

Tagging Single-Nucleotide Polymorphisms in Antioxidant Defense Enzymes and Susceptibility to Breast Cancer

Aranca Cebrian,¹ Paul D. Pharoah,¹ Shahana Ahmed,¹ Paula L. Smith,² Craig Luccarini,¹ Robert Luben,³ Karen Redman,² Hannah Munday,¹ Douglas F. Easton,² Alison M. Dunning,¹ and Bruce A.J. Ponder¹

¹Cancer Research UK Human Cancer Genetics Research Group, Department of Oncology, University of Cambridge,

²Cancer Research UK Genetic Epidemiology Group, and ³Department of Public Health and Primary Care, Strangeways Research Laboratories, Cambridge, United Kingdom

Abstract

It is generally believed that the initiation of breast cancer is a consequence of cumulative genetic damage leading to genetic alterations and provoking uncontrolled cellular proliferation and/or aberrant programmed cell death, or apoptosis. Reactive oxygen species have been related to the etiology of cancer as they are known to be mitogenic and therefore capable of tumor promotion. The aim of this study was to assess the role of common variation in 10 polymorphic genes coding for antioxidant defense enzymes in modulating individual susceptibility to breast cancer using a case-control study (N cases = 4,474 and N controls = 4,580). Both cases and controls were from the East Anglian region of the United Kingdom. We have identified a set of 54 single nucleotide polymorphisms (SNPs) that efficiently tag all the known SNPs in the 10 genes and are also expected to tag any unknown SNPs in each gene. We found no evidence for association of common variants in *SOD1*, *SOD2*, *GPX1*, *GPX4*, *GSR*, *TXNRD1*, and *TXN2*. There was borderline evidence for association of variants in *CAT* *g27168a* [P (2 *df*) = 0.05], *TXN* *t2715c* [P (2 *df*) = 0.007], and *TXNRD2* A66S and *TXNRD2* *g23524a* ($P_{\text{trend}} = 0.074$ and 0.046, respectively). For *TXNRD2* A66S [AS versus AA: odds ratio (OR), 1.05; 95% confidence intervals (95% CI), 0.96-1.15; SS versus AA: OR, 1.12; 95% CI, 0.98-1.29], there are bioinformatics data to suggest that it is functional but confirmation in independent data sets is required before they can be regarded as definitive breast cancer susceptibility alleles. Even if confirmed, these four alleles would account for just 0.32% of the excess familial risk of breast cancer. (Cancer Res 2006; 66(2): 1225-33)

Introduction

Breast cancer is the most common cancer among women in industrialized countries (1). A family history is well established as a risk factor for breast cancer (2) and twin studies suggest that most of the excess familial risk is due to inherited factors (3). However, germ line mutations in so-called high-penetrance cancer susceptibility genes, such as *BRCA1* and *BRCA2*, account for <25% of the

excess risk (4). These findings suggest that less penetrant alleles may make a substantial contribution to breast cancer incidence (5).

The molecular mechanisms underlying the development of breast cancer are not well understood. However, it is generally believed that the initiation of breast cancer, like other cancers, is a consequence of cumulative genetic damage leading to genetic alterations that result in activation of proto-oncogenes and inactivation of tumor suppressor genes. These in turn are followed by uncontrolled cellular proliferation and/or aberrant programmed cell death (apoptosis; ref. 6). Reactive oxygen species have been related to the etiology of cancer as they are known to be mitogenic and therefore capable of tumor promotion (7-9). Transient fluctuations in reactive oxygen species serve important regulatory functions, but when present at high and/or sustained levels, reactive oxygen species can cause severe damage to DNA, protein, and lipids. In view of these findings, reactive oxygen species are considered as an important class of carcinogens. The effect of reactive oxygen species is balanced by the antioxidant action of nonenzymatic antioxidants (e.g., glutathione, vitamins A, C, and E, and flavonoids) as well as antioxidant enzymes. A variety of cancer cells are known to exhibit reduced levels of antioxidant enzymes when compared with their normal counterpart (10). In addition, low levels of dietary antioxidants have been hypothesized to be an important determinant of cancer risk. However, although the results of many epidemiologic studies of diet and cancer would be consistent with this hypothesis, direct evidence for an effect of antioxidant levels on cancer risk has been elusive.

There are three main types of antioxidant defense enzymes: the superoxide dismutases (SOD), including manganese-containing SOD (SOD2, also known as MnSOD) and cytosolic CuZnSOD (SOD1), catalase (CAT), and the peroxidases (GPX1 and GPX4; Fig. 1). All of them function to protect the cell from damage due to reactive oxygen species. In addition, several other enzymes are implicated in oxidative damage repair. The reduction of oxidized glutathione (GSSG) produced by action of GPXs is catalyzed by glutathione reductase. Accumulating evidence suggests that, in addition to their "antioxidant" functions, these enzymes participate in cell signaling processes (11). Transforming growth factor β 1 has been shown to suppress the expression of antioxidant enzymes in some cells leading to increased cellular oxidative stress (12, 13). Recently, it has also been shown that some of these enzymes are associated with modification of histone acetylation and histone acetyltransferase activity, both of which have critical roles in eukaryotic gene transcription (14), and others are associated with decreased cell proliferation in vascular endothelial cells (15) and increased proliferation of fibroblasts (16).

The thioredoxins (TXN and TXN2) and thioredoxin reductases (TXNRD1 and TXNRD2) are also involved in antioxidant defense

Note: B.A.J. Ponder is a Gibb Fellow of Cancer Research UK.

Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Requests for reprints: Aranca Cebrian, Strangeways Research Laboratories, Worts Causeway, Cambridge CB1 8RN, United Kingdom. Phone: 44-1223-740-684; Fax: 44-1223-740-147; E-mail: aranca@srl.cam.ac.uk

©2006 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-05-1857

through the thioredoxin redox cycle. These proteins are responsible for mediating numerous cytoplasmic functions and are implicated in control of cell growth in which the redox function is essential for growth stimulation and apoptosis (17–19). TXN is also a key enzyme for DNA synthesis by directly serving as an electron donor to ribonucleotide reductase (20). In addition, the level of TXNRDs in tumor cells is often 10-fold or even greater than in normal tissues and tumor proliferation seems to be crucially dependent on an active thioredoxin system (21, 22).

Despite compelling evidence that oxidative stress is an important mechanism in carcinogenesis and the importance of antioxidant defense enzymes to control the cell redox level and to combat the accumulation of reactive oxygen species, few studies have examined genetic variation in the genes coding for these enzymes and their relationship to cancer risk. Only two genes related to antioxidant defense (*SOD2* and *GPX1*) have been analyzed in genetic association studies of several different types of cancer including lung cancer, breast cancer, colorectal cancer, prostate cancer, and bladder cancer, and the results from these have been contradictory (23–31). Furthermore, these have evaluated only one or two genetic variants [single nucleotide polymorphisms (SNPs)] at each candidate locus.

The aim of this study was to evaluate the association between common variants in 10 genes coding for antioxidant defense enzymes (*SOD1*, *SOD2*, *CAT*, *GPX1*, *GPX4*, *GSR*, *TXN*, *TXN2*, *TXNRD1*, and *TXNRD2*) and susceptibility to breast cancer. We have used a case-control study design and genotyped SNPs which tag all known common variants present in each gene in our population.

Materials and Methods

Patients and controls. Cases were drawn from SEARCH (breast), an ongoing population based study, with cases ascertained through the East Anglian Cancer Registry. All patients diagnosed with invasive breast cancer below age 55 years since 1991 and still alive in 1996 (prevalent cases, 48), together with all those diagnosed <70 years between 1996 and the present (incident cases, median age: 52 years), were eligible to take part. All study participants completed an epidemiologic questionnaire and provided a blood sample for DNA analysis. Sixty-seven percent of eligible breast cancer patients returned a questionnaire and 64% provided a blood sample. Controls were randomly selected from the Norfolk component of European Prospective Investigation of Cancer. European Prospective Investigation of

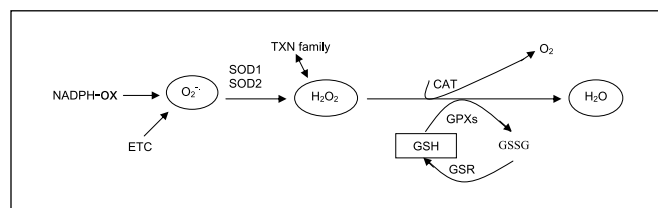


Figure 1. The three main types of antioxidant defense enzymes protect the cell from damage due to reactive oxygen species generated by different processes such as the electron transport chain (ETC) or the plasma membrane-localized NAD(P)H oxidoreductase complex (NADPH-ox). The superoxide dismutases (*SOD1* and *SOD2*) dismutate the superoxide radical into H_2O_2 . The glutathione peroxidases (*GPX*) and catalase (*CAT*) reduce H_2O_2 into water and oxygen. Glutathione redox cycle (*GSH/GSSG*) provides the cell with reduced glutathione (*GSH*) to act as cosubstrate for the peroxidases but to also participate in detoxification reactions and react nonenzymatically with OH^{\cdot} and peroxynitrite. The reduction of GSSG is catalyzed by glutathione reductase (*GSR*) in a process that requires NADPH. Accumulating evidence suggests that, in addition to their antioxidant functions, these enzymes participate in cell signaling processes (11). The thioredoxin (*TXN*) redox cycle is also involved in antioxidant defense. The cycle contains a family of redox-active proteins responsible for mediating numerous cytoplasmic functions and implicated in control of cell growth (19, 21, 22).

Cancer is a prospective study of diet and cancer being carried out in nine European countries. The European Prospective Investigation of Cancer-Norfolk cohort comprises 25,000 individuals, ages 45 to 75 years, resident in Norfolk, East Anglia—the same region from which the cases have been recruited. European Prospective Investigation of Cancer participants were recruited through general practice age-sex registers. Forty-five percent of invited individuals provided a blood sample and took part.

Controls are not matched to cases, but are broadly similar in age, being of ages 42 to 81 years. The ethnic background of both cases and controls as reported on the questionnaires is similar, with >98% being white. The study is approved by the Eastern Region Multicentre Research Ethics Committee and all patients gave written informed consent.

The total number of cases available for analysis was 4,474, of whom 27% were prevalent cases. The samples have been split into two sets to save DNA and reduce genotyping costs: the first set ($n = 2,271$ cases and 2,280 controls) is genotyped for all SNPs and the second set ($n = 2,203$ cases and 2,280 controls) is then tested for those SNPs that show marginally significant associations in set 1 ($P_{\text{heterogeneity}}$ or $P_{\text{trend}} < 0.1$ for univariate analyses). If $P < 0.1$ for comparison of haplotype frequencies, all SNPs within the haplotype block are genotyped in set 2. This staged approach substantially reduces genotyping costs without significantly affecting statistical power. 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 0.95, the staged/full study has 86/88% power to detect a dominant allele with minor allele frequency (MAF) of 0.05 that confers a relative risk of 1.5 or 87/89% power to detect a dominant allele with MAF of 0.25 that confers a relative risk of 1.3. Power to detect recessive alleles is less—53/60% for an allele with MAF of 0.25 and risk 1.5, and 71/75% for an allele with MAF 0.5 and risk 1.3. Cases with high yields of genomic DNA were selected for set 1 from the first 3,500 recruited, with set 2 comprising the remainder of these plus the next 974 incident cases recruited. As the prevalent cases were the first recruited, the proportion of prevalent cases was somewhat higher in set 1 than in set 2 (33% versus 20%). Median age at diagnosis was similar in both sets (51 and 52 years, respectively). Median time from diagnosis to blood draw was slightly longer for set 2 (15 months) than for set 1 (9 months). There was no significant difference in the morphology, histopathologic grade, or clinical stage of the cases by set or by prevalent/incident status. Research Ethics Committee and all patients gave written informed consent.

Selection of tagging SNPs. The principal hypothesis underlying this experiment was that there are one or more common SNPs in the genes of interest that are associated with an altered risk of breast cancer. Thus, the aim of the SNP tagging was to identify a set of SNPs (stSNP) that efficiently tag all the known SNPs. We postulate that such SNPs are also likely to tag any hitherto unidentified SNPs in the gene. The selection of tagging SNPs is most reliable where the gene has been resequenced in a sample of individuals. The National Institute of Environmental Health Sciences (NIEHS) Environmental Genome Project (EGP) Project is currently resequencing all the exons, 5' and 3' untranslated region (UTR), and between 50% to 100% of introns including in all of them the splice sites for candidate genes for cancer across a panel of 90 individuals representative of U.S. ethnicities, including 24 European Americans, 24 African Americans, 12 Mexican Americans, 6 Native Americans, and 24 Asian Americans (PDR90). It is known that there is greater genetic diversity in individuals of African origin (32) but ethnic group identifiers for the PDR90 samples are not available. We have identified 28 of the samples most likely to be African American in this population by comparing the genotypes for the NIHPDR90 samples with the genotypes for the same SNPs from the National Heart, Lung, and Blood Institute Variation Discovery Resource project African American panel.⁴ Data from the remaining 62 individuals were used to identify stSNPs. NIEHS resequencing data were only available for *SOD2*, *CAT*, *GPX1*, *GPX4*, *GSR*, and *TXN*. For the remaining genes (*SOD1*, *TXNRD1*, *TXN2*, and *TXNRD2*), we instead used data from the International MapMap

⁴ http://pga.gs.washington.edu/finished_genes.html.

Table 1. Number of common SNPs and number of tagging SNPs identified in each gene

Gene	Size of gene (kbp)	Database	Average SNP density (SNP/kbp)	No. SNPs	No. LD blocks	No. tag SNPs
<i>SOD1</i> *	9.2	Celera/HapMap	1.85	5	1	—
<i>SOD2</i>	11.0	NIEHS	0.34	32	1	3
<i>CAT</i>	33.1	NIEHS	0.37	89	1	5
<i>GPX1</i>	1.18	NIEHS	0.29	4	1	2
<i>GPX4</i>	3.0	NIEHS	0.21	14	1	4
<i>GSR</i>	48.4	NIEHS	0.73	66	2	LD1 3 LD2 4
<i>TXN</i>	12.5	NIEHS	0.34	37	2	LD1 6 IB 2 LD2 6
<i>TXNRD1</i>	63.3	HapMap	3.90	16	1	4
<i>TXN2</i>	14.6	HapMap	2.08	7	1	3
<i>TXNRD2</i>	66.4	HapMap	2.21	30	3	LD1 3 LD2 4 LD3 3

*Only three SNPs in LD in HapMap and two SNPs more in Celera. We used these five SNPs to choose the variants tagging the common haplotypes for 96 samples of our population. LD1 to LD3, LD blocks 1 to 3, respectively; IB, interblock.

Project (01-03-2005: HapMap last public release used in this study), which has genotyped a large number of SNPs in 30 parent-offspring trios. These samples were collected in 1980 from U.S. residents with northern and western European ancestry by the Centre d'Etude du Polymorphisme Humain. In the case of *SOD1*, only three HapMap SNPs were available, and we therefore used Applied Biosystems (Foster City, CA) SNPbrowser (National Center for Biotechnology Information build 34 genome) to identify additional SNPs within the gene (Table 1).

The best measure of the extent to which a one SNP tags another SNP is the pairwise correlation coefficient (r_p^2) because the loss in power incurred by using a marker SNP in place of a true causal SNP is directly related to this measure. We aimed to define a set of tagging SNPs such that all known "common" SNPs (defined as $MAF > 0.05$) had an estimated r_p^2 of >0.8 with at least one tagging SNP. However, some SNPs are poorly correlated with other single SNPs but may be efficiently tagged by a haplotype defined by multiple SNPs, thus reducing the number of tagging SNPs needed. As an alternative, therefore, we aimed for the correlation between each SNP and a haplotype of tagging SNPs (r_s^2) to be at >0.8 . In this article, we have used r_s^2 as the main criterion to determine tag SNPs but have also presented the tagging efficiency in terms of r_p^2 . Using this design, and assuming a minimum r^2 of 0.8, this study had $>85\%$ power to detect, at a significance level of $P < 0.0001$, any dominant susceptibility allele with a frequency of 5% or greater conferring a relative risk of at least 1.4 or a recessive allele with frequency 10% or greater conferring a relative risk of at least 2.

Because tagging SNP selection is problematic when there is extensive haplotype diversity, where necessary, we divided a gene into haplotype blocks and selected the stSNPs for each block separately. It is possible to use a variety of formal definitions of haplotype blocks but we simply used the graphical representations of the pattern of linkage disequilibrium (LD) based on D' and selected blocks such that the common haplotypes in each block accounted for at least 80% of all haplotypes observed using the Haploview program (33).

Genotyping. We genotyped all samples for the selected tag SNPs using the ABI PRISM 7900 sequence detection system or TaqMan (Applied Biosystems). We carried out PCR on DNA (10 ng) using TaqMan universal PCR master mix (Applied Biosystems), forward and reverse primers, and FAM- and VIC-labeled probes designed by Applied Biosystems (ABI Assay-by-Designs) in a 5- μ L reaction. Sequences of primers and probes are available on request. Amplification conditions on MJ Tetrad thermal cyclers

(Genetic Research Instrumentation, MJ Research, Cambridge, MA) were as follows: 1 cycle of 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. We read the completed PCRs on an ABI PRISM 7900 Sequence Detector in end point mode using the Allelic Discrimination Sequence Detector Software (Applied Biosystems). For the software to recognize the genotypes, we included two nontemplate controls in each 384-well plate. Cases and controls were arrayed together in twelve 384-well plates and a 13th plate contained eight duplicate samples from each of the 12 plates to ensure a good quality of genotyping (the concordances was $>99\%$ for all SNPs). Failed genotypes were not repeated (the rate for failed genotypes did not exceed 8.3% for any of the SNPs under study).

Statistical methods. For each polymorphism, deviation of the genotype frequencies from those expected under Hardy-Weinberg equilibrium was assessed in the controls by a χ^2 test. Genotype frequencies in cases and controls were compared using a χ^2 test with 2 degrees of freedom (2 *df*, $P_{\text{heterogeneity}}$) and the Armitage trend test (χ^2 on 1 *df*) for the trend in breast cancer risk with number of rare alleles (P_{trend}). The relative risks of breast cancer for heterozygotes and for rare homozygotes, relative to common homozygotes, were estimated as odds ratios (OR) with associated 95% confidence intervals (95% CI). Any SNP with a P_{trend} or $P_{\text{heterogeneity}} \leq 0.1$ in set 1 was subsequently genotyped in set 2, and the results were combined to test their association with breast cancer in the U.K. population.

Haplotype frequencies were estimated and compared in cases and controls using the estimation-maximization algorithm implemented in the HaploScore program (34). Haplotypes with a frequency of <0.05 were pooled. The HaploScore program computes score statistics (and hence significance levels) to test for associations between individual haplotypes and disease status along with the global test of association.

The potential phenotypic effect of specific SNPs was examined using Putative Phenotypic Alterations caused by SNPs (PupaSNP).⁵ This is a web-based search tool for SNPs with potential phenotypic effect at transcriptional level. PupaSNP inputs lists of genes (or generates them from chromosomal coordinates) and retrieves SNPs that could affect conserved regions that the cellular machinery uses for the correct processing of genes

⁵ <http://pupasnp.bioinfo.cnio.es>.

Table 2. Breast cancer risks associated with tagging SNPs, genotyped in set 1, in the 10 genes encoding antioxidant enzymes included in this study

Gene and SNPs*	MAF	OR [95% CI] (heterozygotes vs common homozygotes)	OR [95% CI] (rare homozygotes vs common homozygotes)	<i>P</i> (2 <i>df</i>) ^{†, ‡}	<i>P</i> _{trend} (1 <i>df</i>) [‡]
<i>SOD1</i>					
rs202445, a-6366g	0.17	0.96 [0.84-1.09]	1.30 [0.91-1.85]	0.26	0.79
rs4998557, g2809a	0.12	0.97 [0.84-1.13]	0.50 [0.26-0.97]	0.10	0.24
rs2070424, a7237g	0.07	0.91 [0.76-1.09]	0.44 [0.16-1.26]	0.17	0.13
<i>SOD2</i>					
rs1799725, A16V [§] (t>c)	0.49	1.07 [0.93-1.25]	1.01 [0.85-1.20]	0.58	0.86
rs2842958, c5774t	0.21	0.99 [0.87-1.13]	0.99 [0.74-1.33]	0.99	0.90
<i>CAT</i>					
rs1001179, t-330c	0.22	0.91 [0.80-1.03]	1.01 [0.77-1.32]	0.30	0.32
rs1049982, c-20t	0.32	1.08 [0.95-1.23]	0.90 [0.73-1.12]	0.18	0.97
rs769217, D389D (c>t)	0.22	1.09 [0.96-1.23]	0.95 [0.72-1.24]	0.37	0.53
rs511895, g27168a	0.42	0.88 [0.77-1.01]	1.06 [0.89-1.26]	0.06	0.90
rs7104301, a33077g	0.26	0.96 [0.84-1.09]	1.02 [0.80-1.24]	0.75	0.71
<i>GPXI</i>					
rs3448, a-1040g	0.25	0.98 [0.87-1.11]	1.24 [0.97-1.58]	0.17	0.36
rs1050450, P200L (g>a)	0.31	0.93 [0.83-1.05]	0.96 [0.79-1.18]	0.53	0.40
<i>GPX4</i>					
rs757229, c-1999g	0.46	1.18 [1.03-1.36]	1.15 [0.97-1.36]	0.06	0.08
rs4807542, P12P (g>a)	0.16	1.13 [0.99-1.29]	0.96 [0.68-1.36]	0.18	0.20
rs4807543, a2064c	0.05	0.95 [0.78-1.15]	3.12 [0.84-11.53]	0.17	0.95
rs713041, t2572c	0.46	1.01 [0.88-1.16]	0.81 [0.68-0.96]	0.01	0.03
<i>GSR</i>					
LD1					
rs1002149, t-515g	0.16	1.08 [0.95-1.24]	0.86 [0.59-1.26]	0.33	0.58
rs8190924, g12508a	0.05	1.05 [0.87-1.28]	1.81 [0.53-6.19]	0.56	0.43
rs2978663, a19278g	0.37	0.97 [0.85-1.10]	1.05 [0.87-1.27]	0.64	0.83
LD2					
rs4628224, t28414a	0.19	1.03 [0.90-1.17]	0.93 [0.70-1.23]	0.76	0.97
rs8191009, c39396t	0.20	1.11 [0.98-1.26]	0.94 [0.69-1.29]	0.23	0.32
rs3926402, g43288a	0.39	1.01 [0.89-1.15]	0.94 [0.78-1.12]	0.69	0.62
rs2911678, a45359t	0.22	1.09 [0.96-1.23]	0.97 [0.73-1.28]	0.37	0.42
<i>TXN</i>					
LD1					
rs2301241, a-795g	0.41	1.04 [0.91-1.19]	1.04 [0.87-1.24]	0.82	0.61
rs4135165, g1153a	0.09	1.06 [0.91-1.24]	0.94 [0.49-1.81]	0.73	0.54
rs1410051, a1363g	0.24	1.04 [0.91-1.18]	0.94 [0.73-1.22]	0.74	0.94
rs4135168, a1830g	0.26	0.98 [0.86-1.11]	0.83 [0.65-1.06]	0.33	0.23
rs4135172, g2316a	0.08	1.05 [0.89-1.23]	0.88 [0.45-1.72]	0.80	0.73
rs4135179, t2715c	0.23	0.86 [0.76-0.98]	1.24 [0.96-1.61]	0.01	0.64
IB					
rs4135192, g5126a	0.33	0.89 [0.79-1.01]	0.97 [0.79-1.20]	0.21	0.29
rs2418076, g5363a	0.27	1.09 [0.96-1.24]	0.96 [0.76-1.21]	0.30	0.56
LD2					
rs4135208, a9469g	0.33	0.99 [0.87-1.13]	0.96 [0.78-1.18]	0.94	0.75
rs4135211, a10141c	0.07	1.13 [0.95-1.34]	0.85 [0.37-1.93]	0.35	0.26
rs4135215, t11216c	0.16	1.00 [0.87-1.14]	0.84 [0.59-1.21]	0.64	0.60
rs4135221, g11478a	0.13	0.94 [0.82-1.08]	0.94 [0.59-1.51]	0.68	0.40
rs4135225, a12024g	0.33	1.00 [0.88-1.14]	1.06 [0.86-1.31]	0.83	0.69
rs2776, g12496a	0.50	0.94 [0.81-1.09]	0.92 [0.77-1.09]	0.60	0.34
<i>TXNRD1</i>					
rs4964778, g-19040c	0.18	1.02 [0.90-1.16]	0.87 [0.63-1.21]	0.66	0.82
rs4964779, g-10243a	0.12	0.95 [0.82-1.11]	1.09 [0.68-1.74]	0.76	0.72
rs4564401, g-1876a	0.07	1.10 [0.93-1.30]	0.84 [0.39-1.79]	0.49	0.42
rs10861201, t13855g	0.23	1.00 [0.88-1.14]	0.98 [0.76-1.28]	0.99	0.96

(Continued on the following page)

Table 2. Breast cancer risks associated with tagging SNPs, genotyped in set 1, in the 10 genes encoding antioxidant enzymes included in this study (Cont'd)

Gene and SNPs*	MAF	OR [95% CI] (heterozygotes vs common homozygotes)	OR [95% CI] (rare homozygotes vs common homozygotes)	<i>P</i> (2 <i>df</i>) ^{†,‡}	<i>P</i> _{trend} (1 <i>df</i>) [‡]
<i>TXN2</i>					
rs2281082, a4134c	0.19	1.00 [0.88-1.14]	1.06 [0.76-1.47]	0.94	0.83
rs8140110, a8249g	0.09	1.04 [0.89-1.22]	1.53 [0.76-3.12]	0.44	0.36
rs8139906, g8624c	0.18	1.00 [0.88-1.14]	1.23 [0.87-1.75]	0.51	0.52
<i>TXNRD2</i>					
LD1					
rs740603, a-15851g	0.46	1.01 [0.88-1.16]	0.98 [0.83-1.16]	0.92	0.84
rs2020917, a442g	0.28	1.02 [0.90-1.16]	1.15 [0.93-1.43]	0.45	0.27
rs4485648, a9921g	0.20	0.86 [0.76-0.98]	1.12 [0.84-1.50]	0.04	0.24
LD2					
rs5748469, A66S (c>a)	0.33	1.10 [0.98-1.25]	1.25 [1.03-1.52]	0.05	0.02
rs756661, g23524a	0.47	0.85 [0.74-0.97]	0.86 [0.72-1.01]	0.05	0.04
rs732262, g44669a	0.09	1.12 [0.96-1.32]	0.76 [0.41-1.40]	0.22	0.38
rs1548357, t50964c	0.27	1.11 [0.98-1.26]	1.28 [1.03-1.59]	0.04	0.01
LD3					
rs2073750, g55969a	0.21	1.05 [0.93-1.19]	1.00 [0.75-1.32]	0.74	0.59
rs3788306, a57317g	0.30	1.03 [0.91-1.17]	1.04 [0.85-1.29]	0.86	0.59
rs20773752, I370T (a>g)	0.28	0.89 [0.78-1.00]	1.06 [0.84-1.34]	0.10	0.41

*We have given the dbSNP rs number for each SNP, which is the unique identifier and the best standardized naming system; also, the nucleotide position in intronic variants or codon number in coding SNPs is given related to START codon. The common allele appears before the nucleotide position and the rare allele after it.

[†]*P* (2 *df*): genotype frequencies in cases and controls were compared using a χ^2 test with 2 *df*.

[‡]SNPs with *P*_{trend} or *P* (2 *df*) ≤ 0.1 (in bold) were chosen for testing in set 2.

[§]Previously associated with breast and prostate cancer (24, 29).

^{||}Previously associated with lung cancer (25), breast cancer (28), and bladder cancer (31).

(intron/exon boundaries or exonic splicing enhancers), predicted transcription factor binding sites, or which cause changes in amino acids in proteins. The program uses the mapping of SNPs in the genome provided by Ensembl.

Results

Genotype distributions in the controls did not differ significantly from those expected under Hardy-Weinberg equilibrium for any of the SNPs in the 10 genes under study (genotype frequencies for cases and controls genotyped are shown in Supplementary Tables S1 and S2). Table 2 shows MAF for all SNPs in controls, together with the genotype specific risks and results of the significance tests for each SNP. There was no evidence for association of genotype with age in controls and, as expected, age-adjusted risks were similar to unadjusted risks (data not shown). Table 3 shows the estimated haplotype frequencies for each gene in cases and controls with the results of the global test for association and the haplotype specific risks. Table 4 shows the results of the genotyping for the 10 SNPs that were genotyped in both sets 1 and 2.

SOD1. Data for only three SNPs were available in HapMap (rs2070424, rs10432782, and rs1041740) and these were in perfect LD ($r_p^2 = 1$) so that they defined a single pair of haplotypes. We also screened the whole coding region in 96 individuals by dHPLC in to reduce the chance of missing any important functional variants. No further variants were found in any of the five exons.

The Applied Biosystems SNPbrowser enabled us to identify two further SNPs from the Celera database (rs202445 and rs4998557) located at the putative promoter and in intron 1, respectively. We then genotyped all five SNPs in 96 samples from our case series. Four common haplotypes were identified and these were tagged by three of the SNPs with a minimum r_h^2 of 0.98. There was marginal evidence of association for *g2809a* ($P_{\text{heterogeneity}} = 0.099$) in set 1 but no association was observed in the combined data set ($P_{\text{heterogeneity}} = 0.46$).

SOD2. Three stSNPs tagged the 32 common SNPs in the EGP samples with a minimum $r_s^2 = 0.85$, and 18 of 32 SNPs were tagged with $r_p^2 > 0.7$. One of the stSNPs was found to be rare in our population and therefore was excluded from the analysis. No association with breast cancer was found for either SNP, including the coding polymorphism A16V ($P_{\text{heterogeneity}} = 0.58$), which had previously been reported to be associated with breast cancer (24) and prostate cancer (29).

CAT. The 89 common SNPs were tagged by five stSNPs ($r_s^2 = 0.83$; 36 of 89 SNPs were tagged with $r_p^2 > 0.7$). One of these, *g27168a*, showed some evidence of difference in genotype frequency distribution between cases and controls ($P = 0.06$), an effect that became slightly more significant in sets 1 and 2 combined ($P = 0.05$). This effect was due to a slight reduction in risk for *ga* heterozygotes (compared with *gg* homozygotes; OR, 0.94; 95% CI, 0.86-1.03) and a small increased risk in *aa* homozygotes (OR, 1.09; 95% CI, 0.96-1.23).

Table 3. Haplotype frequencies in cases and controls in our hypothesis generating set

Gene	Haplotypes	Control freq	Case freq	OR [95% CI]	Global test
<i>SOD1</i>	aga	0.70	0.71	1	Global-stat = 6.76, <i>df</i> = 4, <i>P</i> = 0.15
	gga	0.17	0.18	1.01 [0.64-1.59]	
	aag	0.07	0.06	0.93 [0.52-1.65]	
	aaa	0.05	0.05	0.91 [0.57-1.46]	
<i>SOD2</i>	tc	0.50	0.50	1	Global-stat = 1.52, <i>df</i> = 3, <i>P</i> = 0.70
	cc	0.29	0.29	1.01 [0.91-1.12]	
	ct	0.21	0.21	1.00 [0.90-1.12]	
<i>CAT</i>	tccgg	0.26	0.26	1	Global-stat = 8.35, <i>df</i> = 5, <i>P</i> = 0.30
	cccaa	0.22	0.21	0.97 [0.86-1.09]	
	tttga	0.22	0.23	1.03 [0.92-1.17]	
	tccaa	0.15	0.16	1.07 [0.94-1.23]	
<i>GPX1</i>	ttcga	0.06	0.05	0.91 [0.74-1.12]	Global-stat = 1.92, <i>df</i> = 3, <i>P</i> = 0.59
	ag	0.42	0.43	1	
	aa	0.31	0.31	0.96 [0.87-1.06]	
	gg	0.26	0.27	1.00 [0.90-1.11]	
<i>GPX4</i>	cgac	0.44	0.41	1	Global-stat = 7.35, <i>df</i> = 4, <i>P</i> = 0.12
	ggat	0.28	0.29	1.07 [0.93-1.22]	
	gaat	0.13	0.15	1.17 [1.03-1.34]	
	cgat	0.06	0.07	1.11 [0.91-1.34]	
<i>GSR</i> <i>LD1</i>	tga	0.58	0.58	1	Global-stat = 1.02, <i>df</i> = 3, <i>P</i> = 0.80
	tgg	0.25	0.24	1.00 [0.91-1.11]	
	ggg	0.12	0.12	1.03 [0.88-1.21]	
<i>LD2</i>	tcaa	0.37	0.36	1	Global-stat = 3.24, <i>df</i> = 4, <i>P</i> = 0.52
	tcgt	0.22	0.22	1.06 [0.96-1.18]	
	acga	0.19	0.18	1.00 [0.90-1.12]	
<i>TXN</i> <i>LD1</i>	ttga	0.18	0.19	1.07 [0.95-1.21]	Global-stat = 4.18, <i>df</i> = 6, <i>P</i> = 0.65
	ggaagt	0.33	0.32	1	
	agaagc	0.20	0.20	1.03 [0.91-1.17]	
<i>LD2</i>	agaggt	0.14	0.14	1.03 [0.90-1.19]	Global-stat = 1.99, <i>df</i> = 6, <i>P</i> = 0.92
	aagagt	0.09	0.10	1.15 [0.97-1.35]	
	aggggt	0.09	0.08	0.92 [0.77-1.09]	
	ggaaat	0.08	0.08	1.03 [0.86-1.23]	
	aatgga	0.32	0.32	1	
	gatgag	0.17	0.17	0.99 [0.87-1.13]	
	aatgaa	0.15	0.14	0.95 [0.81-1.11]	
gacgag	0.14	0.14	0.97 [0.83-1.13]		
<i>TXNRD1</i>	aataag	0.10	0.10	1.02 [0.85-1.21]	Global-stat = 1.31, <i>df</i> = 5, <i>P</i> = 0.93
	actgag	0.07	0.08	1.08 [0.89-1.32]	
	gggt	0.58	0.58	1	
<i>TXN2</i>	cggg	0.17	0.17	1.00 [0.86-1.16]	Global-stat = 1.56, <i>df</i> = 4, <i>P</i> = 0.81
	gagt	0.11	0.11	0.99 [0.70-1.42]	
	ggat	0.07	0.08	1.07 [0.82-1.40]	
	gggg	0.06	0.05	0.99 [0.77-1.27]	
<i>TXNRD2</i> <i>LD1</i>	aag	0.64	0.63	1	Global-stat = 10.93, <i>df</i> = 5, <i>P</i> = 0.09
	cag	0.19	0.19	1.02 [0.92-1.14]	
	agc	0.09	0.10	1.08 [0.93-1.25]	
	aac	0.09	0.08	1.01 [0.86-1.17]	
<i>LD2</i>	gaa	0.42	0.42	1	Global-stat = 10.93, <i>df</i> = 5, <i>P</i> = 0.09
	aga	0.26	0.28	1.04 [0.94-1.16]	
	aag	0.17	0.16	0.96 [0.84-1.08]	
	aaa	0.10	0.10	0.95 [0.82-1.11]	
	cagt	0.46	0.43	1	
	aggc	0.17	0.19	1.17 [1.05-1.31]	
	cggt	0.16	0.15	1.01 [0.90-1.13]	
aggt	0.10	0.11	1.11 [0.96-1.29]		

(Continued on the following page)

Table 3. Haplotype frequencies in cases and controls in our hypothesis generating set (Cont'd)

Gene	Haplotypes	Control freq	Case freq	OR [95% CI]	Global test
<i>LD3</i>	agac	0.05	0.04	1.00 [0.81-1.24]	Global-stat = 1.19, <i>df</i> = 5, <i>P</i> = 0.94
	gga	0.30	0.30	1	
	gaa	0.28	0.28	0.98 [0.89-1.08]	
	gag	0.21	0.20	0.94 [0.83-1.07]	
	aaa	0.14	0.14	1.00 [0.87-1.16]	
	aag	0.07	0.07	1.00 [0.83-1.22]	

GPX1. Four common SNPs were identified from the EGP data, two of which tagged the others with a minimum r_s^2 of 0.89 (all of them were tagged with $r_p^2 > 0.7$). One of these, a nonsynonymous amino acid substitution, P200L (rs1050450), had been previously associated with lung cancer (23), breast cancer (28), and bladder cancer (31). However, neither stSNP was significantly associated with breast cancer ($P_{\text{heterogeneity}} = 0.17$ and 0.53) in our study.

GPX4. The 14 common SNPs were tagged by four stSNPs with a minimum r_s^2 of 0.81 (10 of 14 SNPs were tagged with $r_p^2 > 0.7$). Two of the three genotyped SNPs, *c-1999g* (rs757229) and *t2572c* (rs713041), located in the 3'-UTR, showed evidence of association in set 1 ($P = 0.06$ and $P = 0.01$, respectively). However, neither of these SNPs was associated with breast cancer in set 2 alone

($P_{\text{heterogeneity}} = 0.43$ and 0.20, respectively) or in the combined data ($P_{\text{heterogeneity}} = 0.22$ and 0.23, respectively; Table 4).

GSR. There were 66 common SNPs spanning two LD blocks. Three tagging SNPs were chosen to tag 43 SNPs in block 1 (minimum $r_s^2 = 0.89$) and four SNPs were chosen to tag 23 SNPs in block 2 (minimum $r_s^2 = 0.81$). Thirty-six of 66 SNPs were tagged with $r_p^2 > 0.7$. No association with a risk of developing of breast cancer was found for any of these SNPs.

TXN. Thirty-seven common SNPs were identified from the EGP data. Inspection of the pairwise LD values indicated two LD blocks. Six SNPs tagged the common SNPs ($n = 14$) for LD block 1 with a minimum r_s^2 of 0.94 and six tagging SNPs were chosen for 20 SNPs in LD block 2 with a minimum r_s^2 of 0.94. Twenty-one of 37 SNPs

Table 4. Results of testing the significant SNPs in set 2 and combined results between set 1 and set 2

Gene and SNPs	MAF	OR [95% CI] (heterozygotes vs common homozygotes)	OR [95% CI] (rare homozygotes vs common homozygotes)	<i>P</i> (2 <i>df</i>)* [†]	<i>P</i> _{trend} (1 <i>df</i>) [†]	
<i>SOD1</i>						
rs4998557, g2809a	Set 2	0.11	1.10 [0.95-1.27]	1.08 [0.66-1.77]	0.43	0.22
	Set 1 + 2	0.11	1.03 [0.93-1.15]	0.81 [0.55-1.20]	0.46	0.95
<i>CAT</i>						
rs511895, g27168a	Set 2	0.41	1.00 [0.87-1.14]	1.11 [0.93-1.32]	0.40	0.31
	Set 1 + 2	0.41	0.94 [0.86-1.03]	1.09 [0.96-1.23]	0.05	0.42
<i>GPX4</i>						
rs757229, c-1999g	Set 2	0.48	0.99 [0.86-1.14]	0.91 [0.77-1.07]	0.43	0.27
	Set 1 + 2	0.47	1.08 [0.98-1.20]	1.02 [0.90-1.15]	0.22	0.67
rs713041, t2572c	Set 2	0.44	1.10 [0.96-1.26]	1.15 [0.97-1.36]	0.20	0.08
	Set 1 + 2	0.45	1.06 [0.96-1.16]	0.97 [0.86-1.09]	0.23	0.76
<i>TXN1</i>						
rs4135179, t2715c	Set 2	0.22	0.92 [0.81-1.05]	1.04 [0.80-1.35]	0.40	0.51
	Set 1 + 2	0.22	0.89 [0.81-0.97]	1.14 [0.95-1.37]	0.007	0.43
<i>TXNRD2</i>						
rs4485648, a9921g	Set 2	0.20	0.99 [0.87-1.12]	1.08 [0.80-1.46]	0.84	0.88
	Set 1 + 2	0.20	0.92 [0.84-1.01]	1.10 [0.90-1.36]	0.10	0.46
rs5748469, A66S	Set 2	0.35	1.00 [0.88-1.13]	1.01 [0.84-1.22]	0.99	0.92
	Set 1 + 2	0.34	1.05 [0.96-1.15]	1.12 [0.98-1.23]	0.23	0.07
rs756661, g23524a	Set 2	0.45	0.98 [0.86-1.12]	0.93 [0.78-1.10]	0.68	0.42
	Set 1 + 2	0.46	0.92 [0.83-1.00]	0.89 [0.79-1.00]	0.11	0.046
rs732262, g44669a	Set 2	0.09	0.94 [0.80-1.10]	1.03 [0.49-2.17]	0.73	0.49
	Set 1 + 2	0.09	1.03 [0.92-1.15]	0.86 [0.54-1.37]	0.73	0.89
rs1548357, t50964c	Set 2	0.30	1.00 [0.88-1.13]	0.92 [0.74-1.14]	0.74	0.57
	Set 1 + 2	0.29	1.05 [0.96-1.15]	1.08 [0.93-1.26]	0.39	0.17
rs20773752, I370T	Set 2	0.27	0.97 [0.85-1.09]	0.98 [0.78-1.22]	0.87	0.65
	Set 1 + 2	0.28	0.93 [0.85-1.01]	1.02 [0.86-1.19]	0.19	0.37

**P* (2 *df*): genotype frequencies in cases and controls were compared using a χ^2 test with 2 *df*.

[†]Significant SNPs at the 0.05 level using the heterogeneity test or the test for trend are in bold.

were tagged with $r_p^2 > 0.7$. Two SNPs between the blocks that were not in strong LD with any other SNPs were also genotyped. The SNP *t2715c* (rs4135179), located in block 1, showed a significant difference in genotype distribution between cases and controls ($P = 0.01$), a difference that remained significant in the combined data sets ($P_{\text{heterogeneity}} = 0.007$; Table 4). As with *CAT g27168a*, this was due to an apparent reduction in risk for *tc* heterozygotes (OR, 0.89; 95% CI, 0.81-0.97) and a small increase in risk for *cc* homozygotes (OR, 1.14; 95% CI, 0.95-1.37).

TXNRD1. There were data on eight SNPs in HapMap covering the gene at a density of 8 kbp per SNP. Eight SNPs defined five common haplotypes which were tagged by four SNPs (minimum $r_h^2 = 0.88$, minimum $r_s^2 = 0.89$). Five of eight SNPs were tagged with $r_p^2 > 0.7$. No differences in genotype frequencies between cases and controls were observed for any of the tag SNPs.

TXN2. There were data on seven SNPs in HapMap covering the gene at a density of 2 kbp per SNP. Three of seven SNPs in HapMap were chosen to tag four common haplotypes (minimum $r_h^2 = 0.80$, minimum $r_s^2 = 0.85$). Six of seven SNPs were tagged with $r_p^2 > 0.7$. No significant association was observed between any of these SNPs and risk of breast cancer.

TXNRD2. Thirty SNPs were available in HapMap covering three LD blocks at a density of 2.1 kbp per SNP. There were three tagging SNPs for six SNPs in block 1 (minimum $r_h^2 = 1.00$, minimum $r_s^2 = 1.00$), four tagging SNPs for 14 SNPs in block 2 (minimum $r_h^2 = 0.92$, minimum $r_s^2 = 0.93$), and three tagging SNPs for five SNPs in block 3 (minimum $r_h^2 = 0.80$, minimum $r_s^2 = 0.81$). Twenty-one of 30 SNPs were tagged with $r_p^2 > 0.7$. Five of the 10 studied SNPs were associated with breast cancer and genotyped in set 2. None of these associations was significant in set 2 alone, but in the combined data, the S allele of A66S was associated with a borderline increased risk of breast cancer in an apparently codominant manner [AS versus AA: OR, 1.05 (95% CI, 0.96-1.15); SS versus AA: OR, 1.12 (95% CI, 0.98-1.29); $P_{\text{trend}} = 0.07$]. Similarly, *g23524a* was associated with a codominant decrease in risk [*ga* versus *gg*: OR, 0.92 (95% CI, 0.83-1.00); *aa* versus *gg*: OR, 0.89 (95% CI, 0.79-1.00); $P_{\text{trend}} = 0.05$]. A marginal association was only observed for TXNRD2 LD block 2 when haplotype frequencies were compared ($P = 0.09$; Table 3). However, the results were not confirmed when we combined set 1 and set 2 data ($P = 0.33$; data not shown).

Discussion

We have assessed the effect of 54 stSNPs in 10 polymorphic genes coding for antioxidant defense enzymes on risk of breast cancer in a large population-based case-control study. Only two of these SNPs (*SOD2 A16V* and *GPX1 P200L*) have been investigated previously in cancer susceptibility and there are no published data on other variants in these genes or variants in other antioxidant defense genes.

In our initial set, 10 SNPs were significant at the 0.10 level, of which five were significant at the 0.05 level. After genotyping in the second set, two SNPs remained significant at the 0.05 level using the heterogeneity test (*CAT g27168a* and *TXN t2715c*). However, in both cases there was a reduced risk for heterozygous individuals and an increased risk for rare homozygous individuals. The test for trend was significant for *TXNRD2 g23524a* and borderline for *TXNRD2 A66S* in the combined data set. There was no evidence for any of the SNPs that the genotype specific risks were different for different disease subgroups—cases were stratified by incident and

prevalent, age at diagnosis (age <50 versus age 50+), histologic subtype (lobular versus ductal versus other), histologic grade, or clinical stage. The observed main effects may be real or may be the result of chance or bias. Chance seems to be the most likely explanation. Despite the large sample size, none of the associations in the combined set 1 and set 2 data were highly significant. Furthermore, for two of the SNPs (*CAT g27168a* and *TXN t2715c*), there was a reduced risk for heterozygotes and an increased risk in rare homozygotes, which seems to be biologically implausible. Both of these SNPs are intronic and neither is strongly correlated with other known SNPs more likely to be functional. On the other hand, *TXNRD2 A66S* results in a nonsynonymous amino acid substitution. The PupaSNP webtool (see Materials and Methods) predicts that A66S affects several protein domains and the base change may also modify splice recognition. TXNRD2 belongs to the family of flavin adenine dinucleotide (FAD)-binding proteins and codon 66 is in the only conserved sequence motif present in all members of the family. This motif is located in the FAD-binding domain (IPR001327) and the NAD-binding site (IPR000205). In the conserved sequence $\text{xhxhGxGxxGxxxhxxh(x)}_g\text{hxhE(D)}$, h represents hydrophobic residues which provide hydrophobic interactions between an α -helix and a β -sheet (35). h indicates the position of alanine in codon 66; the change to serine, a nonhydrophobic residue, could affect the folding of the protein. Functional studies are required to test if this hypothesis is correct and to show how the TXNRD2 protein is affected. The other significant *TXNRD2* variant (*g23524a*) is located in intron 4, 62 bp from a splicing site. Further functional studies will be necessary to test the biological effect of this SNP.

An alternative explanation for a (false) positive association is bias due to hidden population stratification. This occurs when allele frequencies differ between population subgroups and cases and controls are drawn differentially from those subgroups. However, it seems unlikely that population stratification is relevant here because the cases and controls were drawn from the same ethnic groups (>98% white). Furthermore, we have found no evidence for association between 23 unlinked markers (209 tests) in the controls, which suggest that there is unlikely to be significant substructure in our population.⁶ It is also possible that selection bias may result in a false positive as response rates were different in cases and controls (64% versus 45%). Nevertheless, genotype at these loci is unlikely to be related to study participation and selection bias is more a theoretical than real problem.

We have found no evidence that common variations in *SOD1*, *SOD2*, *GPX1*, *GPX4*, *GSR*, *TXN2*, or *TXNRD1* are associated with breast cancer. The SNPs under study were not selected because of their predicted effects on structure and function but because they adequately tagged all known common variants. Nevertheless, it is possible that important, unidentified variants in the genes were not efficiently tagged. For *SOD1*, *SOD2*, *GPX1*, *GPX4*, and *GSR*, resequencing data are likely to have identified most of the common SNPs. Some may have been missed by chance or because of sequencing errors, but given the density of SNPs observed, it is unlikely that any missed SNPs will not have been well tagged. *TXNRD1*, *TXN2*, and *TXNRD2* have not been resequenced in large numbers of individuals and it is therefore not known certain whether or not other common variants exist. However, where a high density of SNPs has been identified, it is anticipated that any

⁶ E.L. Goode and N.J. Wareham, unpublished data.

further SNPs will be well tagged by the known haplotypes. HapMap data available for *TXN2* and *TXNRD2* covered the genes at a density of one SNP every 2 kbp, and for *TXNRD1* HapMap data were at a density of one SNP every 4 kbp. It is also possible that the populations used for selecting tag SNPs—mixed American ethnicities for EGP after exclusion of African Americans and Centre d'Etude du Polymorphisme Humain trios for HapMap—do not adequately represent the population from which our study has been drawn. However, the haplotype frequencies estimated from our data are similar to those estimated using both EGP and HapMap data.

Several previous studies have assessed single variants in *SOD2* and *GPX1* for associations with cancers of various types. There have been two reports of a positive association of the A16V variant *SOD2* gene with cancer (24, 29). Subsequent studies failed to confirm these findings (26, 30) and we too found no evidence for an association of this variant with breast cancer. The previously reported association with *GPX1* L200P (31) has not been replicated by Cox et al. (36) in the prospective Nurses' Health study and we also found no evidence for association with this variant. Our failure to replicate these associations is unlikely to be due to lack of statistical power. For A16V, Ambrosone et al.

(24) reported a dominant protective relative risk of 0.25. Our study had 97% power to detect an allele with this frequency with a type I error rate of 0.0001 even if the true relative risk was 0.75. For *GPX1* L200P, Hu and Diamond (28) reported a relative risk of 1.9 per allele. We had 90% power to detect a codominant allele of this frequency conferring a relative risk of 1.2 with a type I error rate of 0.0001.

In conclusion, we found no evidence for association between common variants in *SOD2*, *SOD1*, *GPX1*, *GPX4*, *GSR*, *TXNRD1*, and *TXN2* and breast cancer risk. There was some evidence for association of variants in *CAT*, *TXN*, and *TXNRD2*. These results would be worth replicating in other studies, particularly as there is some evidence for a functional effect of *TXNRD2* A66S.

Acknowledgments

Received 5/27/2005; revised 9/23/2005; accepted 11/11/2005.

Grant support: Cancer Research UK.

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 Oluseun Ajai and Don Conroy for technical support and the European Prospective Investigation of Cancer management team (K-T. Khaw, S. Oakes, S. Bingham, and J. Russell) for access to control DNA.

References

- McPherson K, Steel CM, Dixon JM. ABC of breast diseases, breast cancer—epidemiology, risk factors, and genetics. *BMJ* 2000;321:624–8.
- Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 2001;358:1389–99.
- Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78–85.
- Easton DF. How many more breast cancer predisposition genes are there. *Breast Cancer Res* 1999;1:14–7.
- Houlston RS, Peto J. The search for low-penetrance cancer susceptibility alleles. *Oncogene* 2004;23:6471–6.
- Mitrunen K, Hirvonen A. Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism. *Mutat Res* 2003;544:9–41.
- Emerit I. Reactive oxygen species, chromosome mutation, and cancer: possible role of clastogenic factors in carcinogenesis. *Free Radic Biol Med* 1994;16:99–109.
- Chiarugi P. Reactive oxygen species as mediators of cell adhesion. *Ital J Biochem* 2003;52:31–5.
- Kang J, Zheng R. Dose-dependent regulation of superoxide anion on the proliferation, differentiation, apoptosis and necrosis of human hepatoma cells: the role of intracellular Ca^{2+} . *Redox Rep* 2004;9:37–48.
- Oberley LW, Buettner GR. Role of superoxide dismutase in cancer: a review. *Cancer Res* 1997;39:1141–9.
- Thannickal VJ, Fanburg BL. Reactive oxygen species in cell signalling. *Am J Physiol Lung Cell Mol Physiol* 2000;279:L1005–28.
- Kayanoki Y, Fujii J, Suzuki K, Kawata S, Matsuzawa Y, Taniguchi N. Suppression of antioxidative enzyme expression by transforming growth factor- β 1 in rat hepatocytes. *J Biol Chem* 1994;269:15488–92.
- Islam KN, Kayanoki Y, Kaneto H, et al. TGF- β 1 triggers oxidative modifications and enhances apoptosis in HIT cells through accumulation of reactive oxygen species by suppression of catalase and glutathione peroxidase. *Free Radic Biol Med* 1997;22:1007–17.
- Kang J, Chen J, Shi Y, Jia J, Zhang Y. Curcumin-induced histone hypoacetylation: the role of reactive oxygen species. *Biochem Pharmacol* 2005;69:1205–13.
- White AC, Das SK, Fanburg BL. Reduction of glutathione is associated with growth restriction and enlargement of bovine pulmonary artery endothelial cells produced by transforming growth factor- β 1. *Am J Respir Cell Mol Biol* 1992;6:364–8.
- Cantin AM, Larivee P, Begin RO. Extracellular glutathione suppresses human lung fibroblast proliferation. *Am J Respir Cell Mol Biol* 1990;3:79–85.
- Powis G, Kirkpatrick DL, Angulo M, Baker A. Thioredoxin redox control of cell growth and death and the effects of inhibition. *Chem Biol Interact* 1998;111–2:23–4.
- Townsend DM, Tew KD, Tapiero H. The importance of glutathione in human disease. *Biomed Pharmacother* 2003;57:145–55.
- Saitoh M, Nishitoh H, Fujii M, et al. Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1. *EMBO J* 1998;17:2596–606.
- Holmgren A. Thioredoxin. *Annu Rev Biochem* 1985;54:237–71.
- Becker K, Gromer S, Schirmer RH, Muller S. Thioredoxin reductase as a pathophysiological factor and drug target. *Eur J Biochem* 2000;267:6118–25.
- Streicher KL, Sylte MJ, Johnson SE, Sordillo LM. Thioredoxin reductase regulates angiogenesis by increasing endothelial cell-derived vascular endothelial growth factor. *Nutr Cancer* 2004;50:221–31.
- Moscov JA, Schmidt L, Ingram DT, Gnarr J, Johnson B, Cowan KH. Loss of heterozygosity of the human cytosolic glutathione peroxidase I gene in lung cancer. *Carcinogenesis* 1994;15:2769–73.
- Ambrosone CB, Freudenheim JL, Thompson PA, et al. Manganese superoxide dismutase (MnSOD) genetic polymorphisms, dietary antioxidants, and risk of breast cancer. *Cancer Res* 1999;59:602–6.
- Ratnasingham D, Tangrea JA, Andersen MR, et al. Glutathione peroxidase codon 198 polymorphism variant increases lung cancer risk. *Cancer Res* 2000;60:6381–3.
- Levine AJ, Elkhoully E, Diep AT, Lee ER, Frankl H, Haile RW. The MnSOD A16V mitochondrial targeting sequence polymorphism is not associated with increased risk of distal colorectal adenomas: data from a sigmoidoscopy-based case control study. *Cancer Epidemiol Biomarkers Prev* 2002;11:1140–1.
- Kote-Jarai Z, Durocher F, Edwards SM, et al. Association between the GCG polymorphism of the selenium dependent GPX1 gene and the risk of young onset of prostate cancer. *Prostate Cancer Prostatic Dis* 2002;5:189–92.
- Hu YJ, Diamond AM. Role of glutathione peroxidase 1 in breast cancer: loss of heterozygosity and allelic differences in the response to selenium. *Cancer Res* 2003;63:3347–51.
- Woodson K, Tangrea JA, Lehman TA, et al. Manganese superoxide dismutase (MnSOD) polymorphism, α -tocopherol supplementation and prostate cancer risk in the α -tocopherol, β -carotene cancer prevention study (Finland). *Cancer Causes Control* 2003;14:513–8.
- Tamimi RM, Hankinson SE, Spiegelman D, Colditz GA, Hunter DJ. Manganese superoxide dismutase polymorphism, plasma antioxidants, cigarette smoking, and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2004;13:989–96.
- Ichimura Y, Habuchi T, Tsuchiya N, et al. Increased risk of bladder cancer associated with a glutathione peroxidase 1 codon 198 variant. *J Urol* 2004;172:728–32.
- Crawford DC, Carlson CS, Rieder MJ, et al. Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. *Am J Hum Genet* 2004;74:610–22.
- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70:425–34.
- Dym O, Eisenberg D. Sequence-structure analysis of FAD-containing proteins. *Protein Sci* 2001;10:1712–28.
- Cox DG, Hankinson SE, Kraft P, Hunter DJ. No association between GPX1 Pro198Leu and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2004;13:1821–2.

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Tagging Single-Nucleotide Polymorphisms in Antioxidant Defense Enzymes and Susceptibility to Breast Cancer

Arancha Cebrian, Paul D. Pharoah, Shahana Ahmed, et al.

Cancer Res 2006;66:1225-1233.

Updated version Access the most recent version of this article at:
<http://cancerres.aacrjournals.org/content/66/2/1225>

Supplementary Material Access the most recent supplemental material at:
<http://cancerres.aacrjournals.org/content/suppl/2006/01/24/66.2.1225.DC1>

Cited articles This article cites 36 articles, 9 of which you can access for free at:
<http://cancerres.aacrjournals.org/content/66/2/1225.full#ref-list-1>

Citing articles This article has been cited by 7 HighWire-hosted articles. Access the articles at:
<http://cancerres.aacrjournals.org/content/66/2/1225.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://cancerres.aacrjournals.org/content/66/2/1225>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.