Counting potentially functional variants in \textit{BRCA1}, \textit{BRCA2} and \textit{ATM} predicts breast cancer susceptibility

Nichola Johnson$^{1,†}$, Olivia Fletcher$^{1,†}$, Claire Palles$^1$, Matthew Rudd$^2$, Emily Webb$^2$, Gabrielle Sellick$^2$, Isabel dos Santos Silva$^4$, Valerie McCormack$^4$, Lorna Gibson$^4$, Agnes Fraser$^4$, Angela Leonard$^4$, Clare Gilham$^3$, Sean V. Tavtigian$^5$, Alan Ashworth$^1$, Richard Houliston$^2$ and Julian Peto$^{3,4,*}$

$^1$The Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB, UK, $^2$Section of Cancer Genetics, $^3$Cancer Research UK Epidemiology & Genetics Unit, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK, $^4$Non Communicable Disease Epidemiology Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK and $^5$Genetic Susceptibility Group, International Agency for Research on Cancer, 150 Cours Albert-Thomas, 69372 Lyon Cedex 08, France

Received January 8, 2007; Revised February 20, 2007; Accepted February 28, 2007

Rare inactivating mutations in \textit{BRCA1}, \textit{BRCA2}, \textit{ATM}, \textit{TP53} and \textit{CHEK2} confer relative risks for breast cancer between about 2 and more than 10, but more common variants in these genes are generally considered of little or no clinical significance. Under the polygenic model for breast cancer carriers of multiple low-penetrance alleles are at high risk, but few such alleles have been reliably identified. We analysed 1037 potentially functional single nucleotide polymorphisms (SNPs) in candidate cancer genes in 473 women with two primary breast cancers and 2463 controls. Twenty-five of these SNPs were in \textit{BRCA1}, \textit{BRCA2}, \textit{ATM}, \textit{TP53} and \textit{CHEK2}. Among the 1037 SNPs there were a few significant findings, but hardly more than would be expected in this large experiment. There was, however, a significant trend in risk with increasing numbers of variant alleles for the 25 SNPs in \textit{BRCA1}, \textit{BRCA2}, \textit{ATM}, \textit{TP53} and \textit{CHEK2}. Among the 1037 SNPs there were a few significant findings, but hardly more than would be expected in this large experiment. There was, however, a significant trend in risk with increasing numbers of variant alleles for the 25 SNPs in \textit{BRCA1}, \textit{BRCA2}, \textit{ATM}, \textit{TP53} and \textit{CHEK2} ($P_{\text{trend}} = 0.005$). For the 21 of these with minor allele frequency $<$10% this trend was highly significant ($P_{\text{trend}} = 0.00004$, odds ratio for 3 or more SNPs $= 2.90$, 95% CI $1.69$–$4.97$). The individual effects of most of these risk alleles were undetectably small even in this well powered study, but the risk conferred by multiple variants is readily detectable and makes a substantial contribution to susceptibility. A risk score incorporating a suitably weighted sum of all potentially functional variants in these and a few other candidate genes may provide clinically useful identification of women at high genetic risk.

INTRODUCTION

A large proportion of all breast cancers arise in a genetically susceptible minority of women (1), most of whom are not carriers of \textit{BRCA1} or \textit{BRCA2} (2–4). The variation in familial breast cancer risk unaccounted for by high-risk mutations in known genes can be explained by a polygenic model in which a large number of ‘low-penetrance’ genes that confer small risks individually act in combination to cause wide variation in risk in the population (5). Many reported associations for individual single nucleotide polymorphisms (SNPs) have not been replicated in subsequent studies (6–8) but for a minority the evidence is consistent (9). Several whole genome association studies to identify common

*To whom correspondence should be addressed at: Non Communicable Disease Epidemiology Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK. Tel: +1 02079272455; Fax: +1 02075806897; Email: julian.peto@lshtm.ac.uk

$^†$The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author 2007. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
haplotypes associated with breast cancer risk are currently underway, but such studies have low power to detect the effects of low-penetrance alleles with minor allele frequency (MAF) less than ~10%.

We have analysed 1037 non-synonymous SNPs in 2463 controls and 473 breast cancer cases with two primary breast cancers. Association studies of women with two primary breast cancers are at least four times as efficient as those using unselected cases, and the more extreme odds ratios (ORs) measured in such studies are less likely to be due to the combined effects of chance and bias. Twenty-five of our SNPs were within five genes (BRCA1, BRCA2, ATM, TP53 and CHEK2) in which rare inactivating mutations have been unequivocally associated with breast cancer risk (3,10,11). No SNPs were included in PTEN, or in two further genes (BRIP1 and PALB2) for which similar results have recently been reported (12–14). We have examined the combined effects of these 25 SNPs both overall and within each gene.

RESULTS

Figure 1 shows the MAF OR for cases with two primary breast cancers and minor allele frequencies for 1037 SNPs, with 5% (grey) and 1% (black) significance limits. The y-axis shows the MAF (%) and the x-axis shows the OR for two primary breast cancers (upper scale) and a first primary breast cancer (lower scale). The OR for a first primary breast cancer is calculated as the square root of the observed OR in women with two primaries.

Figure 1. Observed ORs for two primary breast cancers and minor allele frequencies for 1037 SNPs, with 5% (grey) and 1% (black) significance limits. The y-axis shows the MAF (%) and the x-axis shows the OR for two primary breast cancers (upper scale) and a first primary breast cancer (lower scale). The OR for a first primary breast cancer is calculated as the square root of the observed OR in women with two primaries.

OR per SNP = 1.08, 95% CI 1.02–1.14, \( P_{\text{trend}} = 0.005 \). The OR per SNP was 1.07 (95% CI 1.00–1.14, \( P_{\text{trend}} = 0.04 \)) for BRCA1, 1.08 (95% CI 0.94–1.24, \( P_{\text{trend}} = 0.27 \)) for BRCA2 and 1.27 (95% CI 1.04–1.56, \( P_{\text{trend}} = 0.02 \)) for ATM.

There is a highly significant trend with the number of uncommon (MAF < 10%) SNPs carried (Table 2: OR per uncommon SNP = 1.30, 95% CI 1.15–1.47, \( P_{\text{trend}} = 0.00004 \)) and the OR for 3 or more uncommon SNPs is 2.90 (95% CI 1.69–4.97 \( P = 0.0001 \)). The two common (MAF > 10%) variants in BRCA1 (K1183R and S1613G) are in strong linkage disequilibrium (16) so most homozygotes carry at least four variant alleles. The homozygote OR of ~1.4 for these two BRCA1 SNPs is thus diluted in the trend based on all four common SNPs (OR per common SNP = 1.04, \( P_{\text{trend}} = 0.25 \)). Over half (51.5%) of cases carried one or more of the 21 uncommon SNPs (Table 2: OR for one or more uncommon SNPs = 1.37, 95% CI 1.11–1.68 \( P = 0.002 \)), implying an attributable fraction of cases with two primaries due to these 21 SNPs of 14%, or ~7% of unselected breast cancers. Table 2 suggests a much higher attributable fraction when common SNPs are included, but the trend is irregular, and only 24 cases carried none.

The marked trend in Table 2 is due almost entirely to uncommon SNPs in BRCA1, BRCA2 and ATM. The only CHEK2 variant analysed (I157T) is too rare to appreciably affect our results, and the single TP53 SNP (P72R) was ‘common’ (MAF 26.3%). Table 3 shows each of the SNPs carried by the 21 cases with three or more uncommon SNPs. No single SNP makes a dominant contribution to the overall trend in Table 2. BRCA1 Q356R (MAF 5.4%) has the highest heterozygote OR and is one of the more common of our ‘uncommon’ (MAF < 10%) SNPs. When this SNP is removed from the trend analysis, the OR per uncommon SNP is still significant (OR per SNP = 1.19, \( P_{\text{trend}} = 0.01 \)) and the OR for 3 or more uncommon SNPs is slightly increased (OR for 3 or more SNPs = 3.15, \( P = 0.001 \)). The trend in Table 2 also remains highly significant (\( P_{\text{trend}} = 0.0004 \) or less) when any one of the other 20 uncommon (MAF < 10%) SNPs is removed.
DISCUSSION

Our data suggest that many of the low-penetrance ‘polygenes’ that are thought to underlie familial breast cancer are non-synonymous SNPs in known susceptibility genes whose individual effects are too small to be detected even in well-powered association studies. The relative risks conferred by rare inactivating mutations are of the order of 10 for BRCA1, BRCA2 and TP53 and about two for ATM, CHEK2.
**Table 2.** ORs for two primary breast cancers corresponding to increasing numbers of variant alleles for 25 SNPs in BRCA1, BRCA2, ATM, CHEK2 and TP53

<table>
<thead>
<tr>
<th>No of variant alleles</th>
<th>All SNPs</th>
<th>Uncommon (MAF &lt; 10%) SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
<td>Cases OR (95% CI)</td>
</tr>
<tr>
<td>0</td>
<td>193 (7.91)</td>
<td>24 (5.15) 1.00 (Ref)</td>
</tr>
<tr>
<td>1</td>
<td>369 (15.12)</td>
<td>67 (14.38) 1.46 (0.89–2.40)</td>
</tr>
<tr>
<td>2</td>
<td>493 (20.20)</td>
<td>85 (18.24) 1.39 (0.86–2.25)</td>
</tr>
<tr>
<td>3</td>
<td>516 (21.15)</td>
<td>112 (24.03) 1.75 (1.09–2.80)</td>
</tr>
<tr>
<td>4</td>
<td>392 (16.07)</td>
<td>76 (16.31) 1.56 (0.95–2.55)</td>
</tr>
<tr>
<td>5</td>
<td>252 (10.33)</td>
<td>41 (8.80) 1.31 (0.76–2.24)</td>
</tr>
<tr>
<td>6</td>
<td>140 (5.74)</td>
<td>32 (6.87) 1.84 (1.04–3.26)</td>
</tr>
<tr>
<td>7</td>
<td>65 (2.66)</td>
<td>17 (3.65) 2.10 (1.06–4.16)</td>
</tr>
<tr>
<td>8</td>
<td>16 (0.66)</td>
<td>8 (1.72) 4.02 (1.56–10.38)</td>
</tr>
<tr>
<td>9+</td>
<td>4 (0.16)</td>
<td>4 (0.86) 8.04 (1.89–34.26)</td>
</tr>
<tr>
<td>Total</td>
<td>2440 (100)</td>
<td>466 (100) 1.08 (1.02–1.14) per SNP</td>
</tr>
</tbody>
</table>

**Table 3.** Individual SNPs carried by each of the 21 cases with two primary breast cancers with three or more uncommon (MAF < 10%) SNPs

<table>
<thead>
<tr>
<th>No of SNPs</th>
<th>BRCA1</th>
<th>BRCA2</th>
<th>ATM</th>
<th>CHEK2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>D693N × 2</td>
<td>T1915M</td>
<td>F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Q356R, D693N</td>
<td>N289H, T1915M</td>
<td>F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D693N</td>
<td>N289H, T1915M</td>
<td>L1420F, F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q356R × 2</td>
<td>T1915M, E2856A</td>
<td>L1420F</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D693N</td>
<td>T1915M</td>
<td>P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D693N</td>
<td>N289H, R2034C</td>
<td>S707P, L1420F</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q356R</td>
<td>T1915M, E2856A</td>
<td>S707P</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q356R</td>
<td>T1915M</td>
<td>F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q356R</td>
<td>T1915M</td>
<td>F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q356R</td>
<td>T1915M</td>
<td>F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q356R</td>
<td>T1915M</td>
<td>F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q356R</td>
<td>T1915M</td>
<td>F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q356R</td>
<td>T1915M</td>
<td>F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Q356R</td>
<td>T1915M</td>
<td>F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D693N</td>
<td>N289H</td>
<td>P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D693N</td>
<td>N289H</td>
<td>F858L, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D693N</td>
<td>N289H</td>
<td>S707P, P1054R</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D693N</td>
<td>N289H</td>
<td>S707P, P1054R</td>
<td></td>
</tr>
</tbody>
</table>

**BRIP1 and PALB2** (2,10–13) but several authors have concluded that polymorphic variants in these genes contribute little, if anything, to the identification of women at substantial risk (11–13). Our relative risk estimate of 2.90 (95% CI 1.69–4.97) for bilateral breast cancer in carriers of three or more uncommon (MAF < 10%) alleles in these genes corresponds to a relative risk of 1.7 for unselected cases. This is slightly lower than the range (2.0–2.4) for carriers of inactivating mutations in **CHEK**, **ATM**, **BRIP1** and **PALB2** (10–13).

The trend statistic in Table 2 based on all 25 SNPs tests the composite hypothesis that functional SNPs in some or all of these five genes act cumulatively to increase the risk of breast cancer. By restricting our analysis to these five genes we have avoided the multiple testing and resulting insensitivity of genome-wide analysis of individual variants or haplotypes (i) by counting SNPs only in the five genes in which inactivating mutations are known to contribute to breast cancer susceptibility [two more such genes, **BRIP1** and **PALB2** have now been identified (12,13), but no SNPs in these were included in our screen]; (ii) by including all genotyped SNPs in these genes and (iii) by observing an effect which is much larger for multiple variants than for individual SNPs. The overall significance level ($P_{\text{trend}} = 0.005$) for the total number of SNPs thus constitutes quite strong evidence for a gene-dosage effect. The much more significant ($P_{\text{trend}} = 0.00004$) trend for the number of uncommon SNPs (MAF < 10%) constitutes our only subgroup analysis, and is consistent with the evolutionary argument that rarer variants will on average confer larger risks. In contrast to other multi-genic studies referenced in (17) we included all genotyped SNPs within each gene and ignored linkage disequilibrium, as our underlying assumption is that each potentially functional SNP in these genes may have some effect.

Our data are compatible with a polygenic model in which individual alleles each of which has a small effect combine either additively or multiplicatively to produce much larger risks in carriers of multiple risk alleles. The multiplicative polygenic model fitted by Pharoah et al. (5) predicts that half of all breast cancers occur in the 12% of women at highest risk. All such women could be identified if all the relevant genes were known, but knowing half of them (and a smaller proportion if those that are commoner or cause higher risks are discovered first) will identify 20% of women accounting for half the cases (5). The lower scale in Figure 1 shows ORs for a first primary, calculated as the square root of the second primary OR (15,18). The OR for a first primary is rarely as high as 1.2 for SNPs with MAF >5%, or 1.4 for SNPs with MAF >1%. In either case, several hundred independent variants would have to be discovered before women at substantial risk could be identified. The OR for two primaries of 1.30 per SNP in Table 2 implies an even lower OR (1.14 per SNP) for a first primary breast cancer (15). Discovering several hundred variants with MAFs of 5% or less conferring such modest ORs would prove impossible if the relevant SNPs were distributed at random throughout the genome. Our study of 473 bilateral breast cancer cases and 2463 controls had equivalent power to a study of more than 3500 unselected breast cancer cases and 3500 controls, but only one of the variants we tested...
(BRCA1 Q356R) would be considered statistically significant after adjustment for multiple testing for the 25 SNPs in these genes, and none if the 1037 SNPs in the full screen were taken into account.

Our bilateral cases are a genetically enriched series, and of the order of 5% of our cases but only about 0.2% of our controls will be carriers of a high-penetrance BRCA1 or BRCA2 mutation (19–21). Our cases have not been tested for penetrant BRCA1 mutations, but the trend in OR in Table 2 is unlikely to be due to linkage disequilibrium between penetrant BRCA mutations and these SNPs, as no single high-risk BRCA mutation is common enough in Caucasians for linkage with the minor allele of a SNP to substantially influence our observed trend.

The 59 SNPs with significantly ($P < 0.05$) increased ORs (Fig. 1) could all be due to chance in this large experiment. Despite this absence of clear non-Bayesian evidence in our data for any individual SNP, our analysis suggests that multiple potentially functional variants in BRCA1, BRCA2 and ATM contribute substantially to breast cancer risk. The observed trend may be largely due to a subset of the SNPs we genotyped, and a suitably weighted sum of all potentially functional variants in these genes should predict individual risk considerably better than our simple summation of these selected SNPs. Variants in BRCA1 are classified as probably neutral or deleterious on the basis of evolutionary conservation, their predicted effects on protein structure and the assumption that deleterious variants in trans confer an embryonic lethal phenotype (22). This approach was not designed to characterize variants that cause risks of 2-fold or less. All five of the SNPs in BRCA1 that we analysed were recently classified as ‘known neutral’ on this basis, and many more remain unclassified (22). Such sequence-based criteria might however be adapted to provide quantitative measures that correlate, albeit imperfectly, with the ORs of less penetrant variants. If so, our crude trend analyses could be refined by multiple regression in larger data sets to estimate coefficients for a small number of parameters such as MAF, Grantham Variation and Grantham Deviation (22) in a variant-specific risk score. These scores would be multiplied by a single coefficient for each gene and added to give a cumulative risk score. The total number of coefficients to be estimated would thus be only slightly greater than the number of genes included in the risk score, irrespective of the number of variants. Only variants conferring ORs too small to be estimated individually need be analysed in this way. Well-characterized variants such as CHEK2*1100delC and high-risk BRCA4 mutations should be assigned their known ORs. The addition of a further candidate gene by including all its variants in the weighted risk score would be evaluated by a likelihood ratio test on only one degree of freedom.

One common SNP in BRCA2 that we studied (N372H) and one uncommon SNP in ATM (S49C) that was not included in our SNP set have been tested individually in pooled analyses of more than 15 000 cases and 15 000 controls by the Breast Cancer Association Consortium (BCAC) (8,9). The ORs for a first primary were 1.01 (0.98–1.05) per allele for BRCA2 N372H and 1.13 (95% CI 0.99–1.30) per allele for ATM S49C. These estimates are consistent with our observed average ORs of 1.04 (95% CI 0.98–1.10) for common SNPs and 1.30 (95% CI 1.15–1.47) for uncommon SNPs [corresponding to ORs of 1.02 (95% CI 0.99–1.05) and 1.14 (95% CI 1.07–1.21) respectively for a first primary].

Several known SNPs in BRCA1, BRCA2 and ATM and many rare unclassified variants were not examined, and several inactivating mutations in these genes also increase breast cancer risk (11,22). The combined effects of all functional variants in these and a few further candidate genes such as BRIP1 and PALB2 (12–14) may thus account for a substantial fraction of genetic variation. Identification of all variants in candidate genes by sequencing is complementary to whole-genome analysis of common (MAF > 10%) tagging SNPs, which identifies common variants or haplotypes conferring small individual risks but cannot detect the cumulative effect of multiple variants in individual genes. Clinically useful identification of women at high risk is likely to require the combined findings of both approaches. Our observation needs to be confirmed in larger series but if our estimate of the population attributable fraction of unselected breast cancers (7%) due to this subset of uncommon missense variants is approximately correct, the contribution of missense variants in these and other relevant genes may exceed the combined effects of their inactivating mutations.

MATERIALS AND METHODS

Study population

Cases. Breast cancer cases with two primary breast cancers (416 bilateral cases and 57 asynchronous ipsilateral cases) were ascertained through the English cancer registries. Eligible cases had two sequential or simultaneous primary breast cancer registrations confirmed by interview as separate cancers. All cases had their first breast cancer diagnosed after 1 January 1971 and before age 65. The median age at diagnosis of first primary was 49 year (range 26–65).

Controls. This analysis was part of a collaboration in which cases for three other cancers (colorectal cancer, lung cancer and chronic lymphocytic leukaemia) were analysed by comparison with a common set of controls (23–25). Two thousand four hundred and seventy-five controls who were friends or non-blood relatives of patients with malignancies were recruited through the cases as part of the British Breast Cancer study ($n = 367$), ongoing National Cancer Research Network genetic epidemiological studies ($n = 1075$) or the Royal Marsden Hospital Trust/Institute of Cancer Research Family History and DNA Registry ($n = 1033$). None of the controls had a personal history of malignancy at the time of ascertainment. Cases and controls were recruited between 1 January 1999 and 31 December 2004, and all were Caucasians resident in the UK. Written informed consent was obtained from all participants, and the study was approved by the South East Multicentre Research Ethics Committee.

SNP selection

Full details of the methods for SNP selection have been published elsewhere (26). Briefly, candidate cancer genes were identified by interrogating the Gene Ontology Consortium
database (http://www.geneontology.org) (27), Kyoto Encyclopedia of Genes and Genomes database (http://www.genome.jp/kegg) (28), Stratagene’s Interaction Explorer Pathway Assist Program (http://www.iobion.com/news/hotnews.html?cmd=Retrieve&dopt=Abstract), National Center for Biotechnology Information (NCBI) Entrez Gene database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene) (29) and the CancerGene database (http://caroll.vjf.cnrs.fr/cancergene/HOME.html). A total of 9537 validated nsSNPs with MAF data were identified within 21506 LocusLink annotated genes in NCBI dbSNP Build 123 (http://www.ncbi.nlm.nih.gov/SNP/). Filtering this list and linking it to 7080 candidate cancer genes yielded 3666 validated nsSNPs with MAF > 0.01 in Caucasian populations. The functional impact of nsSNPs was predicted using the Grantham matrix (30) and the in-silico computational tools PolyPhen (http://www.bork.embl-heidelberg.de/PolyPhen/) (31) and SIFT (version 2.1; http://blocks.fhcrc.org/sift/SIFT.html) (32). Using the Predicted Impact of Coding SNPs (PICS) database (http://www.icr.ac.uk/cancgen/molgen/MolPopGen_PICS_database.html) and published work on resequencing of DNA repair genes (33–37) we prioritized a set of 1476 nsSNPs for the current study. Full details of the 40 SNPs that were in BRCA1, BRCA2 and ATM are given in Supplementary Material, Table S5 (online). As we had an a priori interest in the IGF axis and the oestrogen synthesis and metabolism pathways, an additional 60 SNPs that lay within 2 kb of the transcriptional startsite of a gene within these pathways were included. These 1336 SNPs lay within or proximal to 1017 candidate cancer genes. For those SNPs yet to be documented in the latest release of NCBI dbSNP (Build 125), we have submitted complete genotype information including MAF to NCBI and assigned the resultant dbSNP ‘ss’ designations accordingly. Annotated flanking sequence information for each SNP was derived from the University of California Santa Cruz (UCSC) Human Genome Browser (Assembly hg17; http://genome.ucsc.edu/cgi-bin/hgGateway).

Genotyping
DNA was extracted from blood samples using conventional methodologies and quantified using PicoGreen (Invitrogen). Genotyping of samples was performed using customized Illumina Sentrix Bead Arrays according to the manufacturer’s protocols. DNA samples with GenCall scores <0.25 at any locus were considered ‘no calls’. A DNA sample was deemed to have failed if it generated genotypes at <95% of loci (12 samples). SNPs were excluded if <99% of DNA samples generated a genotype at the locus (290 SNPs).

Statistical methods
Each SNP was tested for departure from Hardy Weinberg equilibrium in each of the control populations and in all controls combined. Of the 1536 selected SNPs, 290 that were genotyped in <99% of samples and a further 200 that showed no variation in these cases or controls were excluded. Nine of the remaining 1046 SNPs showed significant departure from Hardy Weinberg equilibrium ($P \leq 0.001$) and were also excluded, leaving 1037 SNPs in the final analysis. Twenty-five of these SNPs (24 missense, one nonsense) were within the five main genes linked to breast cancer susceptibility (3) (5 in BRCA1, 9 in BRCA2, 9 in ATM, 1 in TP53, 1 in CHEK2). Multiple SNP analyses are restricted to the 466 (98.5%) cases and 2440 (99.1%) controls in whom these 25 SNPs were all successfully genotyped. Unadjusted ORs with two-sided exact $P$-values and likelihood ratio trend tests were calculated using Stata 9 (Stata Corporation, TX, USA).

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS
We thank Amie Deffenbaugh and Lynn Anne Burbridge for queries of Myriad Genetic Laboratories’ BRACAnalysis database, and the English cancer registries through whom we ascertained the women with two primary breast cancers. The study was funded by Cancer Research UK and Breakthrough Breast Cancer.

Conflict of Interest statement. The authors declare that there are no competing financial interests or other conflicts of interest.

REFERENCES
22. Tavtigian, S.V., Deffenbaugh, A.M., Yin, L., Judkins, T., Scholl, T.,

23. Rudd, M.F., Webb, E.L., Matakidou, A., Sellick, G.S., Williams, R.D.,

    Fletcher, O., Penegar, S., Withey, L., Qureshi, M. et al. (2006) Search
    for low penetrance alleles for colorectal cancer through a scan of 1467
    non-synonymous SNPs in 2575 cases and 2707 controls with validation by
    kin-cohort analysis of 14 704 first-degree relatives. Hum. Mol. Genet.,
    15, 3263–3271.

26. Rudd, M.F., Williams, R.D., Webb, E.L., Schmidt, S., Sellick, G.S. and
    Houlston, R.S. (2005) The predicted impact of coding single nucleotide
    polymorphisms database. Cancer Epidemiol. Biomarkers Prev., 14,
    2598–2604.

27. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry,
    Gene ontology: tool for the unification of biology. The Gene Ontology

    The KEGG resource for deciphering the genome. Nucleic Acids Res., 32,
    D277–D280.


32. Ng, P.C. and Henikoff, S. (2001) Predicting deleterious amino acid

    (2000) Identification of single nucleotide polymorphisms in human DNA

34. Kuschel, B., Auranen, A., McBride, S., Novik, K.L., Antoniou, A.,
    et al. (2002) Variants in DNA double-strand break repair genes and breast

    Identification of 127 amino acid substitution variants in screening 37
    DNA repair genes in humans. Cancer Epidemiol. Biomarkers Prev., 11,
    1054–1064.

36. Fearnhead, N.S., Wilding, J.L., Winney, B., Tonks, S., Bartlett, S.,
    Bicknell, D.C., Tomlinson, I.P., Mortensen, N.J. and Bodmer, W.F.
    (2004) Multiple rare variants in different genes account for multifactorial
    USA, 101, 15992–15997.

    Identifying functional genetic variants in DNA repair pathway using
    protein conservation analysis. Cancer Epidemiol. Biomarkers Prev., 13,
    801–807.