Common Single-Nucleotide Polymorphisms in DNA Double-Strand Break Repair Genes and Breast Cancer Risk

Karen A. Pooley,1 Caroline Baynes,2 Kristy E. Driver,2 Jonathan Tyrer,2 Elizabeth M. Azzato,2,4 Paul D.P. Pharoah,2 Douglas F. Easton,1 Bruce A.J. Ponder,3 and Alison M. Dunning2

1Cancer Research UK Genetic Epidemiology Unit and 2Cancer Research UK Department of Oncology, Strangeways Research Laboratory, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, United Kingdom; and 3Division of Cancer Epidemiology and Genetics, National Cancer Institutes, NIH, Rockville, Maryland

Abstract

The proteins involved in homologous recombination are instrumental in the error-free repair of dsDNA breakages, and common germ-line variations in these genes are, therefore, potential candidates for involvement in breast cancer development and progression. We carried out a search for common, low-penetrance susceptibility alleles by tagging the common variation in 13 genes in this pathway in a two-stage case-control study. We genotyped 100 single-nucleotide polymorphisms (SNP), tagging the 655 common SNPs in these genes, in up to 4,470 cases and 4,560 controls from the SEARCH study. None of these tagging SNPs was associated with breast cancer risk, with the exception of XRCC2 rs3218536, R188H, which showed some evidence of a protective association for the rare allele [per allele odds ratio, 0.89; 95% confidence intervals (95% CI), 0.80-0.99; P trend = 0.03]. Further analyses showed that this effect was confined to a risk of progesterone receptor positive tumors (per rare allele odds ratio, 0.78; 95% CI, 0.66-0.91; P trend = 0.002). Several other SNPs also showed receptor status-specific susceptibility and evidence of roles in long-term survival, with the rare allele of BRIP1 rs2191249 showing evidence of association with a poorer prognosis (hazard ratio per minor allele, 1.20; 95% CI, 1.07-1.36; P trend = 0.002). In summary, there was little evidence of breast cancer susceptibility with any of the SNPs studied, but larger studies would be needed to confirm subgroup effects. (Cancer Epidemiol Biomarkers Prev 2008;17(12):3482–9)

Introduction

The human genome encodes proteins that safeguard its own integrity by continuously monitoring and repairing dsDNA damage caused by ionizing radiation, a known risk factor for breast cancer (1-5). This vigilance is critical at a cellular level to maintain a high degree of replication fidelity, as accumulation of DNA damage is a major step toward tumorigenesis (6-9).

The homologous recombination pathway and its associated upstream gatekeeper molecules oversee the error-free DNA reconstruction of double-stranded lesions (10-15). Indeed, the most penetrant breast cancer genes identified thus far, BRCA1, BRCA2, and CHEK2, play major roles in DNA repair, although common variants in these genes confer no significant risk of breast cancer (16-18).

Materials and Methods

Breast Cancer Case-Control Series. Cases were drawn from the SEARCH (Breast) Study, an ongoing population-based study with cases ascertained through the Eastern Cancer Registration and Information Centre.4 All

Received 6/30/08; revised 8/15/08; accepted 9/11/08.

Grant support: Cancer Research UK. B.A.J. Ponder is a Cancer Research UK Gibb Fellow, P.D.P. Pharoah is a Cancer Research UK Senior Clinical Research Fellow, and D.F. Easton is a Cancer Research UK Principal Research Fellow. EMA is funded by the NIH-University of Cambridge Graduate Partnership Program and the National Cancer Institute.

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

Requests for reprints: Karen A. Pooley, Cancer Research UK Genetic Epidemiology Unit, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, United Kingdom. Phone: 44-1223-741618; Fax: 44-1223-740147. E-mail: karenp@srl.cam.ac.uk

Copyright © 2008 American Association for Cancer Research. doi:10.1158/1055-9965.EPI-08-0594
women diagnosed with invasive breast cancer under the age of 55 y between January 1, 1991 and June 30, 1996 and who were alive at the start of the study (prevalent cases, median age of 48 y) as well as women under the age of 70 y who were diagnosed from 1996 onwards (incident cases, median age of 54 y) were eligible for inclusion. Approximately 67% of eligible patients have enrolled in the study. Women taking part in the study were asked to provide a 20 mL blood sample for DNA analysis and to complete a comprehensive epidemiologic questionnaire. Eligible patients who did not take part in the study were similar to participants, except, as might be expected, the proportion of clinical stage III/IV cases was somewhat higher in nonparticipants (Supplementary Table S1A). Controls were randomly selected from the Norfolk component of European Prospective Investigation of Cancer (EPIC; ref. 33). EPIC is a prospective study of diet and cancer being carried out in nine European countries. The EPIC-Norfolk cohort comprises 25,000 individuals resident in Norfolk, East Anglia, the same region from which the cases have been recruited. Controls are not matched to cases but are broadly similar in age, having a median age of between 42 and 81 y. Ethical approval was obtained from the Eastern Multi-centre Research Ethics Committee and informed consent was obtained from each patient.

The total number of cases available for analysis was 4,470, of which 27% were prevalent cases. The samples have been split into two sets to save DNA and reduce genotyping costs: the first set \((n = 2,270\) cases and 2,280 controls) is genotyped for all single-nucleotide polymorphisms (SNP) and the second set \((n = 2,200\) cases and 2,280 controls) is then tested for those SNP that show marginally significant associations with breast cancer risk in Set 1 \((P \text{ trend} \leq 0.10)\) or, similarly, associations in the subgroup analyses with hormone receptor status or long-term prognosis \((P \text{ trend} \leq 0.05)\). This staged approach substantially reduces genotyping costs without significantly affecting statistical power (Supplementary Table S1B). Cases were randomly selected for Set 1 from the first 3,500 recruited, with Set 2 comprising the remainder of these plus the next 970 incident cases recruited. As the prevalent cases were recruited first, the proportion of prevalent cases was somewhat higher in Set 1 than Set 2 (33% versus 20%). Median age at diagnosis was similar in both sets (51 and 52 y old, respectively). There was no significant difference in the morphology, histopathologic grade, or clinical stage of the cases by set or by prevalent/incident status (34).

**Power.** The statistical power of the study depends on the susceptibility allele frequency, the risks conferred, and the genetic mode of action (dominant, recessive, and codominant). The staged approach substantially reduces genotyping costs without significantly affecting statistical power (Supplementary Table S1B; ref. 35). For example, assuming that the causative SNP is tagged with \(r^2 = 0.8\), a type I error rate of 0.0001, and genotyping success rate of 95%, the staged study has 86% power (versus full study power of 88%) to detect a dominant allele with minor allele frequency (MAF) = 0.05 that confers a relative risk of 1.5. Power to detect a dominant allele with MAF = 0.25 that confers a relative risk of 1.3 is 87% in the staged study and 89% in the full study. Power to detect recessive alleles is less: 53% in the staged study (versus 60% full study) for an allele with MAF = 0.25 and risk of 1.5 and 71% (versus 75%) for an allele with MAF = 0.5 and risk of 1.3.

**Selection of Tagging SNPs.** The aim of the SNP tagging approach is to identify a set of SNPs that efficiently captures all the known SNPs and any unknown SNPs in the gene. The best measure of the extent to which one SNP tags another SNP is the square of the pairwise correlation coefficient, \(r^2\), because the loss of power incurred by using a marker SNP in place of a true causal SNP is directly related to this value. We attempted to define a set of tagging SNPs such that all known common SNPs (MAF > 0.05) in each of the genes studied had an estimated \(r^2 \geq 0.8\) with at least one tagging SNP. However, some SNPs are poorly correlated with other single SNPs but may be efficiently tagged by multiple SNPs (multimarker tags), thus reducing the number of tagging SNPs needed (36). Therefore, as an alternative, we aimed for the correlation between each SNP and a specific multimarker haplotype \((r^2)\) to be >0.8. SNPs were selected using the “aggressive tagging” option implemented in the Tagger program, which combines the simplicity of the pairwise approach with the additional efficacy of the multimarker method and easy visualization through haplotype human. Tagging SNPs were identified from one of two sources, depending on which had the higher SNP density per kb. The International HapMap Project\(^6\) has genotyped several million randomly selected SNPs, in four different populations, including 30 mother-father-child trios of Utah residents with ancestry from northern and western Europe, collected by the Centre d’Etude du Poly Morphisme Humain. Tag SNPs chosen using HapMap data have been shown to capture all known common variation very well (38).

The other data used were the National Institute of Environmental Health Sciences (NIEHS) EGP GeneSNP resource,\(^7\) which has resequenced candidate genes for cancer across panels of individuals representative of U.S. ethnicities. The original panel (P1-PDR90) of 90 individuals consisted of 24 European Americans, 24 African-Americans, 12 Mexican Americans, 6 Native Americans, and 24 Asian Americans, but the ethnic group identifiers were not available. It is known that there is greater genetic and haplotype diversity in individuals of African origin. To reduce this, we have identified and excluded 28 of the samples with the greatest African ancestry by comparing the genotypes of the PDR90 subjects with the genotypes of the National Heart, Lung, and Blood Institute Variation Discovery Resource Project African American Panel\(^8\) for the same SNPs. Data from the remaining 62 individuals were used to identify tag SNPs. Ideally, samples with Hispanic American, Asian American, and Native American ancestry should also be removed, but as there is less genetic diversity among these groups and between these groups and the European Americans, they cannot be excluded with any certainty.

---

\(^6\) http://www.broad.mit.edu/mpg/tagger/

\(^7\) National Center for Biotechnology Information Build 36 Release 22 April 2007 (http://www.hapmap.org/).

\(^8\) http://egp.gs.washington.edu/
DNA Damage Genes and Breast Cancer Risk

**Taqman Genotyping.** Genotyping was done by Taqman assay using the ABI PRISM 7900HT Sequence Detection System according to the manufacturer’s instructions. Primers and probes were supplied directly by Applied Biosystems as Assays-by-Design. All assays were carried out in 384-well plate format, with each plate including negative controls (with no DNA) and positive controls duplicated on a separate quality control plate. Each study set (3.2%) was duplicated for quality control. Assays for which >98% of the duplicated samples did not give identical genotypes were discarded and replaced with alternative assays with the same tagging properties. Additionally, call rates of >90% per plate and >95% for the set overall were required or the assay was discarded and replaced with an alternative tag. Failed genotypes were not repeated.

**Statistical Methods.** For each SNP, the deviation of genotype distribution in controls from Hardy-Weinberg equilibrium was assessed by a \( \chi^2 \) test with 1 degree of freedom (df). The primary tests of association were univariate analyses for each of the SNPs studied. Genotype frequencies in cases and controls were compared using a 2 df \( \chi^2 \) test for heterogeneity (data not shown) and a 1 df Cochran-Armitage \( \chi^2 \) test for trend in risk by allele dose (P trend). Genotype-specific risks were estimated as odds ratios (OR), with associated confidence intervals (CI), using unconditional logistic regression. Genotype distributions were compared between prevalent and incident cases and between subjects in Set 1 and Set 2 with \( \chi^2 \) tests (2 df). No statistically significant differences were found and the results have therefore been combined (data not shown).

In addition to the univariate analyses, we carried out specific haplotype tests for combinations of alleles (multimarker tags) that tagged specific SNPs. We also carried out a general comparison of common haplotypes (frequencies > 0.05) in each gene haplotype block. Rarer haplotypes (<0.05) were pooled into a single category. An in-house program based on the TagSNPs program (39) and carrying out unconditional logistic regression was used for both the multimarker and full gene haplotype analyses. Haplotype-specific risks were estimated with respect to all other common haplotypes present (and the "rare" grouping) as ORs with associated CIs, together with global score statistics for each gene.

SNPs showing any evidence of association in the single SNP or multimarker Cochran-Armitage 1 df tests (P trend ≤ 0.10) were genotyped in Set 2 and unconditional logistic regression was done with the combined data.

Survival analysis was conducted to determine the effect of each common SNP on survival for all Set 1 cases. Standard Cox regression analyses were done to estimate hazard ratios (HR) for each genotype relative to the common homozygote using a 1 df trend test. The proportional hazards assumption was evaluated by visual inspection of standard log-log plots for each SNP, as well as analytically using Schoenfeld residuals. SNP’s showing marginal significant results in Set 1 (P trend ≤ 0.05) were genotyped in Set 2 and the combined data were analyzed in the same way, by Standard Cox regression, with evaluation of the proportional hazards assumption. Time at risk began on the date of blood draw, thus left truncating the data to allow the inclusion of prevalent cases, and ended at the date of death from any cause or on November 30, 2006, the date of last follow-up, whichever occurred first. Follow-up was censored for all cases at 10 y after initial diagnosis, as the number of patients with a longer time of follow-up was relatively small and the follow-up of these patients became less reliable.

Heterogeneity of genotype distribution per SNP with respect to receptor status was assessed using unconditional logistic regression for the cases on which these tumor subtype data were available. Receptor-negative and receptor-positive Set 1 cases were compared separately with the full complement of Set 1 controls. SNPs showing marginally significant results in Set 1 (P trend ≤ 0.05) were genotyped in Set 2 and the combined data were analyzed in the same way, with receptor-negative and receptor-positive cases compared separately to the full complement of Set 1 + Set 2 controls.

All data analyses were done using Intercooled STATA version 8.2 for Windows.

**Estrogen and Progesterone Receptor Status.** Estrogen receptor (ER) and progesterone receptor (PR) expression was determined in cores from paraffin-embedded primary tumor tissue, collected as part of the SEARCH study. Immunohistochemistry was done using the BondMax autostainer on-board retrieval system (Vision Biosystems) and the ER 6f11 (Novocastra) and PR PgR636 (Dako) antibodies. All samples were scored using the Allred scoring system (40) by a pathologist without prior knowledge of sample background or treatment history.

**Results**

**SNP Tagging.** Using the NIEHS and HapMap data sources as detailed in Table 1, there were 655 known common SNPs (MAF > 0.05) in the 13 genes of interest and a subset of 98 tagging SNPs was chosen using the Tagger program. In these genes, 86% of all known SNPs were successfully tagged with \( r^2 > 0.8 \) and a further 12% with \( r^2 > 0.8 \) (Table 1). For each of the genes studied, the common SNPs lay in a single, well-defined linkage disequilibrium block showing no evidence for recombination hotspots.

An additional two rare coding variants in the RAD52 gene were also genotyped as they had been previously studied by others (41, 42).

**SNP Associations.** The genotype frequency distributions and ORs for the SNPs studied and the multimarker tags are shown in Supplementary Table S2. Genotype distributions in controls did not deviate significantly from that expected under Hardy-Weinberg equilibrium (HWE); 5% of SNPs would be expected to give P HWE of <0.05 by chance, whereas we observed 8% SNPs with a P HWE of >0.01 and <0.05 (data not shown). Reevaluation of the raw genotyping data for these SNPs revealed nothing abnormal about the assays or cluster plots, so they seem likely to have been chance findings. The estimated haplotype frequencies for each gene and their associated risks are shown in Supplementary Table S3.

Of the 109 individual tag SNPs plus multimarker combinations studied, 98 showed no suggestion of an association with breast cancer risk in the first stage of the two-stage study (P trend ≤ 0.10). None of the SNPs tested...
was associated with age in controls and age-adjusted results were similar to the unadjusted results (data not shown). Eight individual tags and three multimarker combinations merited further investigation in Set 2. In total, 11 SNPs progressed to the second stage based on these criteria (Table 2).

Table 2. Breast cancer risks associated with the 13 SNPs genotyped in SEARCH Set 1 and Set 2

<table>
<thead>
<tr>
<th>Genes and SNPs</th>
<th>MAF</th>
<th>Base change</th>
<th>Controls</th>
<th>Cases</th>
<th>OR (95% CI)</th>
<th>OR per allele (95% CI)</th>
<th>P trend (1 df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>XRCC2 rs3218536</td>
<td>0.09</td>
<td>GG</td>
<td>3,639</td>
<td>3,590</td>
<td>1*</td>
<td>0.89 (0.80-0.99)</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GA</td>
<td>711</td>
<td>610</td>
<td>0.93 (0.57-1.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>34</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAD51 rs12442560</td>
<td>0.42</td>
<td>AA</td>
<td>1,467</td>
<td>1,399</td>
<td>1*</td>
<td>0.98 (0.90-1.08)</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AC</td>
<td>2,192</td>
<td>2,058</td>
<td></td>
<td>1.07 (0.95-1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CC</td>
<td>835</td>
<td>854</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NBS1 rs9995</td>
<td>0.32</td>
<td>TT</td>
<td>2,036</td>
<td>1,947</td>
<td>1*</td>
<td>1.00 (0.91-1.09)</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TC</td>
<td>1,993</td>
<td>1,898</td>
<td></td>
<td>1.05 (0.92-1.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CC</td>
<td>509</td>
<td>513</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BARD1 rs6435862</td>
<td>0.27</td>
<td>TT</td>
<td>2,397</td>
<td>2,194</td>
<td>1*</td>
<td>1.12 (1.03-1.22)</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TG</td>
<td>1,789</td>
<td>1,835</td>
<td></td>
<td>1.01 (0.86-1.19)</td>
<td></td>
</tr>
<tr>
<td>rs7591615</td>
<td>0.25</td>
<td>CC</td>
<td>2,546</td>
<td>2,508</td>
<td>1*</td>
<td>0.94 (0.86-1.03)</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CT</td>
<td>1,674</td>
<td>1,554</td>
<td></td>
<td>0.99 (0.82-1.18)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TT</td>
<td>261</td>
<td>254</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3768704</td>
<td>0.12</td>
<td>GG</td>
<td>3,516</td>
<td>3,421</td>
<td>1*</td>
<td>0.92 (0.83-1.03)</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GA</td>
<td>927</td>
<td>833</td>
<td></td>
<td>0.96 (0.69-1.33)</td>
<td></td>
</tr>
<tr>
<td>rs2229571</td>
<td>0.39</td>
<td>CC</td>
<td>1,679</td>
<td>1,587</td>
<td>1*</td>
<td>1.02 (0.93-1.12)</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG</td>
<td>2,158</td>
<td>2,081</td>
<td></td>
<td>1.01 (0.88-1.14)</td>
<td></td>
</tr>
<tr>
<td>rs6717301</td>
<td>0.15</td>
<td>CC</td>
<td>3,279</td>
<td>3,168</td>
<td>1*</td>
<td>0.98 (0.89-1.08)</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CG</td>
<td>1,123</td>
<td>1,064</td>
<td></td>
<td>1.00 (0.92-1.08)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG</td>
<td>97</td>
<td>100</td>
<td></td>
<td>1.07 (0.80-1.42)</td>
<td></td>
</tr>
<tr>
<td>rs11896262</td>
<td>0.36</td>
<td>TT</td>
<td>1,868</td>
<td>1,746</td>
<td>1*</td>
<td>1.03 (0.94-1.13)</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TA</td>
<td>2,063</td>
<td>1,992</td>
<td></td>
<td>1.03 (0.90-1.18)</td>
<td></td>
</tr>
<tr>
<td>rs4673896</td>
<td>0.41</td>
<td>CC</td>
<td>1,589</td>
<td>1,492</td>
<td>1*</td>
<td>1.06 (0.97-1.16)</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA</td>
<td>2,132</td>
<td>2,121</td>
<td></td>
<td>0.99 (0.93-1.05)</td>
<td></td>
</tr>
<tr>
<td>rs6759222</td>
<td>0.24</td>
<td>CC</td>
<td>2,639</td>
<td>2,613</td>
<td>1*</td>
<td>0.91 (0.80-1.04)</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA</td>
<td>1,609</td>
<td>1,482</td>
<td></td>
<td>0.94 (0.87-1.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>258</td>
<td>230</td>
<td></td>
<td>0.81 (0.62-1.05)</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The 13 SNPs genotyped in Set 2 were significant in Set 1 with P trend ≤ 0.10.

*Reference group.
Table 3. Estimated haplotype frequencies and associations in \( NBS1 \) and \( CHEK1 \)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Haplotypes*</th>
<th>Frequency</th>
<th>( P ) trend</th>
<th>OR per allele (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBS1</td>
<td>00010000</td>
<td>0.27</td>
<td>0.65</td>
<td>1.02 (0.93-1.13)</td>
</tr>
<tr>
<td></td>
<td>00001000</td>
<td>0.16</td>
<td><strong>0.02</strong></td>
<td>0.84 (0.74-0.94)</td>
</tr>
<tr>
<td></td>
<td>00010100</td>
<td>0.13</td>
<td>0.82</td>
<td>1.01 (0.89-1.15)</td>
</tr>
<tr>
<td></td>
<td>10010010</td>
<td>0.09</td>
<td>0.90</td>
<td>0.99 (0.96-1.21)</td>
</tr>
<tr>
<td></td>
<td>00010110</td>
<td>0.08</td>
<td>0.67</td>
<td>1.03 (0.89-1.20)</td>
</tr>
<tr>
<td></td>
<td>00010010</td>
<td>0.08</td>
<td>0.31</td>
<td>1.03 (0.93-1.27)</td>
</tr>
<tr>
<td></td>
<td>Rare</td>
<td>0.19</td>
<td>1.08</td>
<td>0.96-1.20</td>
</tr>
</tbody>
</table>

Global score statistic: \( P = 0.11, \chi^2 = 10.39, df = 6 \)

| CHEK1 | 00000000    | 0.39      | 0.05          | 1.09 (1.00-1.19)       |
|       | 1101001    | 0.14      | 0.55          | 0.96 (0.86-1.09)       |
|       | 1100001    | 0.13      | 0.58          | 0.97 (0.85-1.10)       |
|       | 0000100    | 0.13      | 0.94          | 1.00 (0.88-1.13)       |
|       | 1000000    | 0.12      | 0.24          | 0.93 (0.82-1.05)       |
|       | 1010001    | 0.06      | 0.78          | 0.98 (0.82-1.16)       |
|       | Rare       | 0.57      | 0.93          | 0.72-1.20               |

Global score statistic: \( P = 0.55, \chi^2 = 4.94, df = 6 \)

NOTE: All values are per haplotype versus all other haplotypes.
*SNPs are ordered 5’-3’ as in Supplementary Table S2, 0, common allele; 1, rare allele.
† Values in bold type are nominally significant with \( P \leq 0.05 \).

None of the multimarker combinations showing evidence of an association in Set 1 was replicated in the second stage of the study.

Haplotype Associations. We looked for evidence of any association between the common haplotypes derived from the SNPs genotyped and breast cancer risk. In this way, we were able to maximize our chances of detecting any association with poorly tagged or as yet undiscovered SNPs in each gene studied.

There was little evidence of an association between any of the common haplotypes in the 13 genes studied and differences in breast cancer risk (Supplementary Table S3). None of the global score statistics for each gene was significant. In NBS1, a protective effect was observed with the “ancestral” haplotype [carrying none of the minor alleles of the tag SNPs used]; \( h0000000 \) had an OR of 0.84 (95% CI, 0.74-0.94), \( P \) trend = 0.002, and in CHEK1, carrying the ancestral haplotype, \( h0000000 \), conferred an increase in breast cancer risk (OR, 1.09; 95% CI, 1.00-1.19; \( P \) trend = 0.05; Table 3). These results may highlight the association of an as yet unidentified SNP and possible incompleteness of the SNP data available for these genes.

Survival Analysis. The Set 1 case-only data for all tag SNPs were analyzed to look for any association with long-term prognosis (full data not shown). Vital status was available for all 2,270 Set 1 cases and the median time of follow-up was 7.75 years, with a range of between <1 and 10 years. The total person-years at risk were 13,851, with an all-cause mortality of 359 deaths over this time (16% of Set 1).

Four SNPs in three different genes showed some evidence of association with survival (\( P \) trend \( \leq 0.05 \) and were genotyped in Set 2 (Table 4). Follow-up data were available for all 2,200 Set 2 cases, with a broadly similar median and range to Set 1. The total person-years at risk for the full data set were 25,049, with an all-cause mortality of 639 deaths over this time (14% of Set 1 + Set 2).

In the full study, the rare variants of \( BRIP1 \) SNP rs2191249 and \( RAD51C \) rs16943176 were associated with an increased risk of death per minor allele carried compared with the common homozygote (\( P \) trend = 0.002 and 0.03, respectively; Fig. 1). Neither of these effects was attenuated after adjustment in a multivariate model for age, stage, and grade (data not shown).

Receptor Status. We examined the association of all the tag SNPs studied in Set 1 with risk of specifically developing tumors that were positive or negative for either ER or PR (full data not shown). ER and PR status was available for 1,383 and 900 Set 1 cases, respectively.

Seven SNPs showed evidence of an association with risk in one of the ER status subgroups and nine SNPs exhibited some effect on breast cancer risk dependent on PR status (\( P \) trend \( \leq 0.05 \)). These 16 SNPs were genotyped in Set 2 and the combined data were subdivided and analyzed as with Set 1 (Supplementary Table S4). ER and PR status was available for 2,698 and 1,596 cases, respectively, for the combined data set.

The \( RAD50 \) rs1777256 SNP had no overall effect on breast cancer risk, but when cases were divided with respect to ER status, carrying the minor allele of this SNP leads to a significantly increased risk of developing an ER+ tumor (\( P \) trend = 0.01) but had no effect on the risk of ER- disease. Similarly, the rare alleles of two \( BARD1 \) SNPs, rs1048108 and rs2229571, were associated with risk of PR- tumors (\( P \) trend = 0.02 and 0.01, respectively).

Subgroup analysis of the \( XRCC2 \) rs3218536 SNP, found to be associated with breast cancer risk in the full tagging study, showed that the minor allele confers significant protection against developing PR+ tumors (OR, 0.78; 95% CI, 0.66-0.91; \( P \) trend = 0.002) but had no

Table 4. Survival analyses—associated SNPs

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>MAP*</th>
<th>Base change</th>
<th>Set 1</th>
<th>Set 1 + Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HR (95% CI)</td>
<td>( P ) trend</td>
</tr>
<tr>
<td>RAD54</td>
<td>rs9793263</td>
<td>0.41</td>
<td>G vs A</td>
<td>1.20 (1.03-1.40)</td>
<td>0.02</td>
</tr>
<tr>
<td>RAD51C</td>
<td>rs9916423</td>
<td>0.06</td>
<td>T vs C</td>
<td>0.67 (0.46-0.96)</td>
<td>0.04</td>
</tr>
<tr>
<td>RAD51C</td>
<td>rs16943176</td>
<td>0.21</td>
<td>C vs G</td>
<td>1.25 (1.04-1.46)</td>
<td>0.02</td>
</tr>
<tr>
<td>BRIP1</td>
<td>rs2191249</td>
<td>0.26</td>
<td>C vs A</td>
<td>1.18 (1.00-1.39)</td>
<td>0.04</td>
</tr>
</tbody>
</table>

NOTE: The four SNPs genotyped in Set 2 were significant in Set 1 with \( P \) trend \( \leq 0.05 \).
*Common versus rare in SEARCH.
† HR (95% CI) per copy of the rare allele.
effect on the risk of developing PR− tumors. Neither of the two SNPs associated with survival showed any susceptibility to breast cancer in these subgroups analyses.

**Discussion**

The DNA double-strand break repair genes are perennial candidates for investigation with respect to cancer susceptibility and all known highly penetrant breast cancer–causing mutations are in genes within this pathway. Here, we have attempted a comprehensive SNP tagging study of the genes involved in homologous recombination. Using 100 single variants, we have been able to tag 98% of the 655 known common SNPs in these genes and have evaluated their potential association with breast cancer risk. We have found only one borderline association of the SNPs tested: XRCC2 rs3218536, R188H, $P_{\text{trend}}=0.03$. All $P$ values were unadjusted for multiple testing.

The R188H mutation has been studied on numerous occasions (29, 43-45). Recently, a large meta-analysis (17), which included the data presented here, found no overall association between this SNP and a risk of breast cancer ($P_{\text{trend}}=0.33$), thus leading us to conclude that the borderline effect seen here either is a false positive or may be the diluted effect of an underlying larger subgroup association.

None of the previously reported SNP associations with breast cancer susceptibility has been confirmed by this study. The widely yet conflictingly reported XRCC3 T241M mutation (rs861539; refs. 29, 45-49) was not significantly associated in this case-control comparison or in the recently published meta-analysis, including our study and eight others (17).

Other previously reported SNPs, RAD52 rs4987208 (Y415X; ref. 40), RAD52 rs4987207 (S346T; ref. 41), and RAD51 rs1801320 (50, 51), were not replicated in this study. Several publications document genetic association studies of the NBS1 E185Q coding variant (43, 51, 52). Whereas we have not typed this variant itself, we have perfectly tagged it ($r^2_p=1.0$) by genotyping rs6470524, and no association was evident with this tag.

This study was well powered to detect main effects but larger numbers with survival and receptor status would strengthen the subgroup analyses. The reason our study failed to confirm the findings of the existing literature was because these were generally much smaller sample numbers and likely to be false positives, as indicated by the Breast Cancer Association Consortium study (17). The point estimates of the smaller studies often fell within the 95% CIs achieved by our study, and although these prior publications had effects in the same direction as SEARCH, it is unlikely that the true effect is outside our CIs.

The ancestral haplotypes of NBS1 and CHEK1 also showed some evidence of association with breast cancer risk and this may bring us onto the question of efficacy of the study: How certain are we that we have evaluated all the common SNPs and haplotypes? As the tagging record shows in Table 1, we have achieved our goals to

**Figure 1.** Kaplan-Meier curves describing the effect on patient survival with time for the three possible genotypes for the two SNPs associated with long-term prognosis. Traces are notated with genotype followed by number of deaths/total genotype carriers.

![Kaplan-Meier curves](image-url)
varying yet sufficiently in-depth extents, so we can be confident that we have either directly genotyped or comprehensively tagged any other known, previously published common SNPs in these genes. We have studied both of the genes with associated ancestral haplotypes particularly thoroughly: 95% of the common SNPs in NBS1 were tagged with \( r^2 > 0.8 \) and 97% of the known common variants in CHEK1 were covered. The study has fulfilled its aim to tag all the known variation in the data source used, in both cases, the resequencing data of the NIEHS GeneSNPs project. As detailed in Table 1, these genes have been resequenced to a reasonable extent, 69% coverage for NBS1 and 75% coverage for CHEK1, with the missing regions for both genes tending to be the central portions of large introns. These ancestral haplotype associations indicate that there are unknown variants not found in our tagging data source that we may not have managed to tag sufficiently well. We have not been able to explain all the unknown variation as comprehensively as we have explained the known variation. Due to lack of resequencing data in the intronic regions, we may not have tagged SNPs involved in splicing or receptor binding rather than missing a nonsynonymous association in the coding regions of the genes.

Another explanation of our ancestral haplotype association may be the presence of much rarer alleles with a dominant effect present on that background. We cannot be certain we have excluded such mutations, with MAF < 0.05, from association as we do not have sufficient statistical power to do so (Supplementary Table S1B). Similarly, our power to detect SNPs with MAF = 0.05 to 0.25 acting in a recessive manner is also a limitation of this style of study. There may also be SNPs we have not tagged at all in each of the genes studied. These ‘singleton’ variants are not correlated with any other common SNPs in the region and may also appear on a variety of haplotype backgrounds (i.e., are hypermutable). As these SNPs do not share identity by descent with any of the other SNPs studied, they cannot be interrogated by association and so could only confer a change in breast cancer risk if they did so directly and functionally.

Several SNPs proved worthy of investigation in the full study as they were associated with a poorer long-term prognosis (Table 4). The most notable of the survival effects was in the BRIP1 gene, where carrying the rare allele of rs2191249 was indicative of a worse postdiagnosis outcome than for patients carrying the common variants (\( P \) trend = 0.002): HR per minor allele of 1.20 (95% CI, 1.07–1.36; Fig. 1).

BRIP1 is a DEAH COOH-terminal helicase that interacts directly with the COOH-terminal BRCT repeats of BRCA1, mediating its response to cellular trauma and aiding genome surveillance through the “RAD50 complex” (53). The BRIP1 gene has been shown to contain rare truncating mutations associated with breast cancer susceptibility and Fanconi’s anemia (27, 54), a chromosome instability disorder partly characterized by a predisposition to cancer, so it is notable that we now have evidence of common variants affecting cancer outcome.

Overall mortality was used as an end point for the survival analysis as these data were available on all the SEARCH cases. Mortality due to breast cancer was reported in 597 of the 716 total deaths. However, actual cause of mortality is often misreported on death certificates, especially in older participants. The study was not originally designed for survival analyses; we have reliable all-cause mortality data and have used it but follow-up in better designed studies would be important in confirming our findings.

Stratification by hormone receptor status resulted in several SNPs showing borderline associations in ER or PR subgroups, despite having exhibited little evidence for direct breast cancer susceptibility (Supplementary Table S4). The most marked was the XRCC2 coding SNP, rs3218536, which was of borderline significance (\( P = 0.03 \)) in the full case-control study but showed a much stronger association with the risk of specifically developing PR+ tumors (OR per rare histidine allele, 0.78; 95% CI, 0.66–0.92; \( P = 0.002 \); PR+ versus PR− \( P \) difference = 0.13). Thus, individuals carrying the common arginine allele would be at a significantly increased risk of developing a receptor-positive tumor, whereas their risk of developing receptor-negative tumors would be unchanged. There is a 93% correlation between PR+ status and ER+ status in our cases, and thus, the \( P \) value for risk of developing a “receptor-positive” tumor (i.e., ER+ and PR+ tumor) is similar to that for PR+ tumors alone (\( P = 0.001 \)). This SNP could therefore be a marker conferring a natural, inheritable protection to some hormone receptor-positive tumors.

In summary, there is little evidence for the association of any known common SNPs in the 13 genes studied with breast cancer susceptibility. We are well powered to detect modest effects, so it is unlikely that moderate-risk, common, low-penetrance alleles exist in these genes. Further work may be necessary to fully explain any subgroup analyses or survival trends.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

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.

We thank the 9,000 women who took part in this research, the SEARCH (Breast) Study team (Patricia Harrington, Clare Jordan, Hannah Munday, Barbara Perkins, Mitul Shah, and Judy West) and the EPIC management team (Sheila Bingham, Nicholas Day, Kay-Tee Khaw, and Nick Wareham) for their work, and Fiona Blows for doing immunohistochemistry work.

References

Common Single-Nucleotide Polymorphisms in DNA Double-Strand Break Repair Genes and Breast Cancer Risk


Cancer Epidemiol Biomarkers Prev 2008;17:3482-3489.

Updated version
Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/17/12/3482

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2008/12/08/17.12.3482.DC1

Cited articles
This article cites 54 articles, 15 of which you can access for free at:
http://cebp.aacrjournals.org/content/17/12/3482.full#ref-list-1

Citing articles
This article has been cited by 5 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/17/12/3482.full#related-urls

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, use this link
http://cebp.aacrjournals.org/content/17/12/3482.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.