) haplotypes that account for 44.4% of the 54 haplotypes. Three htSNPs were selected and they are located in the high LD block (Figure 1) that has .

On the basis of the genotype data of 90 subjects in the NIEHS SNP database, we calculated the power and correlation coefficient r² with the 2LD program (40) and Statistical Analysis System (SAS) Genetics software (Version 9.1, SAS Institute, Cary, NC) between any pair of these candidate SNPs and found that locus 626G>A was in complete LD with 8360G>C and 40419A>G (D² = 1.00, P = 0.00 and r² = 1.00 for both) and locus 8360G>C was in complete LD with 40419A>G. As a result, genotyping for locus 8360G>C only would be sufficient for these three variants. We then selected five htSNPs with 924T>C, 8360G>C, 30537G>C, 626G>A and 31129>G>T and performed a pilot study to genotype 118 non-Hispanic white cancer-free controls to further confirm the LD of these five loci. We found that the SNP 3816G>A was in complete LD with 8360G>C (D² = 1.00, P = 0.00 and r² = 1.00), and the SNP 31129>G>T was in incomplete LD with 8360G>C (D² = 0.91, P = 0.00 and r² = 0.67). As a result, the SNPs 3816G>A and 31129>G>T were eliminated from further genotyping. Because 8360G>C is the mostly studied nsSNP, with the FORCE IN command of Stram’s tagSNPs program we replaced 626G>A for 8360G>C, ensuring the P = 1.00. Selection of htSNPs

As shown in the GenBank SNP database (http://www.ncbi.nlm.nih.gov), there are >317 SNPs that have been identified in the NBS1 gene, of which 249 SNPs are also reported in the Environmental Genome Project (EGP) SNP database of the National Institute of Environmental Health Sciences (NIEHS EGP; http://egp.gs.washington.edu). On the basis of the available resequencing data, only 84 SNPs have a minor allele frequency ≥5% in 90 tested individuals, for which 54 sets of haplotypes were reconstructed. A large block of high LD diversity with limited haplotype diversity was marked with a triangle diagram. We used the BEST 1.0 program (36) and Stram’s tagSNPs program (37,38) with the squared correlation R² > 0.80 to select a minimal set of representative tagging SNPs (htSNPs) from the 84 SNPs included in the common [i.e. the frequency ≥5%] (39) haplotypes that account for 44.4% of the 54 haplotypes. Three htSNPs were selected and they are located in the high LD block (Figure 1) that has .

On the basis of the genotype data of 90 subjects in the NIEHS SNP database, we calculated the D² values and correlation coefficient r² with the 2LD program (40) and Statistical Analysis System (SAS) Genetics software (Version 9.1, SAS Institute, Cary, NC) between any pair of these candidate SNPs and found that locus 626G>A was in complete LD with 8360G>C and 40419A>G (D² = 1.00, P = 0.00 and r² = 1.00 for both) and locus 8360G>C was in complete LD with 40419A>G. As a result, genotyping for locus 8360G>C only would be sufficient for these three variants. We then selected five htSNPs with 924T>C, 8360G>C, 30537G>C, 626G>A and 31129>G>T and performed a pilot study to genotype 118 non-Hispanic white cancer-free controls to further confirm the LD of these five loci. We found that the SNP 3816G>A was in complete LD with 8360G>C (D² = 1.00, P = 0.00 and r² = 1.00), and the SNP 31129>G>T was in incomplete LD with 8360G>C (D² = 0.91, P = 0.00 and r² = 0.67). As a result, the SNPs 3816G>A and 31129>G>T were eliminated from further genotyping. Because 8360G>C is the mostly studied nsSNP, with the FORCE IN command of Stram’s tagSNPs program we replaced 626G>A for 8360G>C, ensuring the P = 1.00. Genotyping analysis

From each blood sample, the leukocyte cell pellet obtained from theuffy coat by centrifugation of 1 ml of whole blood was used for genomic DNA...
The NBS1 924T>C polymorphism was genotyped with the polymerase chain reaction–restriction fragment length polymorphism (PCR–RFLP) method by using the following self-designed primers to amplify the SNP-containing fragment: 5'-CGTCCAATTGTAAAGCCAGAA-3' (forward) and 5'-AGGCCTGTTTTACAATCTCTG-3' (reverse). PCR was performed in 10 μl reaction systems containing 2 mM MgCl₂, 0.04 mM deoxynucleotide triphosphates, 1.0 units of BclI (New England BioLabs) overnight, and the products were separated in 3% MetaPhor agarose gel. The 924TT genotype produced two bands (197, 172 and 25 bp), whereas the 924TG genotype produced a single DNA band (197 bp), whereas the 924GG genotype produced two bands (172 and 49 bp), whereas the 8360G allele and q =1 − p was tested by a goodness-of-fit χ²-test to compare the observed genotype frequencies with the expected genotype frequencies in cancer-free controls. Unconditional logistical regression analysis was used to calculate odds ratios (ORs) and their 95% confidence intervals (CIs) with and without adjustment for age, ethnicity, smoking status and alcohol use. The logistic regression model was also used for the trend test. The 2LD program and Proc ALLELE in SAS/Genetics software were used to detect the LD of any pair of SNPs. Proc HAPLOTYPE in SAS/Genetics software using the expectation-maximization (EM) algorithm was applied to generate maximum likelihood estimates of haplotype frequencies (41). It was assumed that the Hardy–Weinberg equilibrium is applicable to the constructed haplotypes based on observed genotypes of NBS1. The PHASE program using Bayesian methods with Markov chain Monte Carlo algorithm was also used to construct haplotypes (42,43), and the results from PHASE and SAS were compared. Haplotype and diplotype pairs were combined on the basis of haplotype information. The associations between the diplotypes and breast cancer risk were evaluated with logistic regression models. The joint effects of the haplotypes and breast cancer risk were evaluated with logistic regression on the basis of haplotype information. The associations between the diplotypes and breast cancer risk were evaluated with logistic regression based on observed genotypes of NBS1. The PHASE program using Bayesian methods with Markov chain Monte Carlo algorithm was also used to construct haplotypes (42,43), and the results from PHASE and SAS were compared. Haplotype and diplotype pairs were combined on the basis of haplotype information. The associations between the diplotypes and breast cancer risk were evaluated with logistic regression models. The joint effects of the NBS1 polymorphisms, smoking and drinking on breast cancer risk were calculated. All statistical tests were two-sided, a P-value of 0.05 being considered significant, by using SAS software (version 8.2; SAS Institute, Cary, NC).

Results

Characteristics of the study population

This analysis included 421 breast cancer cases and 423 cancer-free controls. Because we used frequency-matching, there were no significant differences in the distributions of age, smoking status and alcohol usage between the cases and controls (P = 0.545, 0.397 and 0.241, respectively). However, case subjects were slightly older (mean age ± standard deviation = 45.4 ± 6.6 years, range = 22–55 years) than control subjects (mean age ± standard deviation = 44.9 ± 8.2 years, range = 21–55 years), but the difference

60 s, and then a final extension step of 72°C for 5 min, there were 35 cycles at 95°C for 45 s and 72°C for 10 min. The amplified fragments were digested with BclI (New England BioLabs, Beverly, MA, USA) overnight, and the products were separated in 3% MetaPhor agarose gel. The 924TT genotype produced two bands (121 and 20 bp), whereas the 924TC genotype produced a single band (141 bp), and the heterozygotes displayed all three bands (141, 121 and 20 bp).

The NBS1 8360G>C polymorphism was also analyzed by the PCR–RFLP method. We amplified the SNP-containing fragment by using the following self-designed primers: 5'-TCAGCCAGATGGCAGACTC-3' (forward) and 5'-TCCTGAAACAAGCATTAAAGAGG-3' (reverse). PCR was performed in 10 μl reaction systems containing 2 mM MgCl₂, 0.04 mM deoxynucleotide triphosphates, 1.0 units of Taq polymerase and the manufacturer's buffer [20 mM Tris–HCl (pH 8.4) and 50 mM KCl]. The PCR profile consisted of an initial melting step at 95°C for 5 min, 35 cycles of 95°C for 30 s, 61°C for 45 s and 72°C for 1 min. The amplified fragments were digested with BglII (New England BioLabs) overnight, and the products were separated in 2% agarose gel. The 924TT genotype produced two bands (172 and 25 bp), and the heterozygotes displayed all three bands (174, 125 and 49 bp).

The NBS1 30537G>C polymorphism was also analyzed by the PCR–RFLP method. We amplified the SNP-containing fragment by using the following self-designed primers: 5'-CGTCCAATTGTAAAGCCAGAA-3' (forward) and 5'-TCCTGAAACAAGCATTAAAGAGG-3' (reverse). PCR was performed in the 10 μl reaction systems. After an initial denaturation at 94°C for 5 min, there were 35 cycles at 95°C for 45 s, 54°C for 45 s and 72°C for 10 min. The amplified fragments were digested with HinfI (New England BioLabs) overnight, and the products were separated in 2% agarose gel. The 924TT genotype produced two bands (125 and 49 bp), whereas the 924TC genotype produced a single band (174 bp), and the heterozygotes displayed all three bands (174, 125 and 49 bp).

Fig. 1. NBS1 gene structure and the positions of selected htSNPs. NBS1 contains 16 exons and spans over 50 kb. The block of high LD and limited haplotype diversity is highlighted with a triangle. Three SNPs (924T>C, 8360G>C and 30537G>C) were selected as the htSNPs (labeled with arrows) and genotyped in the study: 924T>C locates in 5' untranslated region; 8360G>C (Glu185Gln) locates in exon 5; and 30537G>C is in intron 9.

96 samples (11.37% of 844 samples) for each of three SNPs for repeated assays, and the results were 100% concordant.

Statistical analysis

χ²-tests were used to compare the distribution of demographic variables and selected risk factors between cases and controls. The Hardy–Weinberg equilibrium (p² + 2pq + q² = 1, where p is the frequency of the variant allele and q =1 − p) was tested by a goodness-of-fit χ²-test to compare the observed genotype frequencies with the expected genotype frequencies in cancer-free controls. Unconditional logistical regression analysis was used to calculate odds ratios (ORs) and their 95% confidence intervals (CIs) with and without adjustment for age, ethnicity, smoking status and alcohol use. The logistic regression model was also used for the trend test. The 2LD program and Proc ALLELE in SAS/Genetics software were used to detect the LD of any pair of SNPs. Proc HAPLOTYPE in SAS/Genetics software using the expectation-maximization (EM) algorithm was applied to generate maximum likelihood estimates of haplotype frequencies (41). It was assumed that the Hardy–Weinberg equilibrium is applicable to the constructed haplotypes based on observed genotypes of NBS1. The PHASE program using Bayesian methods with Markov chain Monte Carlo algorithm was also used to construct haplotypes (42,43), and the results from PHASE and SAS were compared. Haplotype and diplotype pairs were combined on the basis of haplotype information. The associations between the diplotypes and breast cancer risk were evaluated with logistic regression models. The joint effects of the NBS1 polymorphisms, smoking and drinking on breast cancer risk were calculated. All statistical tests were two-sided, a P-value of 0.05 being considered significant, by using SAS software (version 8.2; SAS Institute, Cary, NC).
was not statistically significant (Student’s t-test: \( P = 0.329 \)). Additionally, multivariate logistic regression models were used to control these variables for any residual effect of confounding on the main effect of the selected \( \text{NBS1} \) polymorphisms.

\( \text{NBS1} \) genotype distributions and risk of breast cancer

The genotype and allele distributions of the \( \text{NBS1} \) 924T>C, 8360G>C and 30537G>C polymorphisms among the cases and controls are summarized in Table I. The observed genotype frequencies of these three SNPs were all in agreement with the Hardy–Weinberg equilibrium in the control subjects (\( P = 0.515 \) for \( \text{NBS1} \) 924T>C; \( \chi^2 = 0.27, P = 0.603 \) for \( \text{NBS1} \) 8360G>C; and \( \chi^2 = 1.56, P = 0.212 \) for \( \text{NBS1} \) 30537G>C).

\( \text{NBS1} \) genotype frequencies were 40.6\% (GG), 47.0\% (GC), and 12.4\% (CC) in the cases and 48.9\% (GG), 42.8\% (GC) and 8.3\% (CC) in the controls. The difference was statistically significant (\( P = 0.005 \)). For the \( \text{NBS1} \) 30537G>C polymorphism, the 30537CC variant homozygous carriers had a non-significantly increased risk (1.90-fold, 95\% CI = 0.69–5.22), and the heterozygous 30537GC carriers had no risk for breast cancer at all, compared with the 30537GG genotype carriers (Table I). These data suggest that the 8360C polymorphism is a risk allele, since it causes amino acid change, whereas 924C and 30537C polymorphisms may be risk alleles or in linkage with other unknown risk alleles.

\( \text{NBS1} \) haplotypes/diplotypes and risk of breast cancer

To further analyze the combined effect of these three tagging SNPs, we generated the haplotypes on the basis of the observed genotypes. As shown in Table II, the PHASE and SAS Proc HAPLOTYPE programs both showed eight possible \( \text{NBS1} \) haplotypes from the observed genotypes. The five common (\( \geq 5\% \)) haplotype alleles (TGG, TCG, TGC, CCG and TCC) represented 94.6\% of the chromosomes in the controls. The frequency distribution of haplotype alleles between the cases and controls was not significantly different (\( P = 0.189 \)). However, when we used the most common haplotype allele TGG as the reference, the CCG haplotype allele was associated with a 74\% increased risk of breast cancer (OR = 1.74; 95\% CI = 1.17–2.59). Because the frequencies of all minor variant alleles in the three selected htSNPs were associated with increased risk of breast cancer (Table I), and the frequencies of some haplotypes were <5\%, we recombed all possible haplotypes into three groups on the basis of the number of minor variants within the haplotype alleles, assuming that the effects of the minor variant alleles would follow an additive model. Given that an individual’s genotype of the gene is a mix of all possible SNPs, the analysis of diplotype
provide the estimate for an actual effect of such a genotype on the risk of breast cancer. On the basis of the number of variants within the haplotype alleles, we recombinated the haplotypes into three groups. The difference in frequency distributions of these haplotype groups among cases and controls was statistically significant (P = 0.012). Compared with the non-variant haplotype group Hap0, the risk for breast cancer was increased as the number of variant alleles increased in a dose–response manner [OR = 1.11, 95% CI = 0.89–1.39 for Hap1 (i.e. having 1 variant allele) and OR = 1.51; 95% CI = 1.14–2.00 for Hap2 (i.e. having ≥2 variant alleles); P_trend = 0.004] (Table II).

We further examined the distribution of the diplotypes based on the combination of the eight possible haplotypes. As shown in Table III, 25 possible diplotypes were generated. Using the most common TGG/TGG diplotype as the reference, only one diplotype CCG/CCG was associated with a 9.21-fold elevated risk of breast cancer (95% CI = 1.10–77.3). When we divided these 25 diplotypes into three groups (i.e. no variant allele, 1–2 variant alleles and 3–6 variant alleles), the distributions between cases and controls were significantly different (P = 0.002). When the genotypes without any variant allele (i.e. any of 924C, 8360C and 30537C) were used as the reference, the cancer risk increased as the number of variant alleles increased in a dose–response manner (adjusted OR = 1.07, 95% CI = 0.78–1.46 for 1–2 variant alleles; adjusted OR = 2.47; 95% CI = 1.48–4.14 for 3–6 variant alleles; P_trend = 0.006) (Table III).

Stratification analysis of NBS1 diplotypes and risk of breast cancer

To take into account all three variants, we used the combined NBS1 diplotypes for stratification analysis by age, smoking status, alcohol use and tumor histological type. Compared with a zero variant allele, the trend of significantly increased cancer risk with an increased variant number of alleles was also obvious in subgroups of cases who were <45 years (adjusted OR = 1.30, 95% CI = 0.78–2.18 for 1–2 variant alleles; adjusted OR = 2.09, 95% CI = 1.01–4.35 for 3–6 variant alleles; P_trend = 0.048), never smokers (adjusted OR = 1.09, 95% CI = 0.73–1.65 for 1–2 variant alleles; adjusted OR = 2.70, 95% CI = 1.37–5.30 for 3–6 variant alleles; P_trend = 0.021) and never drinkers (adjusted OR = 1.32, 95% CI = 0.85–2.07 for 1–2 variant alleles; adjusted OR = 3.33, 95% CI = 1.58–7.04 for 3–6 variant alleles; P_trend = 0.005) (Table IV).

However, there was no any difference between tumor histological types (i.e. ductal or lobular carcinomas in situ and invasive carcinomas) or among the AJCC stages in risk associated with the NBS1 polymorphism of 924T>C, 8360G>C, 30537G>C and its diplotypes (data not shown). Furthermore, there were no significant differences in the distribution of subjects’ characteristics [i.e. menopausal status, age at first full-term pregnancy, number of live births, use of hormone replacement or oral contraceptives, education level and body mass index (BMI, kg/m²) by the NBS1 genotypes] (data not shown). Taken together, these data...
suggest that the NBS1 variant genotypes and diplotypes may be biomarkers for susceptibility rather than a disease biomarker. Finally, we analyzed the joint effects of the NBS1 polymorphism of 924T>C, 8360G>C, 30537G>C and its diplotypes with smoking and drinking exposure on breast cancer risk. We did not find any evidence of modification or interaction (data not shown).

**Discussion**

We have previously reported that young breast cancer patients tended to have suboptimal DNA repair capacity for removing BPDE-induced DNA adducts (7) or have higher levels of BPDE-induced chromosomal aberrations (8). In the present study, we were interested in the association between the htSNPs of the NBS1 gene and risk for sporadic breast cancer. When we evaluated each of the selected htSNPs separately, the NBS1 8360C (185Gln) variant genotypes were found to be associated with increased risk in an allele-dose response manner, whereas only the homozygous 924CC genotype was associated with an increased risk. The homozygous 30537CC was associated with non-significantly increased risk. However, when the risk alleles (i.e. 924C, 8360C and 30537C) were combined either in the form of haplotypes or diplotypes, the risk for breast cancer appeared to increase as the number of risk alleles increased. Although the functional relevance of these selected htSNPs remains unknown, our findings are biologically plausible, because the htSNPs are either non-synonymous (Glu185Gln) or located in the promoter region.

The NBS1 8360G>C polymorphism is located in the area coding for breast cancer C-terminal (BRCT) domain (108–196 amino acids) (9), which facilitates the interaction between NBS1 and BRCA1 (breast cancer 1), forming a BRCA1-associated genome surveillance complex (BASC), which is responsible for recognition and repair of aberrant DNA (44–47). This might partially explain the possible effect of the 8360G>C polymorphism on breast cancer risk, that is, the amino acid change caused by the variant may interfere with such a protein–protein interaction. However, this hypothesis needs to be further tested.

Several studies have investigated the association between polymorphisms of NBS1 8360G>C (Glu185Gln) and cancer risk, but the results are mixed. A recent large U.S. study of 894 African–American breast cancer cases and 788 controls and 1417 whites breast cancer cases and 1234 controls, between 21 and 74 years, reported that Glu185Gln variant genotypes were not associated with risk of breast cancer (30). A similar result was reported in a large case–control study of 2205 breast cancer cases and 1826 controls, between 45 and 74 years, in Europe (31). We do not know what percentage of the subjects in these studies were 55 years or younger; it is possible that the analysis of mixed age groups may mask the moderate effect of Glu185Gln in young women. A combined study of 221 Finnish sporadic breast cancer cases and Polish familial cases and 319 controls also reported no association between the Glu185Gln polymorphism and breast cancer (28), as did a study of a Chinese population of 220 cases and 310 controls (29), but these studies were relatively small.

For other cancers, however, Medina et al. (48) report that the NBS1 185Gln variant genotypes (8360GC/CC) were associated with a p53 mutation in lung cancer and concluded that the genotypes may play a role in lung carcinogenesis. However, a study in a Chinese population reported that 8360CC (42.3% in controls) was the common genotype and that the 8360GG genotype was associated with increased risk of lung cancer (49). Other studies, including a case–control study of non-small cell lung cancer in Norway (50), a multicenter study of ovarian cancer from the United Kingdom, Denmark and the United States (51) and a study of non-Hodgkin’s lymphoma, have not found any associations of NBS1 185Gln and cancer risk (52).

Compared with other studies, our present study is rather unique. We included relatively young (≤55 years) non-Hispanic white women who were frequency-matched with controls on smoking status because genetic susceptibility is often characterized by an early age of onset and smoking is a potential risk factor for breast cancer at a young age. Therefore, a positive finding between the association of the 8360G>C polymorphism and breast cancer risk may be because of the inclusion of an at-risk population and frequency-matching, which may have enhanced the statistical power of our study. Furthermore, the findings from the present study are consistent with the findings from our two previously published pilot studies in which we found that increased risk of breast cancer in young women was significantly associated with a reduced capacity for repairing DNA damage and with chromosomal aberrations induced by a tobacco carcinogen (7,8).
However, it is likely that our results could be biased owing to possible selection biases in hospital-based case–control studies. Because the genotype frequencies of the NBS1 8360G>C polymorphism estimated from the hospital-based control subjects in our study were very close to the resequencing data in the Cancer Genome Anatomy Project SNP500 Cancer Database for Caucasians (http://snp500cancer.nci.nih.gov), the potential selection bias in genotype distribution is unlikely to be substantial, if any. Nevertheless, larger studies with young women are warranted to further test whether the NBS1 Glu185Gln variant is associated with breast cancer risk.

In this study, we also found that variant homozygote 924CC carriers had a 4.55-fold elevated risk of breast cancer compared with the homozygote 924TT carriers, whereas the 30537CC polymorphism was associated with a non-significantly increased risk of breast cancer compared with the homozygote 30537GG carriers. The significance of these findings, however, is limited because of the small number of cases in the subgroups. To the best of our knowledge, no previous studies have explored the association of NBS1 924T>C or 30537G>C polymorphisms and breast cancer risk. The NBS1 924T>C polymorphism locates in the 5’ untranslated region (−1120 nt) of the NBS1 gene, which has been shown by bioinformatics analysis that it is the transcription factor GATA-1 binding site. The activation domains of GATA-1 are capable of activating transcription in mammalian cells through GATA motifs (53). When this locus changes to 924C, it might cause a loss of the GATA-1-binding cis-element and thus may have an adverse effect on NBS1 gene expression.

The analyses of haplotypes and diplotypes with the combination of the three NBS1 variants showed that an elevated risk of breast cancer was associated with haplotypes or diplotypes in an allele-dose response manner. This suggests that NBS1 924T>C or 30537G>C polymorphisms also have a potential role in the etiology of breast cancer. Our analysis on the haplotype and diplotype provided additional information on these SNPs as markers of genetic susceptibility and a more efficient method for assessing the genetic susceptibility of a candidate gene than any of the polymorphisms. Further stratification analysis showed that the increased risk associated with variant diplotypes was more pronounced in the subgroup of never smokers and never drinkers, consistent with the concept that the susceptible individuals may have been exposed to a low level of carcinogens for a shorter period of time. However, these results may also be biased by the relatively small number of subjects in the subgroups. Further studies with large sample sizes are needed to verify these findings.

Our study has inherited shortcomings of a hospital-based case–control study. First of all, because our cases and controls were recruited from a hospital, they may not be representative of the general population and could have introduced selection bias. Also, because we restricted the subjects to non-Hispanic whites, it is uncertain whether there may be population stratification and whether these results are generalizable to other populations. A recent article reported the comparison of haplotype information between databases of HapMap and NIEHS; the authors indicated that the NIEHS provided more information with resequenced SNPs than that of HapMap, but it lacks the ethnic information of each individual (54). Furthermore, the tagging SNPs in our testing set of 118 non-Hispanic whites were also showing differences in LD compared with those of HapMap. Therefore, potential population stratification may have contributed to the differences observed in the different data sets. Finally, limited information on reproductive factors, particularly hormonal use, in this study population did not allow us to explore the interactions between genes and the environment.

In conclusion, in this hospital-based case–control study of sporadic breast cancer, we found a significant association between NBS1 polymorphisms and breast cancer risk in non-Hispanic whites. Furthermore, the haplotypes of NBS1 generated from the observed genotypes were associated with significantly increased cancer risk in an allele/dose response manner, especially in non-smokers and non-drinkers. This supports the notion that variant genotypes of the NBS1 gene may contribute to the risk of sporadic breast cancer, particularly in young women. However, because we used less stringent \( r^2 \) values (i.e. 0.5 instead of 0.8) to define complete LD between SNPs in the SNP selection (55), we could have suffered some loss of our statistical power in our risk estimates. Furthermore, because the study design was retrospective, the observed associations may have been biased or simply due to chance. Therefore, larger, prospective studies are needed to verify these findings.

Acknowledgements

This study was supported in part by National Institutes of Health grants CA 108364 (L.E.W.), ES11740 (Q.W.) and CA 16672 (M. D. Anderson Cancer Center). We thank Margaret Lung, Maria Torres, and Leaneal Fairly for their assistance in recruiting the subjects; Quiling Shi for statistical support; Zhihun Hu, Luo Wang, Zhenheng Liu, Zhouzheng Gao, Yawei Qiao, Jianzhong He, Kejing Xu and Yinyan Li for their laboratory assistance; Joanne M. Sider and Zhibin Hu, Luo Wang, Zhensheng Liu, Zhaozheng Guo, Yawei Qiao, Jianzhong He, Kejing Xu and Yinyan Li for their laboratory assistance; Joanne M. Sider for manuscript preparation; and Martha Morrison for scientific editing.

Conflict of Interest Statement: None declared.

References


Carcinogenesis, 23, 687–696.