Functional polymorphisms, altered gene expression and genetic association link NRH:quinone oxidoreductase 2 to breast cancer with wild-type p53

Ke-Da Yu[1,2], Gen-Hong Di[1], Wen-Tao Yuan[3], Lei Fan[1], Jiong Wu[1,2], Zhen Hu[1],
Zhen-Zhou Shen[1], Ying Zheng[4], Wei Huang[3] and Zhi-Ming Shao[1,2,*]

[1]Breast Surgery Department, Breast Cancer Institute, Cancer Hospital, Fudan University, Shanghai 200032, P.R.
China, [2]Department of Oncology, Institutes of Biomedical Science, Shanghai Medical College, Fudan University,
Shanghai 200032, P.R. China, [3]Chinese National Human Genome Center at Shanghai, Shanghai 201203, P.R. China
and [4]Shanghai Municipal Center for Disease Control and Prevention, Shanghai 200336, P.R. China

We hypothesized that NRH:quinone oxidoreductase 2 (NQO2) is a candidate susceptibility gene for breast cancer because of its known enzymatic activity on estrogen-derived quinones and its ability to stabilize p53. We performed case–control studies to investigate the contributions of genetic variants/haplotypes of the NQO2 gene to breast cancer risk. In the first hospital-based study ($n = 1604$), we observed significant associations between the incidence of breast cancer and a 29 bp-insertion/deletion polymorphism (29 bp-I/ D) and the rs2071002 ($+237A>C$) polymorphism, both of which are located within the NQO2 promoter region. Decreased risk was associated with the D-allele of 29 bp-I/D [odds ratio (OR), 0.76; $P = 0.0027$] and the $+237C$-allele of rs2071002 (OR, 0.80; $P = 0.0031$). Specifically, the susceptibility variants within NQO2 were notably associated with breast carcinomas with wild-type p53 (the most significant P-value: 3.3×10^{-6}). The associations were successfully replicated in an independent population set (familial/early-onset breast cancer cases and community-based controls, $n = 1442$). The combined P-values of the two studies ($n = 3046$) are 3.8×10^{-7} for 29 bp-I/D and 2.3×10^{-6} for rs2071002. Furthermore, we revealed potential mechanisms of pathogenesis of the two susceptibility polymorphisms. Previous work has demonstrated that the risk-allele I-29 of 29 bp-I/D introduces transcriptional-repressor Sp3 binding sites. Using promoter reporter-gene assays and electrophoretic-mobility-shift assays, our present work demonstrated that the other risk-allele, $+237A$-allele of rs2071002, abolishes a transcriptional-activator Sp1 binding site. Furthermore, an *ex vivo* study showed that normal breast tissues harboring protective genotypes expressed significantly higher levels of NQO2 mRNA than those in normal breast tissues harboring risk genotypes. Taken together, the data presented here strongly suggest that NQO2 is a susceptibility gene for breast carcinogenesis.

INTRODUCTION

A large body of research in molecular epidemiology supports the hypothesis that estrogens and their metabolites are carcinogens in mammary glands (1–3). For example, estrogen-3,4-quinone can form unstable adducts with DNA that lead to mutations both *in vitro* and *in vivo* (4). In women, chronic exposure to estrogen and estrogen-derived quinone or semi-quinone metabolites is achieved through physiological metabolism and relevant environmental factors. In breast tissues, phase I and II detoxification pathways protect against damage caused by reactive metabolites (1). Phase II enzymes are critical, as the rate at which phase I produces intermediates must be balanced by the rate at which phase II

*To whom correspondence should be addressed at: Department of Breast Surgery, Cancer Hospital/Cancer Institute, Breast Cancer Institute, Fudan University, 399 Ling-Ling Road, Shanghai, 200032 China. Tel: +86 2164175590; Fax: +86 2164434556; Email: zhimingshao@yahoo.com

finishes their processing. Therefore, slow or inactive phase II enzymes may be factors in carcinogenesis. Quinone oxidoreductases, phase II detoxification enzymes that neutralize carcinogens, are cytosolic enzymes that catalyze the metabolic reduction of quinones and their derivatives (5). Two isoforms of human quinone oxidoreductases, NAD(P)H:quinone oxidoreductase 1 (NQO1, MIM:125860) and NRH:quinone oxidoreductase 2 (NQO2, MIM:160998), have been identified, and their genes have been cloned (6). Although various studies have established that NQO1 protects against redox-cycling compounds, oxidative stress and breast carcinogenesis (7,8), the function of NQO2, a homologue of NQO1, is still unknown (9,10).

The association between breast cancer and NQO2 remains unclear. Some studies have suggested that greater NQO2 expression may activate certain kinds of chemicals in the brain, leading to oxidative stress and resulting in neuronal damage (11). Other studies have implied that NQO2 can protect against quinone-induced skin carcinogenesis (12). Recently, new evidence (13) has shown for the first time that NQO2 catalyzes the reduction of electrophilic estrogen quinones and thereby acts as a detoxification enzyme. Gaikwad *et al.* successfully demonstrated that estrogen-3,4-quinone binds to NQO2 and established that estrogen quinones are endogenous biological substrates of NQO2. Moreover, they demonstrated that NQO2 is faster at reducing estrogen quinones than its homologue NQO1. Such encouraging findings reveal a possible relationship between breast cancer and NQO2, although no studies to date have addressed this issue. Besides, NQO2 can stabilize p53 protein (14), a known breast tumor suppressor gene product. p53 is recognized as a highly penetrant breast cancer susceptibility gene, and loss of both p53 and BRCA1 results in the rapid and efficient formation of mammary carcinomas (15). The cross-talk between NQO2 and p53 implies a potential modulating effect of NQO2 on breast carcinogenesis. Interestingly, the human NQO2 gene is located on chromosome 6 in the 6pter-q12 segment, within a chromosomal region frequently lost in many breast carcinoma tissues and cells (16–18).

It is logical to assume that NQO2 is a candidate gene for breast cancer susceptibility. Thus far, the genetic contribution of NQO2 to breast cancer susceptibility has not yet to be investigated. Here, we conducted case–control studies by analyzing genomic DNA from two sets of Chinese populations (sporadic and familial/early-onset breast cancer patients, respectively). The results presented here revealed that the functional genetic variants in NQO2 that elevate its promoter activity and increase its expression level attenuate susceptibility to breast cancer. In other words, higher expression of NQO2 provides protection against breast carcinogenesis. Additional data suggest that NQO2 confers specific protection against breast cancer with wild-type p53.
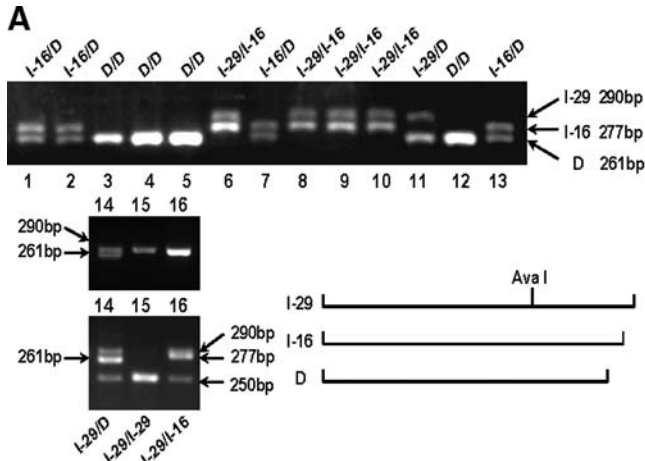
RESULTS

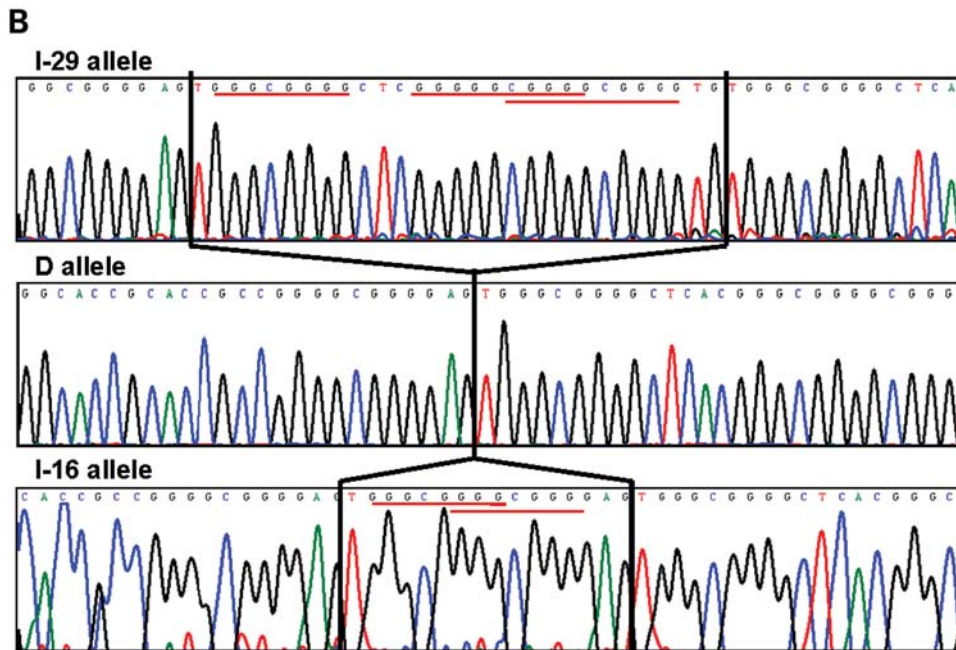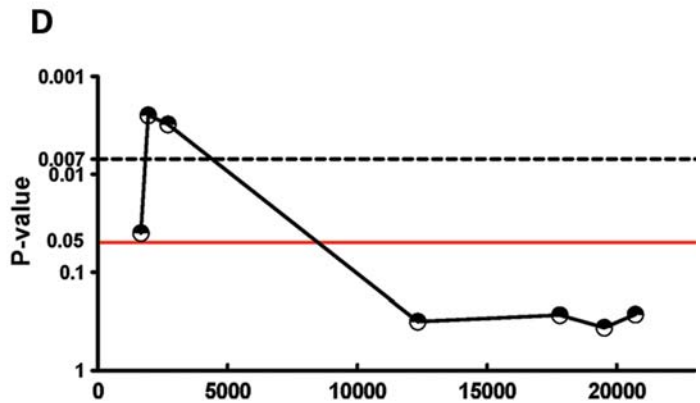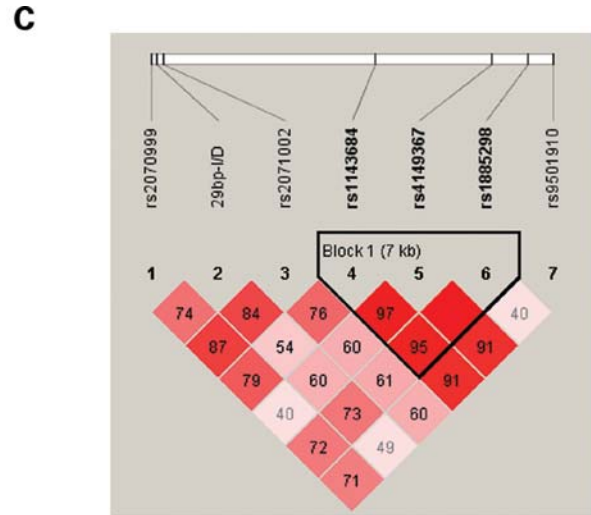### Polymorphisms and haplotypes in NQO2 in a Chinese population

In the present study, we first set out to systematically analyze the genetic variants in NQO2 in a hospital-based population comprising 1604 unrelated Shanghai women (893 sporadic breast cancer cases and 711 cancer-free controls). Ten single-nucleotide polymorphisms (SNPs) and one insertion/deletion polymorphism of 29 base pairs (29 bp I/D) were selected for genotyping. Of the 11 tested polymorphisms, five polymorphisms are tagging SNPs (tSNPs) capturing 84% of the genetic variants within the NQO2 gene locus, and six polymorphisms are potentially functional variants (see detailed information in Materials and Methods). After genotyping, three non-synonymous SNPs (rs17136117 (Gly29Glu), rs28383623 (Arg16Lys) and rs17300141 (Asp58Gly)) were excluded due to non-polymorphism in the Shanghai population, and one (rs3823096, which is in an intron and tags itself and one other SNP in an intron) failed technically. The remaining seven polymorphisms, which were genotyped successfully with genotyping call rates ranging from 90 to 100%, could capture the un-genotyped SNPs very well. The polymorphisms used were rs2070999 ($-338G > A$, the transcriptional start site is designated as $+1$), 29 bp-I/D (-60 to -32), rs2071002 ($+237A > C$), rs1143684 ($+10324C > T$, Phe47Leu), rs4149367 ($+15799C > T$), rs1885298 ($+17514G > T$) and rs9501910 ($+18708G > C$).

The 29 bp-I/D polymorphism, which is the most studied polymorphism located within the NQO2 promoter region, plays an important role in activating gene expression and is associated with Parkinson's disease (11,19). One recent study (19) demonstrated that the previously reported 29 bp-I/D is indeed a tri-allelic polymorphism consisting of a 29 bp-insertion (I-29), a 29 bp-deletion (D) and a 16 bp-insertion (I-16). However, the existence of the tri-allelic polymorphism in the Chinese population had not yet been confirmed. We first randomly selected 40 cases and 40 controls to directly sequence the genetic fragment containing the 29 bp-I/D site. Only one individual of I-16/I-29 genotype was identified. We concluded from this analysis that although there are three alleles in the so-called 29 bp-I/D polymorphism, the occurrence of I-16 is relatively scarce in our population. We next used a sensitive genotyping strategy that permits identification of I-29, I-16 and D alleles in all of the samples. In brief, we used a polymerase chain reaction (PCR) and lengthy electrophoresis [rather than the short-time electrophoresis used previously (11)] method to discriminate the I-29, I-16 and D alleles (19). Because the I-29 and I-16 alleles are difficult to separate through electrophoresis (there are only 13 bp different between them), we employed a PCR-restriction fragment length polymorphism (RFLP) approach to discriminate I-16 from I-29 alleles (typical gel electrophoresis plots and delineation are shown in Fig. 1A; sequence plots of all three alleles are shown in Fig. 1B). As shown in the embedded table in Figure 1A, only three I-16 homozygous individuals and 24 I-16 heterozygous individuals were found, resulting in an I-16 allele frequency of 0.94%. The low occurrence rate of I-16 in our moderately sized study population made it impossible to sufficiently reach statistical power to detect an association between the I-16 allele and breast cancer risk. Additionally, preliminary data have suggested that the biological role of the I-16 allele is somewhat different from that of I-29 (19). We therefore excluded individuals harboring I-16 alleles from further association analysis.

Table 1. Frequency of NQO2 alleles and haplotypes in 893 sporadic breast cancer patients and 711 controls

Name	Risk allele	Location	Allele (*n*, %) All cases	Wt-p53 cases[a]	Controls	All *P*-value[b]	All *P*-value[c]	Wt-p53 *P*-value[b]	Wt-p53 *P*-value[c]
Polymorphism			*n* = 1786	*n* = 488	*n* = 1422				
rs2070999	A	Promoter	492 (27.6)	127 (26.0)	346 (24.4)	**0.040**	N.S.	N.S.	N.S.
29 bp-I/D[d]	I-29	Promoter	1429 (82.3)	423 (87.8)	1083 (78.0)	$\mathbf{2.7 \times 10^{-3}}$	**0.011**	$\mathbf{3.3 \times 10^{-6}}$	$\mathbf{1.0 \times 10^{-5}}$
rs2071002	A	5′-UTR	1191 (67.7)	335 (70.1)	873 (62.6)	$\mathbf{3.1 \times 10^{-3}}$	**0.013**	$\mathbf{3.3 \times 10^{-3}}$	**0.013**
rs1143684	T	Exon, nonsyn.	1175 (66.5)	309 (63.6)	909 (64.8)	N.S.	N.S.	N.S.	N.S.
rs4149367	C	Exon, syn.	1476 (83.6)	410 (84.7)	1156 (82.1)	N.S.	N.S.	N.S.	N.S.
rs1885298	G	Intron	1534 (87.0)	410 (85.4)	1214 (85.9)	N.S.	N.S.	N.S.	N.S.
rs9501910	G	Intron	1060 (64.9)	269 (60.6)	791 (62.9)	N.S.	N.S.	N.S.	N.S.
Haplotype[e]			%						
1	TCG		37.4	34.3	33.0	**0.010**	**0.022**	N.S.	N.S.
2	CCG		33.1	35.6	34.9	N.S.	N.S.	N.S.	N.S.
3	TTG		16.3	15.1	17.8	N.S.	N.S.	N.S.	N.S.
4	TCT		12.9	14.1	14.0	N.S.	N.S.	N.S.	N.S.

Wt, wild-type; UTR, untranslated region; nonsyn., non-synonymous; syn., synonymous; N.S., no significance. The significance of boldface is a *P*-value of <0.05.
[a]Among 497 sporadic patients of invasive breast cancer with available p53 status, 244 were identified as breast cancer with wild-type p53.
[b]Unadjusted *P*-value of two-sided χ^2 test.
[c]After 1000-permutation tests.
[d]The 27 individuals with the I-16 allele (10 in cases and 17 in controls) were excluded from the analysis.
[e]Haplotypes were derived from the structure of the linkage disequilibrium (LD) block consisting of rs1143684, rs4149367 and rs1885298.

Of the other six successfully genotyped dimorphic SNPs which have minor allele frequencies (MAFs) of at least 10%, no significant deviation from Hardy–Weinberg equilibrium (HWE) were found in either cases or controls, using a cutoff value of 0.05, except for a deviation of rs1885298 in the case group. In view of the accuracy of the SNPstream genotyping platform, the HWE departure of rs1885298 in cases, but not in controls, suggests the existence of an association between disease and this polymorphism (20–22). Therefore, we did not remove rs1885298 from single-locus association analysis. Although rs1885298 deviated from HWE in cases (*P* = 0.04), it fulfilled HWE in all samples (*P* > 0.05) that were used for haplotype construction. We have also established haplotypes using only the controls in which the rs1885298 also fulfilled HWE. The results demonstrated that the haplotype block patterns in control subjects are similar to those patterns that are generated using all samples (data not shown). As shown in Figure 1C, all seven polymorphisms were used to characterize the linkage disequi-

librium (LD) pattern and haplotype structure (27 samples harboring the rare I-16 allele of 29 bp-I/D were excluded). Using a method previously described in Gabriel *et al.* (23), rs1143684, rs4149367 and rs1885298 could be parsed into one block. Four common haplotypes (each with a frequency of at least 0.05) were identified in this block (Table 1).

Genetic variants in NQO2 are associated with breast cancer risk in a hospital-based population

Based on the polymorphisms that we identified, we then evaluated the frequency distributions of alleles and haplotypes in 893 cases and 711 controls. As shown in Table 1, three allele genotypes and one haplotype are significantly associated with breast cancer. After Bonferroni correction and permutation tests, the *P*-values of two polymorphisms (29 bp-I/D and +237A>C (rs2071002)) and one haplotype (TCG) continued to be significant using a cutoff level of 0.05 (Table 1 and Fig. 1D). Decreased breast cancer risk was associated

Figure 1. Polymorphisms and haplotypes in NQO2. (**A**) Genotyping of 29 bp-I/D in NQO2 and corresponding frequencies. Typical gel electrophoresis plots of the 29 bp-I/D polymorphism are shown in upper plots. Samples 1–13 were detected by PCR and lengthy electrophoresis run (100 V, 2.5% agarose gel, 120 min). Samples 14–16 were detected by PCR and short-time electrophoresis (100 V, 2.0% agarose gel, 30 min). Under these conditions, the I-29 and I-16 alleles are difficult to separate due to the limited resolution of electrophoresis. For instance, sample 16 is identified as I-29/I-29 by short-time electrophoresis (upper picture in Sample 14–16 plot) but was ultimately shown to be I-29/I-16 by an RFLP assay (lower picture in Sample 14–16 plot). I-29 could be digested by Ava I into 250 and 40 bp fragments, but the I-16 and D alleles could not. Correspondingly, I-29/I-29 could be completely digested, and no band was observed at 290 bp. In contrast, I-29/I-16 had two bands, one 250 bp and the other 277 bp. For I-29 heterozygotes, the I-29 allele could not be completely digested, even with sufficient restriction enzyme, leaving a shadow band at 290 bp. This is due to the inability of Ava I to effectively cut the dimeric I-29/I-16 complex or the dimeric I-29/D complex. The lower embedded table shows the frequencies of diverse 29 bp-I/D alleles and genotypes. Fifteen breast cancer cases were excluded because of poor DNA quality or insufficient quantity for genotyping the 29 bp-I/D alleles. This analysis includes 878 cases and 711 controls. I-29/I-29: homozygous I-29 allele genotype. I-29/I-16: heterozygous I-29 and I-16 genotype. D/D: homozygous genotype for deletion of 29 bp sequence. I-29/D: heterozygous genotype of I-29 and D alleles. (**B**) Sequencing plot of the 29 bp-I/D trimorphism. The 29 bp insertion sequence introduces Sp1/Sp3 binding sites (underlined in red) and thereby enables recruitment of Sp3, which results in decreased promoter activity (Wang *et al.*). The I-16 allele deletes 13 bp from the I-29 allele. (**C**) Pairwise linkage disequilibrium (LD) among selected variants in the NQO2 gene. The value within each diamond represents the pairwise correlation between polymorphisms (measured as D′ (×100)) defined by the upper left and the upper right sides of the diamond. The red-to-white gradient reflects higher to lower LD values, and the diamond without a number corresponds to D′ = 1. (**D**) Significance of each polymorphism tested. The X-axis shows genomic relative position, and the Y-axis shows the *P*-values for the allele. The red line indicates the significance threshold of 0.05, and the black dashed line indicates the significance threshold of the Bonferroni correction for single-locus association (0.05/7 = 0.007). Lines are drawn between points only for the sake of clarity, as the disease association between markers may be not linear.

Downloaded from https://academic.oup.com/hmg/article/18/13/2502/571820 by guest on 20 August 2022

Table 2. Risk associated with NQO2 polymorphisms in 893 sporadic breast cancer cases and 711 controls

Polymorphism	Genotype	Cases (n, %)	Controls (n, %)	Model for individual polymorphism OR (95% CI)[a]	Dominant model OR (95% CI)[a,b]	Recessive model OR (95% CI)[a,b]
rs2070999	GG	472 (52.9)	404 (56.9)	Reference	1.17 (0.96–1.43)	1.46 (0.98–2.18)
	AG	348 (39.0)	266 (37.5)	1.11 (0.91–1.38)		
	AA	72 (8.1)	40 (5.6)	**1.53 (1.02–2.31)**		
29 bp-I/D[c]	I-29/I-29	594 (68.4)	420 (60.5)	Reference	**0.71 (0.57–0.87)**	0.84 (0.51–1.39)
	I-29/D	241 (27.8)	243 (35.0)	**0.70 (0.57–0.87)**		
	DD	33 (3.8)	31 (4.5)	0.75 (0.45–1.25)		
rs2071002	AA	403 (45.8)	264 (37.9)	Reference	**0.72 (0.59–0.89)**	0.82 (0.60–1.11)
	AC	385 (43.8)	345 (49.5)	**0.73 (0.59–0.91)**		
	CC	92 (10.5)	88 (12.6)	**0.69 (0.50–0.96)**		
rs1143684	TT	396 (44.8)	295 (42.1)	Reference	0.89 (0.73–1.09)	0.95 (0.70–1.29)
	CT	383 (43.4)	319 (45.5)	0.89 (0.72–1.10)		
	CC	104 (11.8)	87 (12.4)	0.90 (0.65–1.24)		
rs4149367	CC	613 (69.4)	475 (67.5)	Reference	0.91 (0.74–1.13)	0.68 (0.37–1.25)
	CT	250 (28.3)	206 (29.3)	0.94 (0.76–1.17)		
	TT	20 (2.3)	23 (3.3)	0.67 (0.36–1.23)		
rs1885298	GG	674 (76.4)	523 (74.0)	Reference	0.88 (0.70–1.11)	1.10 (0.57–2.11)
	GT	186 (21.1)	168 (23.8)	0.87 (0.68–1.10)		
	TT	22 (2.5)	16 (2.3)	1.06 (0.55–2.05)		
rs9501910	GG	342 (41.9)	244 (38.8)	Reference	0.88 (0.71–1.09)	0.93 (0.68–1.27)
	CG	376 (46.0)	303 (48.2)	0.88 (0.71–1.11)		
	CC	99 (12.1)	82 (13.0)	0.87 (0.62–1.21)		

The significance of boldface is a *P*-value of <0.05.
[a]OR and 95%CI calculated by logistic regression, adjusted for age, age at menarche, menopause status, body mass index (BMI) and history of oral contraceptive drug use. Major genotype is indicated as Reference.
[b]For the model, homozygotes for the major allele (1/1), heterozygotes (1/2) and homozygotes for the rare allele (2/2) are coded as a continuous numeric variable for genotype (i.e. 0, 1, 2). The dominant model is defined as contrasting genotypic groups 1/1 versus (1/2+2/2), and the recessive model is defined as contrasting genotypic groups 2/2 versus (1/1+1/2).
[c]The 27 individuals with the I-16 allele (10 in cases and 17 in controls) were excluded from this analysis.

with the NQO2 D allele [odds ratio (OR), 0.76; 95% confidence interval (CI), 0.64–0.91, compared with I-29 allele, $P = 0.0027$] and the +237C allele (OR, 0.80; 95% CI, 0.69–0.93, compared with +237A allele, $P = 0.0031$). The 29 bp-I/D and +237A>C genotypes were significantly correlated with breast cancer under a dominant model rather than a recessive model (Table 2). For these two polymorphisms, women homozygous for the more frequent allele were more prevalent in the case group, and variant-type (both homozygotes and heterozygotes) women tended to be in the control group. To confirm the associations of genotype with breast cancer, logistic regression analysis was applied with adjustments for age, age at menarche, menopause status, body mass index (BMI) and history of oral contraceptive drug use. The adjusted OR is 0.71 for the I-29/D and DD genotypes of 29 bp-I/D ($P = 0.001$) and 0.72 for the AC and CC genotypes of +237A>C ($P = 0.002$).

Genetic association links NQO2 specifically to breast cancer with wild-type p53

Preliminary genetic association analysis linked the NQO2 gene to breast cancer susceptibility. Because NQO2 stabilizes p53 by preventing 20S proteasomal degradation (14), we hypothesized that NQO2 may have a greater protective effect when coupled with wild-type p53. Therefore, we further examined the relationship between NQO2 and breast cancer with wild-type p53. Among the 780 patients with sporadic, invasive breast cancer, 497 cases had previously been examined for their p53 status using an immunohistochemistry

(IHC) assay (24). With further analysis, we observed a particularly striking association between susceptibility polymorphisms in NQO2 and breast cancer with wild-type p53 tumors (Table 1). The robust ORs of the D allele (29 bp-I/D) and the C allele (+237A>C) were 0.50 ($P = 3.3 \times 10^{-6}$) and 0.71 ($P = 0.003$), respectively.

An association between NQO2 and breast cancer risk in a second case–control study

To further validate the genetic associations observed in our initial studies, we analyzed the D allele (29 bp-I/D) and the C allele (+237A>C) polymorphisms in another, independent population. Four hundred and three unrelated patients with familial/early-onset breast cancer and 1039 unrelated healthy controls from a community-based population were chosen for a second experiment. In this population set, we confirmed the highly significant associations of breast cancer with both the 29 bp-I/D (OR (95% CI) for alleles: 0.70 (0.56–0.87), $P = 0.0011$; for genotypes: 0.63 (0.49–0.81), $P = 2.4 \times 10^{-4}$ and the +237A>C polymorphisms (alleles, $P = 0.0098$; genotypes, $P = 0.0025$, Table 3). Because the community screening program is ongoing, we have only received epidemiological information for age. The age-adjusted ORs of the DD and +237CC genotypes were 0.52 ($P = 1.5 \times 10^{-4}$) and 0.68 ($P = 0.019$), respectively. Finally, by combining the two studies, both of the polymorphisms obtained highly significant combined *P*-values: 2.7×10^{-6} for alleles and 3.8×10^{-7} for genotypes of 29 bp-I/D and 6.6×10^{-6} for alleles and 2.3×10^{-6} for genotypes of +237A>C (Table 3).

Table 3. Validation of the two susceptibility polymorphisms in NQO2 in 403 familial/early-onset breast cancer cases and 1039 controls

Polymorphism	Allele or genotype	Familial/early-onset breast cancer cases[a] (allele or genotype)				Combined (n = 3046) (allele or genotype)			
		Cases (n, %)	Controls (n, %)	OR (95% CI)	P-value	Cases (n, %)	Controls (n, %)	OR (95% CI)	P-value
29 bp-I/D[b]	I-29	647 (82.7)	1570 (77.1)	Reference		2076 (82.4)	2653 (77.5)	Reference	
	D	135 (17.3)	466 (22.9)	0.70 (0.56–0.87)	1.1×10^{-3c}	442 (17.6)	771 (22.5)	0.73 (0.64–0.84)	2.7×10^{-6c}
	I-29/I-29	273 (69.8)	603 (59.2)	Reference		867 (68.9)	1023 (59.8)	Reference	
	I-29/D+D/D	118 (30.1)	415 (40.8)	0.63 (0.49–0.81)	2.4×10^{-4c}	392 (31.1)	689 (40.2)	0.66 (0.54–0.79)	3.4×10^{-7c}
				0.52 (0.37–0.73)	1.5×10^{-4d}			0.65 (0.56–0.77)	3.8×10^{-7d}
rs2071002	A	530 (65.9)	1258 (60.7)	Reference		1721 (67.1)	2131 (61.5)	Reference	
	C	274 (34.1)	814 (39.3)	0.80 (0.67–0.95)	9.8×10^{-3c}	843 (32.9)	1335 (38.5)	0.78 (0.70–0.87)	6.6×10^{-6c}
	AA	179 (44.5)	372 (35.9)	Reference		582 (45.4)	636 (36.7)	Reference	
	AC+CC	223 (55.5)	664 (64.1)	0.70 (0.55–0.88)	2.5×10^{-3c}	700 (54.6)	1097 (63.3)	0.70 (0.60–0.81)	1.5×10^{-6c}
				0.68 (0.49–0.94)	0.019^{d}			0.68 (0.58–0.80)	2.3×10^{-6d}

The significance of boldface is a P-value of <0.05.
[a]Eleven samples failed genotyping for 29 bp-I/D, and four samples failed genotyping for rs2071002 due to poor DNA quality or insufficient quantity.
[b]Individuals with I-16 allele were excluded from corresponding analysis of 29 bp-I/D. Of the ascertainment population, 27 women (10 in cases and 17 in controls) had I-16; within the replication population, 22 women (4 in cases and 18 in controls) had I-16.
[c]Unadjusted P-value of two-sided χ^2 test.
[d]Adjusted for age.

To further investigate the relationship between the two notable variants, we performed stratified allele study and joint analysis (Table 4). Missing allelic genotypes of 29 bp-I/D and +237A>C, as well as haplotypes derived from these two variants, were established from observed genotypes using a PHASE procedure (25). In the stratified analysis, we revealed that 29 bp-I/D was related to breast cancer in both +237A allele-containing subgroup ($P = 0.007$) and +237C allele-containing subgroup ($P = 0.077$), suggesting an independent role of 29 bp-I/D in breast carcinogenesis. With regard to +237A>C, we observed an association of this SNP with breast cancer in the 29 bp-I allele-containing subgroup. Although we failed to observe an association between this SNP with breast cancer in the D allele-containing subgroup, it might be due to relatively smaller sample-size in D allele subgroup. In the joint analysis, four haplotypes were established from these two genetic variants. We observed associations of the D-C haplotype and D-A haplotype with a pronounced decrease in breast cancer risk (Bonferroni corrected *P*-values of 3.2×10^{-5} and 0.042, respectively). Although the I-29-C haplotype was associated with a relatively decreased risk compared with the I-29-A haplotype ($P = 0.018$), the difference was not statistically significant after Bonferroni correction. We also found that the I-29-A and D-C haplotypes were more frequent than the I-29-C and D-A haplotypes. This observation suggests that there is LD between the I-29 and +237A alleles and between the D and +237C alleles. Our results are in line with the observation that these two variants have a high D′ (0.84), as indicated in Figure 1C. Taken together, the informative stratified and joint analysis imply that, although genetic variants 29 bp-I/D and +237A>C are in strong LD, they are individually associated with breast cancer in an independent way.

Susceptibility polymorphisms in the promoter region of NQO2 affect transcriptional activity

Based on observations made in the epidemiological study, we set out to identify the role of the two susceptibility polymorphisms in pathogenesis. The polymorphisms are adjacently located within the promoter region of NQO2, and recent findings have suggested that the 29 bp-I/D polymorphism regulates NQO2 promoter activity. The regulative effect of the 29 bp-I/D is strongly supported by the demonstration that the GC-rich 29 bp insertion sequence introduces distinct binding sites of transcription factor Sp3, Sp3 in this case acting as a repressor (26). In contrast, no functional research has been conducted on the +237A>C polymorphism (Fig. 2A). We analyzed its flanking sequence using the TESS program to screen for transcription factor binding sites. Results showed that the C allele of +237A>C polymorphism is located in the 5′ untranslated region (UTR) and likely creates a new Sp1 binding site (Fig. 2B). To assess the individual and cooperative effects of the two polymorphisms, we generated four luciferase reporter gene constructs that share identical backbone sequence except for the polymorphisms (Fig. 2C). As shown in Figure 2D, reporter gene expression driven by the +237C-containing NQO2 promoter (I-29-C) was 1.5-fold greater than that driven by the +237A-containing counterpart (I-29-A) ($P < 0.05$). Similarly,

Table 4. Stratified allelic frequencies and joint haplotype frequencies of 29 bp-I/D and rs2071002 (+237A>C) in all of the breast cancer cases and controls

Two-locus analysis[a]		Allele[b]		OR (95% CI)	*P*-value	*P*-value[c]
		Cases (*n*, %)	Controls (*n*, %)			
Stratified analysis						
I-29	+237A	1671 (79.2)	2027 (76.7)	Reference		
	+237C	439 (20.8)	629 (23.3)	**0.85 (0.74–0.97)**	**0.018**	N.A.
D	+237A	48 (10.6)	94 (12.1)	Reference		
	+237C	406 (89.4)	680 (87.9)	1.17 (0.80–1.73)	0.406	N.A.
+237A	I-29	1671 (97.2)	2027 (95.6)	Reference		
	D	48 (2.8)	94 (4.4)	**0.62 (0.43–0.89)**	**0.007**	N.A.
+237C	I-29	439 (52.0)	629 (48.1)	Reference		
	D	406 (48.0)	680 (51.9)	0.86 (0.72–1.02)	0.077	N.A.
Joint haplotype						
I-29	+237A	1671 (65.2)	2027 (59.1)	Reference		
I-29	+237C	439 (17.1)	629 (18.3)	**0.85 (0.74–0.97)**	**0.018**	N.S.
D	+237A	48 (1.9)	94 (2.8)	**0.62 (0.43–0.89)**	**0.007**	**0.042**
D	+237C	406 (15.8)	680 (19.8)	**0.72 (0.63–0.83)**	$\mathbf{5.3 \times 10^{-6}}$	$\mathbf{3.2 \times 10^{-5}}$

N.S., no significance; N.A., not applicable. The significance of boldface is a *P*-value of <0.05.
[a]fastPHASE 1.1 program was used to infer individual's haplotype and missing allelic genotypes based on the observed NQO2 genotypes.
[b]There are a total of 2997 women in this analysis (1282 case and 1715 controls); 49 individuals with the I-16 allele were excluded.
[c]Bonferroni corrected *P*-value.

reporter gene expression driven by the D allele-containing NQO2 promoter (D-A) was significantly greater than that driven by the I-29-containing counterpart (I-29-A). A similar observation regarding 29 bp-I/D has been reported previously (26). Our results verified the inhibitory effect of the 29 bp-insertion sequence on NQO2 gene expression. Promoter activity of the sequences that correspond to the four haplotypes showed similar patterns using either human embryonic epithelial cells (293T) or breast cancer cells (MDA-MB-231 and MCF-7).

Considering the effect that each polymorphism had alone, we next investigated their joint effect. As expected, significantly lower luciferase activity of the I-29-A haplotype was observed when compared with the other three haplotype vectors. However, the D-C haplotype failed to show the maximum promoter activity, as expected (Fig. 2D).

Polymorphism rs2071002 (+237A>C) affected promoter activity by allele-specific binding of the transcriptional factor Sp1

Our results above showed that the +237C allele is associated with significantly increased promoter activity compared with the +237A allele, and this effect may be mediated by the introduction of a transcription factor Sp1 binding site. We used an electrophoretic mobility shift assay (EMSA) to investigate whether the difference in activity between sequences containing +237C or +237A could be attributed to a different affinity of these two alleles in binding to Sp1. As shown in Figure 2E, a much clearer DNA-protein complex was detected with the +237C probe than with the +237A probe. Competition experiments revealed that the +237C-band could be competed away with a 100-fold molar excess of the same, unlabeled probe but could not by the same concentration of unlabeled +237A type probe. We also used a specific Sp1 consensus binding sequence in parallel with the two test sequences to ascertain if the oligonucleotides-complex band using the Sp1 consensus probe was at the same location as when using the +237A/C probes. The corresponding results clearly showed a vastly greater affinity of +237C rather than +237A to Sp1 protein.

Although the results from the *in vitro* EMSA assay supported the presumption that +237C>A affects promoter activity by allele-specific binding of Sp1, a chromatin immunoprecipitation (ChIP) assay could also be used to verify that the nuclear protein

Figure 2. Genetic variants within the NQO2 promoter region alter promoter activity by introducing an Sp1 binding site. (**A**) Sequencing plot of the +237A>C SNP. Bioinformatic analysis predicts that the C allele creates a new Sp1 binding site. (**B**) Schematic representation of the two putative consensus sequences of the Sp1 biding site (plots available at: http://www.cbil.upenn.edu/cgi-bin/tess/tess?request=FCT-DBRTRV-Accno&key=T00759). (**C**) Schematic graphs of luciferase reporter plasmids. To construct vectors of four haplotypes, NQO2 promoter regions (−537 to +529 bp) containing both 29 bp-I/D and +237A>C polymorphisms were cloned into a pGL3-Basic reporter gene. (**D**) Promoter construct activity *in vitro*. Arbitrary units denote fold increase, measured by defining the activity of the empty pGL3-Basic vector as one. Data shown represent mean values, with error bars showing standard error (SE). Significantly low promoter activity is observed in the I-29-A haplotype compared with the other three haplotype constructs ($P < 0.05$ of MDA-MB-231; $P < 0.01$ of MCF-7; $P < 0.001$ of 293T). Comparisons of relative luciferase activity between groups are performed by ANOVA analysis and adjusted by the Games–Howell method if equal variances are not assumed. (**E**) Allele-specific binding of the transcription factor Sp1 in the promoter region that contains the +237A>C allele. Biotin-labeled double-stranded probes (100 fmol) of the +237A allele, the +237C allele or the Sp1 consensus binding sequence were incubated with 300 ng of pure human Sp1 protein. In competition experiments, 100-fold excess of unlabeled probe was added before the addition of the labeled probes. The site of an oligo-nucleotides/Sp1 complex is indicated by an asterisk. (**F**) Relative expression of Sp1, Sp3 and the Sp1/Sp3 ratio in normal breast tissues (N.B.T), HBL-100, MCF-7, MDA-MB-231 and 293T cells. *GAPDH* expression is used for normalization. (**G**) Dose response of the four haplotypes of the NQO2 promoter on Sp1 treatment. 293T cells were transfected with increasing concentrations of Sp1 expression vector. Plasmids were transfected as described in Materials and Methods. *Statistical significance as determined by ANOVA, compared with I-29-A haplotype ($P < 0.001$). $Significant increase (ANOVA), compared with D-A and I-29-C haplotypes ($P < 0.01$). #Significant increase (ANOVA), compared with D-A and I-29-C haplotypes ($P < 0.001$). Arbitrary units denote fold increase, measured by defining the activity of the empty pGL3-Basic vector as one. Data are means ± SE from three independent experiments.
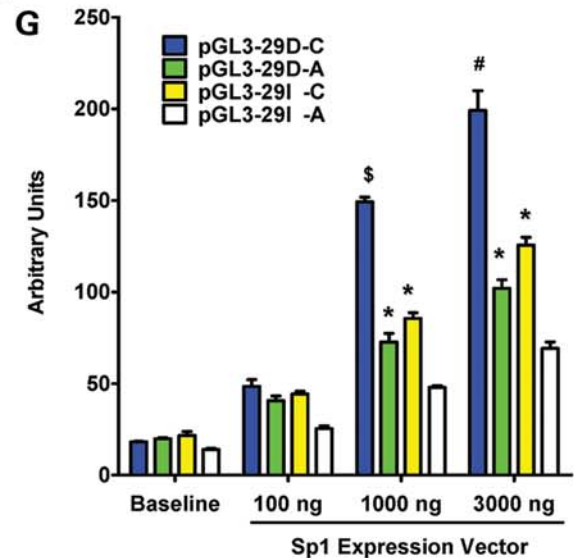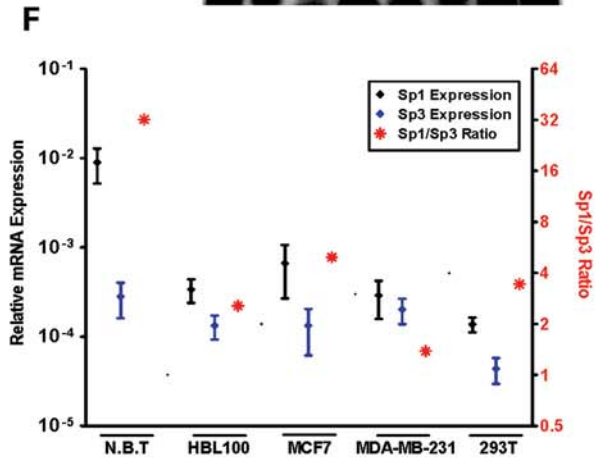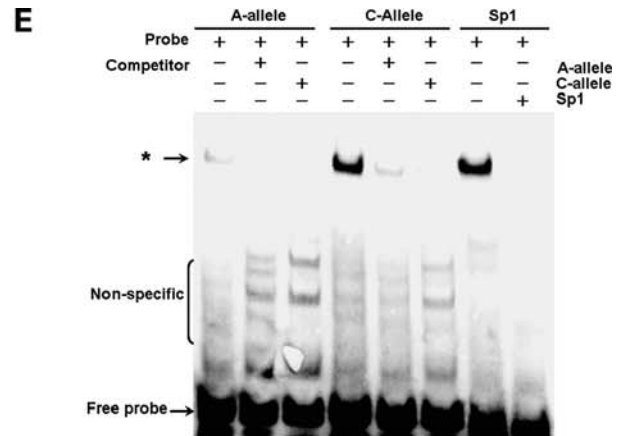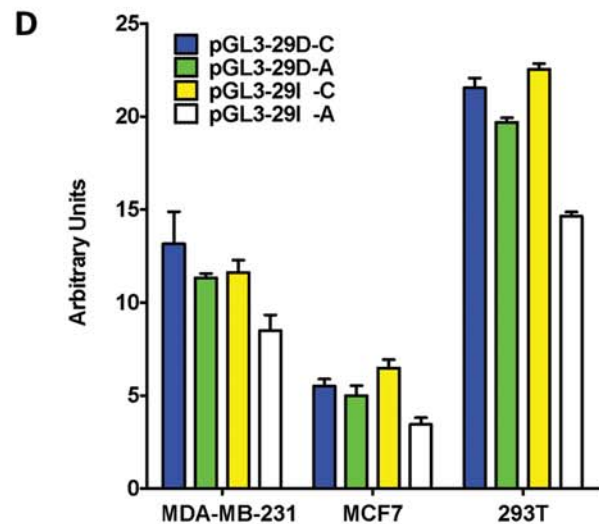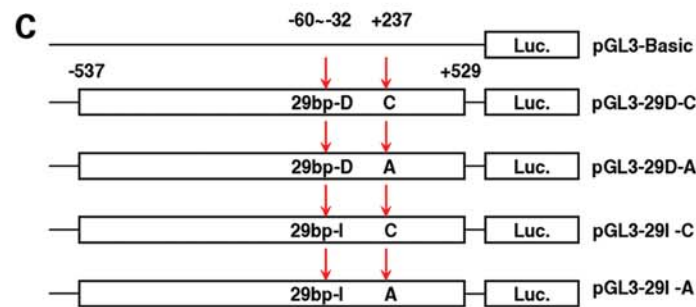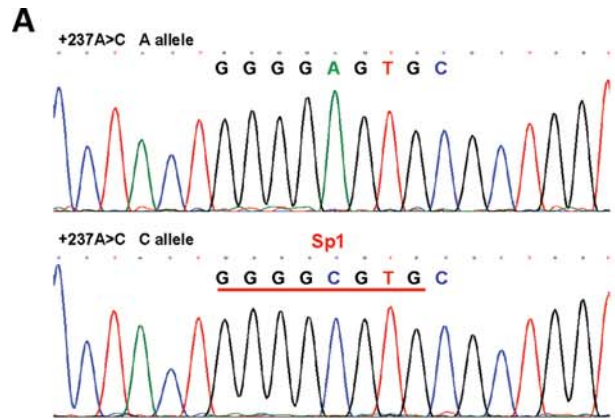
is indeed Sp1 and that Sp1 is indeed on the NQO2 promoter *in vivo*. Unfortunately, we did not use a ChIP assay for two reasons. First, the pivotal role of Sp1 in the regulation of NQO2 expression has been addressed. Sp1 binds to the promoter region of NQO2 *in vivo* (26). Second, there is another Sp1 binding site near the +237A>C polymorphism (within 100 bp), which would prevent a specific detection of Sp1 binding to +237A>C because ChIP can only be used to examine a global effect of Sp1-binding sites within a DNA fragment corresponding to +237A>C and its flanking ~100 bp sequence. Therefore, EMSA is considered to be the optimal test for examining the biolog6ical function of +237A>C.

Knowing that Sp1 (at +237A>C) and Sp3 (at 29 bp-I/D) are involved in the transcriptional regulation of the variant promoter alleles, we speculated that the paradoxical 'non-extreme' activity of the D-C haplotype compared with the other haplotypes may be related to these two transcriptional factors. We quantified the amount of Sp1 and Sp3 in our experimental cells and in normal breast tissues (Fig. 2F). Sp1 expression in most experimental cells was less than that in normal breast tissues. Correspondingly, the Sp1/Sp3 ratio in normal breast tissues was 6–20 times that in experimental cells. We hypothesized that the coexistence of variant 29 bp-I/D and +237A>C alleles requires more Sp1 for transcriptional regulation. Therefore, we transiently co-transfected increasing concentrations of a Sp1 expression vector with each of the four haplotype promoter plasmids to evaluate the responsiveness of the four promoter sequences (Fig. 2G). Because the endogenous Sp1 expression is relatively weak in 293T cells, we chose this cell line for co-transfection assays. Generally, transfection of the Sp1 expression plasmid remarkably increased NQO2 promoter activity in a dose-dependent manner. Specifically, although the D-C promoter showed twice the activity as the I-29-A promoter with low levels of Sp1, addition of the Sp1 vector could significantly induce transcriptional activity of the D-C promoter as much as 3-fold over that of the I-29-A promoter. Interestingly, the D-C promoter was twice as active as the D-A and I-29-C promoters with high levels of Sp1. These results suggest that Sp1 has a substantial role in modulating the biological function of the 29 bp-I/D and +237A>C polymorphisms.

Functional polymorphisms alter NQO2 gene expression

We employed a real-time PCR assay to examine the influence of the functional polymorphisms on NQO2 gene expression. The level of NQO2 mRNA was normalized to that of *GAPDH*. First, we measured the differential expression of NQO2 mRNA in both breast cancer tissues ($n = 38$) and normal breast tissues ($n = 71$) using a standard curves method. The 38 breast cancer tissues and 71 normal breast tissues were obtained from 71 breast cancer patients diagnosed by the Breast Surgery Department. These samples corresponded to 38 paired samples (both cancer tissue and normal tissue) and 33 normal tissue specimens. All 71 women provided blood samples as well, which allowed us to genotype 29 bp-I/D and +237A>C in the DNA samples. The genotype distributions of the 29 bp-I/D and +237A>C polymorphisms were comparable between all 71 samples from women providing normal tissues and the sub-group of 38 women who also provided cancer tissues ($P = 0.92$ and $P = 0.96$, respectively). It is important to note that 38 women were overlapping in these two groups as they provided both cancerous breast tissue and normal breast tissue. Despite the similar genotype distributions, significantly lower NQO2 mRNA expression levels were observed in breast cancer tissues compared with normal breast tissues (Mann–Whitney test $P < 0.0001$, Fig. 3B). Our results are consistent with the observation of a significant down-regulation of NQO2 in hepatocellular cancer tissues relative to normal hepatic tissues (27). Our findings indicate that the decreased NQO2 expression in cancer tissues could be due to multiple mechanisms (e.g. methylation, aberrant regulation) rather than to a regulative effect of the polymorphic alleles specifically. It is also reasonable to speculate that NQO2 may be involved in the development of breast cancer.

Because an abnormal expression pattern of NQO2 in cancer tissues interferes with the real biological effect of the functional polymorphisms, we measured NQO2 levels in normal breast tissue specimens, which represent a physiological condition. Of the 40 'susceptible' and 25 'non-susceptible' normal breast tissues examined, transcript NQO2 expression was significantly variable, and this variability appeared to be related to the 29 bp-I/D polymorphism. ID/DD genotypes displayed markedly higher expression levels when compared to the II genotype in 'susceptible' normal breast tissues ($P = 0.0019$, Fig. 3C). This tendency was also observed with the +237A>C genotypes and NQO2 expression, although with borderline significance ($P = 0.086$, Fig. 3D). In addition, the 'susceptible' specimens harboring the II-AA risk haplotypes had lower expression levels of NQO2 relative to samples with the ID/DD-AC/CC genotypes ($P = 0.0025$, Fig. 3E).

DISCUSSION

In the first case–control study, we observed a significant association between breast cancer and genetic variants/haplotypes within the NQO2 gene. To further confirm our results, we repeated our analysis in another, independent

Figure 3. Differential expression of NQO2 according to genotype. (**A**) Representative real-time PCR amplification plots. Fluorescence intensity measurements are reported on the Y-axis, and the number of PCR cycles is on the X-axis. Upper plot: a normal DD-genotype breast sample compared with a II-genotype sample. These two samples both harbor the AA-genotype of +237A>C. Middle plot: an AC-genotype sample compared with an AA-genotype sample. These two samples both harbor II-genotype. Lower plot: real-time PCR analysis of NQO2 mRNA expression in MCF-7 cells of +237AC-genotype and in MDA-MB-231 cells of +237AA-genotype. These two cell lines both harbor the I-29/D-genotype of 29 bp-I/D. The I allele in Figure 3 represents I-29 allele only. (**B**) Differential NQO2 expression in breast cancer tissues and normal breast tissues. Relative NQO2 mRNA expressions are detected in breast cancer tissues ($n = 38$) and normal breast tissues ($n = 71$). The horizontal lines represent the mean values. The mean NQO2 mRNA expression level in normal breast tissues is five times higher than in breast cancer tissues (*P*-value of Mann–Whitney test <0.0001). (**C**–**E**) Box plots of NQO2 mRNA expression according to 29 bp-I/D (44 individuals of II genotype versus 21 ID+DD) (C), +237A>C (20 AA versus 41 AC+CC) (D), or I-29/D-A/C haplotypes (20 II-AA versus 21 ID/DD-AC/CC) (E). The relative expression of NQO2 was calculated relative to *GAPDH* (NQO2/GAPDH). NQO2 expression between two groups was analyzed by the Mann–Whitney test. N.T., normal tissues; suscep., susceptible.
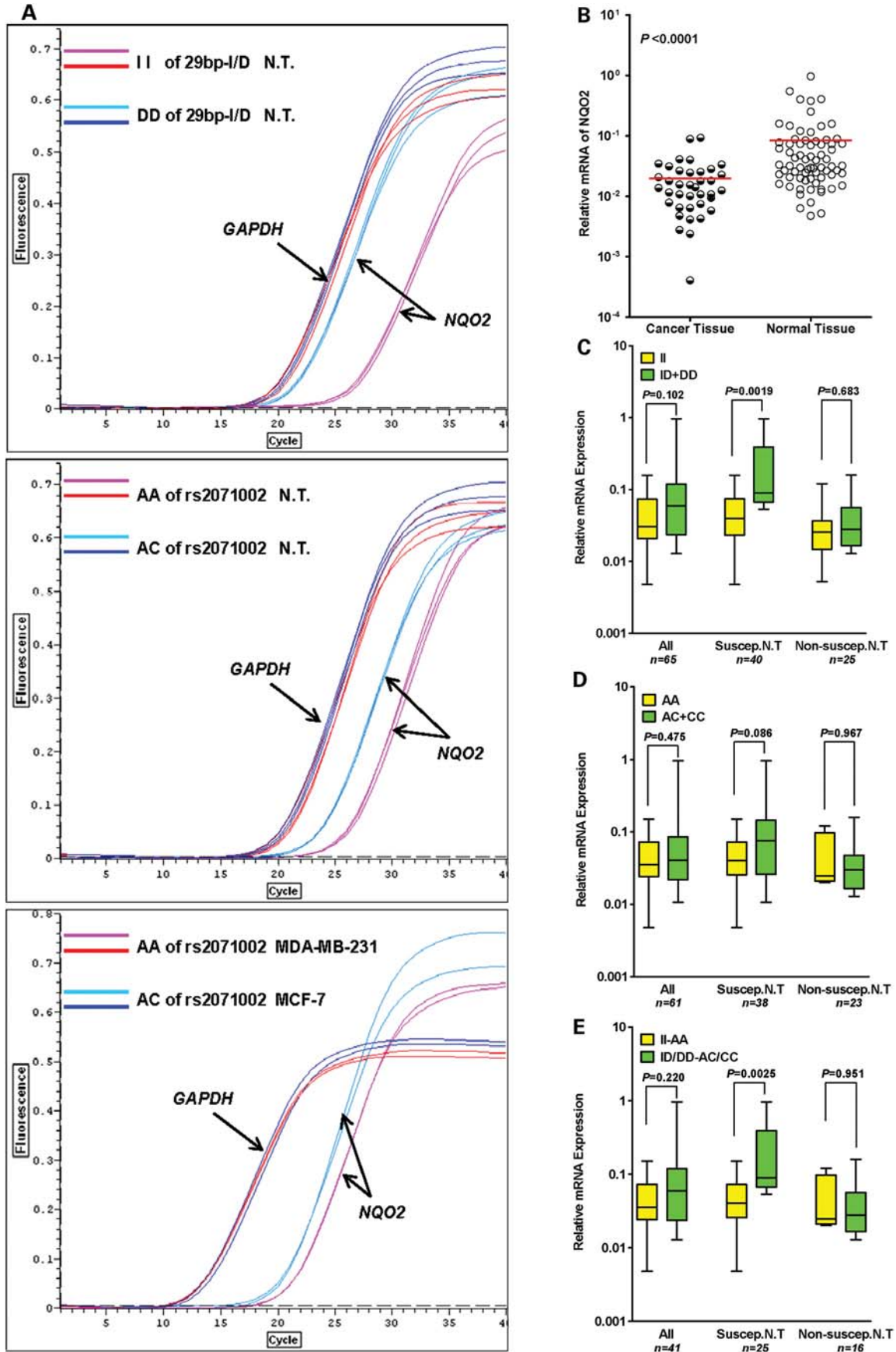
Chinese population (including familial/early-onset breast cancer cases and community-based controls) and observed the same association. Specifically, the susceptibility variants in NQO2 are related to breast carcinomas, especially those with wild-type p53. Furthermore, we reveal a potential mechanism of pathogenesis of the two susceptibility polymorphisms within the promoter region: one risk-allele, I-29 of 29 bp-I/D, introduces transcriptional repressor Sp3 binding sites, and the other risk-allele, +237A of rs2071002 (+237A>C), abolishes a binding site of the transcriptional activator Sp1. Our findings from epidemiological association studies, *in vitro* promoter reporter gene assays, EMSA assays and *ex vivo* NQO2 expression measurements consistently suggest that NQO2 is a susceptibility gene for breast cancer and that high levels of NQO2 expression in breast tissue can protect against oncogenesis. To our knowledge, this is the first study that not only evaluates the association between genetic variants in NQO2 and breast cancer risk, but also provides functional verification of that association.

Breast cancer appears to be a consequence of both genetic and environmental influences. By linkage analyses and association studies, several susceptibility loci and genes, including BRCA1, BRCA2, CHEK2, TP53 and ATM, have been identified (28,29). Considering the close functional relationship between both PALB2 and BRCA2 and also BRIP1 and BRCA1, it is conceivable that mutations and polymorphisms in PALB2 and BRIP1 may account for a proportion of BRCA1/BRCA2-negative breast cancers (30,31). Despite these significant advances, the identified genes and loci have failed to satisfactorily explain rates of inheritance. For most familial as well as sporadic cases, a substantial risk component may be due to the effects of multiple genes with lower penetrance (32). Recently, large scale genome-wide association studies identified several new, independent low-penetrance susceptibility loci that are strongly associated with breast cancer in populations of diverse ethnicity (33,34). Although whole-genome screening of cancer susceptibility loci is an available methodologically, the classic strategy of surveying carcinogenic metabolism-related pathways is still well employed.

We hypothesized that NQO2 is a candidate susceptibility gene for breast cancer because it has a known activity on estrogen-derived quinone metabolites (13) and because it is capable of stabilizing the breast tumor suppressor gene product p53 by preventing 20S proteasomal degradation (14). The hypothesis that NQO2 might be involved in breast cancer initiation prompted us to conduct case–control studies in two independent study populations of Chinese women. Our analysis was designed as an exhaustive genetic association study using a tagging polymorphism strategy. We successfully found two common polymorphisms in the promoter region of NQO2, 29 bp-I/D and +237A>C, both of which showed significant frequency differences between hospital-based sporadic breast cancer cases and controls. Subsequently, we reconfirmed the association of breast cancer with both of the susceptibility polymorphisms in another, independent set of familial/early-onset breast cancer patients and community-based controls. It is important to note that our results have been adjusted for suspected breast cancer risk factors. Adjustments reveal an independent effect of the genetic variants on the significance of our results. In addition,

our previous studies have provided us with data about the spectrum of BRCA1 (35), BRCA2 (35), BRIP1 (36) and PALB2 (37) mutations in our familial/early-onset cases. The samples chosen in the current study are free of the above-mentioned gene mutations. Therefore, the observed increase in breast cancer risk could be largely attributed to the influence of the polymorphic NQO2 gene. The combined analyses of the ascertainment and replication studies demonstrated highly significant combined *P*-values ($10^{-6}-10^{-7}$) of the two susceptibility polymorphisms in NQO2. These findings suggest that NQO2 is associated with sporadic breast cancers as well as familial/early-onset breast cancers. As we have excluded the influence of high-penetrance gene mutations such as BRCA1 and BRCA2, these results also imply that sporadic and familial/early-onset breast cancer may at least partially share similar, low-penetrant genetic susceptibility.

Consistent results from genetic association analysis epidemiologically link the NQO2 gene to breast cancer susceptibility. This association is supported by the putative relationship between NQO2 protein and breast cancer pathogenesis. As mentioned previously, NQO2 could exert its effect in two ways. First, NQO2 is involved in the detoxification of estrogen quinone metabolites (13). Second, NQO2 helps stabilize p53 by preventing its degradation by the 20S pathway (14). Considering the cross-talk between NQO2 and p53, we further speculated that NQO2 could have a stronger anti-tumor effect when coupled with wild-type p53. Although p53 germ-line mutations are highly penetrant in breast cancer, the rate of germ-line mutation is rather low in sporadic and hereditary breast cancer patients (38–40). In contrast, mutated or aberrant p53 occurs at a rate of more than 40% in breast tumor tissues (41,42). We explore the association between NQO2 and breast carcinomas with either mutated or wild-type p53, although aberrant p53 in tumor cells is more likely to relate to breast cancer progression rather than initiation. Because we did not receive all breast tumor DNA samples from the patients studied for p53 mutation detection, we referred to immunodetectable p53 as mutated p53 in this report. Positive p53 immunostaining has been used to indicate p53 mutations that increase protein half-life and lead to intracellular accumulation (43). According to p53 status, we observed a more pronounced association between susceptibility polymorphisms in NQO2 and breast cancer with wild-type p53 tumors. In summary, not only does NQO2 contribute to reducing breast cancer risk at a global level, it has a greater effect when coupled with wild-type p53. However, the sensitivity and specificity of IHC for the detection of p53 mutant forms is ~80%, indicating that not all mutant forms can be detected and that not all cases of detectable overexpression are indicative of mutations (44). Although we have obtained robust *P*-values, this limitation of mutation detection should be noted.

Because polymorphisms in the promoter region usually modulate gene transcription by interacting with *trans*-acting elements, we set out to explore the roles of susceptibility polymorphisms in altered promoter activity. Wang and Jaiswal (26) has demonstrated that the I-29 allele of 29 bp-I/D leads to decreased promoter activity, and our present work demonstrates that the +237A allele of +237A>C can abolish a Sp1 recognition site, resulting in lower promoter activity.

Through EMSA, the DNA/Sp1 complex was almost undetectable with a very low binding affinity to the +237A allele. Therefore, it is logical to speculate that the I-29-A haplotype results in an extremely low level of promoter activity. This reduced activity was indeed shown using promoter reporter assays. Unexpectedly, the D-C haplotype failed to show the maximum promoter activity. We conjectured that the coexistence of variant types of 29 bp-I/D and +237A>C requires a larger amount of the transcriptional activator Sp1. The moderate activity of the D-C haplotype could be due to insufficient Sp1 levels in our experiments. Moreover, low Sp1/Sp3 ratio allows Sp3 to easily antagonize Sp1 (45). The Sp1 vector treatment assay demonstrated that sufficient Sp1 can induce a higher activity in promoters containing the D-C haplotype. We also believe that the D-C haplotype could further induce NQO2 expression in normal breast tissues, where Sp1 is express abundantly and there is a high Sp1/Sp3 ratio. Finally, we evaluated the actual effect of the susceptibility polymorphisms on NQO2 expression in mammary glands under physiological conditions. A significant association of risk-alleles/genotypes and reduced expression of NQO2 in normal breast tissues was observed, suggesting a protective role of NQO2 against breast carcinogenesis. The differential expression caused by genetic variants was observed in 'susceptible' normal breast tissues rather than in 'non-susceptible' normal breast tissues, indicating that the predisposing loci in NQO2 coupled with some other, unknown susceptibility loci together influence NQO2 expression at a global level. In other words, NQO2 is not the only genetic factor that defines breast cancer susceptibility. This implication is in accordance with a polygenic model of breast carcinogenesis.

Our results may be subject to a type I error owing to population selection bias, stratified analysis and/or mutated p53 detection flaws. A considerable proportion of familial/early-onset patients were used in our validation study, and the different features between ascertainment population and replication population may influence our results. However, adequate sample size, conservative corrections, multivariate adjustment, successful replication and functional demonstration are capable of minimizing the impact of a type I error. Furthermore, the I-16 allele could not be sufficiently examined in the current study because it is not common enough to draw convincing conclusions about its contribution to breast cancer. Our study in a Chinese population identified women harboring heterozygous I-29/I-16 and I-16/D alleles, while another report in a western population (most subjects were Caucasian) found only homozygous I-16/I-16 alleles (19). Our findings challenge the previous assumption that the I-16 allele is only stable when paired with another I-16 allele (19). Although we did not perform an in-depth functional examination regarding the I-16 allele, available data from Wang *et al*. (19) showed a significant increase in activity of the I-16-containing promoter compared with the I-29-containing promoter. In view of the obscure role of I-16 as well as its low prevalence, we excluded the individuals harboring the I-16 allele for association analyses and treated 29 bp-I/D as a dimorphism under most circumstances in the present work. We argue that the exact association between the I-16 allele and breast cancer as well as the precise biological function of the I-16 allele should be further investigated. It is also important to point out that, as in other studies of common complex diseases, we cannot completely exclude the possibility that other, unidentified, un-genotyped or low-frequency SNPs in NQO2 may also associate with breast cancer or that the two susceptibility variants may be in LD with variants in additional risk genes in the 6pter-q12 linkage interval. Nevertheless, our studies clearly associate NQO2 polymorphisms and their biological impacts with breast cancer risk.

Taken together, our data suggest that common genetic changes in NQO2 may influence breast cancer risk, likely through an altered gene expression regulation mechanism. Increased NQO2expression can decrease the likelihood of breast cancer, especially in breast tissues with wild-type p53. Finally, the replication of our studies in other populations will further strengthen our understanding of NQO2.

MATERIALS AND METHODS

Study subjects

In the first hospital-based case–control study, 1604 unrelated women, including 893 patients with pathologically confirmed primary sporadic breast cancers (780 invasive breast cancers and 113 ductal cancers *in situ* with or without micro-invasion) and 711 female cancer-free controls were recruited between January 2005 and December 2007 at the Shanghai Cancer Hospital. Subjects were identified as genetically unrelated Han Chinese from Shanghai City and its surrounding regions. Patients with a previous history of other cancers (not breast cancer) or metastatic breast cancers were excluded. The control subjects, matched to the cases based on age and geographical area, were chosen from women who had come to the outpatient department of the Shanghai Cancer Hospital, primarily for the purpose of breast cancer screening. The selected controls were determined to be free of breast cancer by a complete physical examination, bilateral mammography, ultrasonography and biopsy when necessary. Women with a previous history of cancer were also excluded from the control group. A few of the controls underwent biopsies and were diagnosed as having benign breast disease that was pathologically confirmed as fibrocystic changes; such pathological changes do not confer an increased risk for breast cancer according to the current concepts (46). Each participant was personally interviewed by doctors in either the outpatient or inpatient department to obtain epidemiological and clinicopathological information. All of the collected data were entered into a computerized database established by the Department of Breast Surgery of the Shanghai Cancer Hospital as previously described (47). Patients and controls in the first case–control study were comparable in age [mean age ± standard deviation (SD) for the cases was 47 ± 10 years versus 47 ± 9 years for the controls, $P > 0.05$], age at menarche (cases 14.7 ± 1.5 years versus controls 14.9 ± 1.7 years, $P = 0.09$), menopausal status (39.7% of the cases versus 40.9% of the controls were postmenopausal, $P > 0.05$) and BMI (cases 23.2 ± 3.0 versus controls 23.2 ± 2.9, $P > 0.05$). Because most Chinese women are non-smokers and non-drinkers, our study populations were restricted to women who did not smoke cigarettes or drink alcohol.

We validated the polymorphisms that showed statistically significant associations with cancer in the first case–control stage in an independent population of familial/early-onset breast cancer cases and community-based controls. Since 2000, the Shanghai Cancer Hospital has conducted a multi-center hospital-based BRCA mutation screening project in order to gain a full understanding of the contribution of BRCA1/2 mutations to Han Chinese hereditary breast cancer (35,48). About 500 cases coming from five medical centers in China had been collected between 2000 and 2006, and an additional ∼100 patients were recruited from 2006 to 2008. The eligibility criteria have been described previously (35). In this study, we chose 403 unrelated familial (at least one first- or second-degree relative with breast cancer) and early-onset (breast cancer diagnosed below 35 years of age) cases from southeast China, mainly Shanghai City and its surrounding regions and provinces. All of the selected familial/early-onset cases had been tested for BRCA1 (35), BRCA2 (35), BRIP1 (36) and PALB2 (37) germ-line mutations, and no deleterious genetic changes were found. Additionally, 1039 healthy women were chosen as controls from a community-based breast cancer screening project, which was initiated in May 2008 by the Shanghai Cancer Hospital of Fudan University and the Shanghai Municipal Center for Disease Control and Prevention (CDC). This cohort program will be conducted from 2008 to 2010 and will screen more than 15 000 women aged 35–75 years at enrollment for breast cancer. The breast cancer screening methods include physical examination, mammography and breast ultrasonography. We determined that the participants in the cohort were cancer-free if they were not identified as breast cancer patients after screening and they had also reported having no history of cancer at that time. After providing informed consent, each woman was personally interviewed face-to-face by trained interviewers from the Shanghai Municipal CDC using a pre-tested standard questionnaire to obtain information about demographic factors, menstrual and reproductive histories, hormone use, dietary habits, prior disease history, physical activity, tobacco and alcohol use and family history of cancer. All participants were also measured for their current weight and height. After the interview, each subject provided 3–5 ml of venous blood. This community-based breast cancer screening project is currently ongoing and has recruited more than 3000 women. About 1200 blood genomic DNA samples of cancer-free women were extracted through standard procedures. We selected ∼1000 DNA samples as controls for the validation study. Because most data from the questionnaires have yet to enter the database, we only used the age of the controls in this study. Our work was approved by the Ethical Committee of the Shanghai Cancer Hospital of Fudan University, and each participant signed an informed consent document.

DNA/RNA preparation

Genomic DNA was extracted from 3 to 5 ml of the study individuals' peripheral blood leukocytes using Gentra's PureGene DNA Purification kit (Gentra Systems, MN, USA), according to the manufacturer's protocol, and then stored at −20°C. Breast tissue samples obtained surgically from women receiv-

ing a mastectomy, lumpectomy or reduction mammaplasty in the Breast Surgery Department of the Shanghai Cancer Hospital were snap frozen in liquid nitrogen and stored at −80°C. mRNA was extracted from frozen breast tissue specimens, as well as cultured cells, and reverse transcribed. General PCR was performed as described previously (35). Here, we defined normal breast samples from women at high risk of developing breast cancer (e.g. women having a breast cancer history, having one or more first- or second-degree relatives afflicted with breast cancer and/or ovarian cancer, or having been diagnosed with lobular carcinoma *in situ* or atypical hyperplasia) as 'susceptible' normal breast tissues. We defined breast tissue samples from healthy cancer-free women without any family history or personal history of cancer as 'non-susceptible' normal breast tissues.

Selection of genetic variants

SNPs spanning a 22.3 kb region from 2 kb upstream to 0.5 kb downstream of NQO2's transcribed sequence were surveyed in the NCBI-dbSNP and the International HapMap websites. We used the HapMap database of the Han Chinese population (HapMap Data Rel 21a/phaseII Jul 06). We selected tSNPs using the pairwise method under a restriction of MAFs >0.05 and $r^2 \geq 0.8$, aiming to identify a set of tSNPs that efficiently captures all known common variants and is likely to tag most unknown variants. In all, nine tSNPs were identified (rs1885298, rs9501910, rs4149367, rs3823096, rs2071002, rs4149361, rs10046131, rs4149352 and rs3823093) that capture all 31 alleles with a mean r^2 of 0.979. Three tSNPs only tag themselves (rs10046131, rs4149352 and rs3823093), and one tSNP (rs9378352) tags itself and one additional SNP, rs4149361. Each of these five SNPs was located in an intron and excluded from further genotyping. Finally, we chose five representative tSNPs that effectively captured 26/31 SNPs (84%). In addition, variants in NQO2 with potential functional effects (such as causing amino acid changes or alternative splicing) or locations in putative transcription factor binding sites were chosen for genotyping whenever possible. We used potentially functional polymorphisms with MAFs >1%. They consisted of four non-synonymous SNPs, rs17136117 (Gly29Glu), rs28383623 (Arg16Lys), rs17300141 (Asp58Gly) and rs1143684 (Phe47Leu), and two genetic variants, rs2070999 and 29 bp-I/D, with potential regulatory function in the putative promoter region.

Genotyping

In the first hospital-based case–control study, we used the 12-plex SNPstream system (Beckman Coulter) (49) for SNP genotyping, which was carried out by the Chinese National Human Genome Center (Shanghai). The 29 bp-I/D polymorphism was genotyped by PCR- and RFLP-based assays. In brief, we performed PCR (primers: sense, 5′-CTGCCTGG AAGTCAGCAGGGTC-3′; antisense, 5′-CTCTTTACGCAG CGCGCCTAC-3′) and used a lengthy electrophoresis-based method to discriminate the I-29, I-16 and D alleles (11,19). Because the I-29 and I-16 alleles were difficult to separate via electrophoresis because there is only a difference of

13 bp different between them, we employed a RFLP-based approach using the restriction enzyme Ava I to discriminate between the I-16 and I-29 alleles. The I-29 allele was digested into 250 and 40 fragments by Ava I, while I-16 was not digested.

In the validation study, the 29 bp-I/D polymorphism was genotyped using the assay described above; +237A>C was genotyped using a RFLP-based assay. The +237C allele, when amplified using primers (sense) 5′-GTTCCGGGCTG CTTAGGTTGGC-3′ and (antisense) 5′-AAATCCCTGAGA GCGTTCTGGC-3′ to form a 269 bp product, could be digested by Cac8I into 170 and 100 bp fragments, while the PCR product containing the +237A allele could not be digested. To ensure the quality of our genotyping results, the RFLP assay was performed by operators knowing nothing about the subjects' case and control status, and the samples were assayed in a 96-well PCR plate with a positive control, consisting of a DNA sample with known heterozygous genotype. Two research assistants (K.D.Y. and L.F.) independently examined the gel pictures and repeated the assays if they did not reach a consensus on the genotype. An adequate quantity of restriction enzyme was used to completely cleave PCR amplicons. Samples with inconsistent outcomes in two independent tests were directly sequenced. In addition, 10% of the samples were randomly selected for repeated RFLP analysis for both of the polymorphisms, and the results were 100% concordant.

HWE test and LD analysis

HWE was tested by χ^2 tests for each SNP locus. The structure of the LD block, if any, was examined using the method described by Gabriel *et al.* (23). We used the fastPHASE 1.1 program to infer an individual's haplotype based on the observed NQO2 genotypes (25). The Haploview 3.32 program was used for allele genotype analysis and haplotype analysis.

Plasmid constructs

The pGL3-Basic reporter vector from Promega (Madison, WI, USA) was used to construct luciferase reporter plasmids using standard recombination techniques, as previously described (50). Briefly, fragments spanning the −537 to +529 bp promoter region of NQO2 ($\sim$ −500 to +300 bp was sufficient for basal expression of NQO2 (26)) containing haplotypes I-29-C and D-C were cloned from individual DNA samples using Proofstart DNA polymerase (Qiagen, Hilden, Germany), the sense primer 5′-GGGGTACCCCAACCATCC GTGCGTGAC-3′, and the antisense primer 5′-GAAGATCTTC CCCGCTGATTCCTACTTG-3′. These fragments were cloned between the KpnI and BglII sites of the pGL3-Basic vector to generate pGL3-29I-C and precursor-pGL3-29D-C constructs. Both constructs were verified by direct sequencing to identify any mismatched bases besides the 29 bp polymorphism. A site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA) was used to correct those mismatched variants in the precursor-pGL3-29D-C, forming a new construct named pGL3-29D-C. Finally, both pGL3-29I-C and pGL3-29D-C were mutated to generate pGL3-29I-A and pGL3-29D-A plasmids. A human Sp1 expression vector was constructed using the pcDNA™ 3.1 Directional TOPO® Expression Kit (Invitrogen), according to

the manufacturer's instructions. All constructs were verified before use by direct sequencing.

Transient transfection and luciferase assays

Human breast cancer cells (MCF-7 and MDA-MB-231), immortal normal breast epithelial cells (HBL-100) and 293T cells were grown in complete medium consisting of DMEM supplemented with 10% heat-inactivated fetal calf serum in a humidified, 5% CO_2 incubator at 37°C. For the experiments shown in Figure 2D, cells were transfected with 500 ng of plasmid DNA (four haplotype vectors or pGL3-Basic as a negative control) and co-transfected with 10 ng of pRL-SV40 as a control for transfection efficiency. For the experiments shown in Figure 2G, 0.1, 1.0 or 3.0 μg of the Sp1 expression vector were co-transfected with 10 ng of pRL-SV40. Transfections were performed using Lipofectamine2000 (Invitrogen), according to the manufacturer's protocol. Luciferase activity was measured on a VeritasTM microplate luminometer (Turner BioSystems, Sunnyvale, CA, USA) using the Dual-Luciferase Reporter Assay System Kit (Promega, Madison, WI, USA). Each experiment was performed in triplicate at least three times. Luciferase units were calculated using the formula Firefly luciferase units/Renilla luciferase units. Fold increase was reported by defining the activity of the empty pGL3-Basic vector as one.

EMSA

Pure, human Sp1 protein was purchased from Promega. The probes and competitors for the +237A allele, the +237C allele and the Sp1 consensus binding sequence were 5′-gctcctactgggg**A**gtgcgctggtcgg-3′, 5′-gctcctactgggg**C**gtgcgctg gtcgg-3′ and 5′-ATTCGATCGGGGCGGGGCGAGC-3′, respectively. Probes were synthesized as single strands and end-labeled with biotin (Invitrogen). Identical, unlabeled oligonucleotides with same sequences were used as competitors. dsDNA was made, and EMSAs were performed using the LightShift Chemiluminescent EMSA kit (Pierce, Rockford, IL, USA), according to the manufacturer's instructions.

Real-time PCR

Extracted mRNA was subjected to real-time PCR with the SYBR Green fluorescent-based assay (TaKaRa, Japan) as previously described (51) in a fluorescence temperature cycler (Opticon, MJ Research), using the standard curves method. All samples were run in triplicate at least three times. cDNA-specific primers were designed using Primer Premier 5.00 (Premier Biosoft International). The following primers were used for gene expression detection: NQO2 sense, 5′-GAAACCCACGAAGCCTACA-3′, and antisense, 5′-CAG CACCCTATCCATCCAG-3′ (153 bp); Sp1 sense, 5′-AATG GTAATGGTGGTGGTGCCTT-3′, and antisense, 5′-CTGTT CCCCCTGACTGACTCGG-3′ (190 bp); and Sp3 sense, 5′-ATGGCTGCCTTGGACGTGGAT-3′, and antisense, 5′-CGA GCGGTGACGGCTGAGTG-3′ (143 bp). *GAPDH* was used for normalization.

Statistical analysis

Tests of association were conducted using Pearson's χ^2 test or Fisher's exact test, when appropriate. The OR and its 95% CI were also determined. A Student's *t*-test or Mann–Whitney test was used to compare continuous variables between two groups; one-way ANOVA or Kruskal–Wallis analysis was used to compare continuous variables among three or more groups. The Games–Howell procedure was employed for multiple comparisons tests when the variances were unequal. The significances of single-locus or haplotype association results were corrected using the Bonferroni correction and 1000-time permutation test. Logistic regression was used to analyze the association between a single locus and breast cancer risk. A *P*-value less than or equal to 0.05 was considered statistically significant. Statistical analysis was performed using Stata/SE version 10.0 (SAS Institute, Cary, NC, USA) and SPSS Software version 12.0 (SPSS, Chicago, IL, USA).

Web resources

dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
HapMap, http://www.hapmap.org/
NCBI, http://www.ncbi.nlm.nih.gov/
TESS website for in silico analysis, http://www.cbil.upenn.edu/cgi-bin/tess/

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Yager, J.D. and Davidson, N.E. (2006) Estrogen carcinogenesis in breast cancer. *N. Engl. J. Med.*, **354**, 270–282.
2. Le Marchand, L., Donlon, T., Kolonel, L.N., Henderson, B.E. and Wilkens, L.R. (2005) Estrogen metabolism-related genes and breast cancer risk: the multiethnic cohort study. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 1998–2003.
3. Haiman, C.A., Stram, D.O., Pike, M.C., Kolonel, L.N., Burtt, N.P., Altshuler, D., Hirschhorn, J. and Henderson, B.E. (2003) A comprehensive haplotype analysis of CYP19 and breast cancer risk: the Multiethnic Cohort. *Hum. Mol. Genet.*, **12**, 2679–2692.
4. Li, K.M., Todorovic, R., Devanesan, P., Higginbotham, S., Kofeler, H., Ramanathan, R., Gross, M.L., Rogan, E.G. and Cavalieri, E.L. (2004) Metabolism and DNA binding studies of 4-hydroxyestradiol and estradiol-3,4-quinone in vitro and in female ACI rat mammary gland in vivo. *Carcinogenesis*, **25**, 289–297.
5. Talalay, P. (1989) Mechanisms of induction of enzymes that protect against chemical carcinogenesis. *Adv. Enzyme Regul.*, **28**, 237–250.
6. Jaiswal, A.K. (2000) Characterization and partial purification of microsomal NAD(P)H:quinone oxidoreductases. *Arch. Biochem. Biophys.*, **375**, 62–68.
7. Menzel, H.J., Sarmanova, J., Soucek, P., Berberich, R., Grunewald, K., Haun, M. and Kraft, H.G. (2004) Association of NQO1 polymorphism with spontaneous breast cancer in two independent populations. *Br. J. Cancer*, **90**, 1989–1994.
8. Fagerholm, R., Hofstetter, B., Tommiska, J., Aaltonen, K., Vrtel, R., Syrjakoski, K., Kallioniemi, A., Kilpivaara, O., Mannermaa, A., Kosma, V.M. *et al.* (2008) NAD(P)H:quinone oxidoreductase 1 NQO1*2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer. *Nat. Genet.*, **40**, 844–853.
9. Long, D.J. and Jaiswal, A.K. (2000) NRH:quinone oxidoreductase2 (NQO2). *Chem. Biol. Interact.*, **129**, 99–112.
10. Celli, C.M., Tran, N., Knox, R. and Jaiswal, A.K. (2006) NRH:quinone oxidoreductase 2 (NQO2) catalyzes metabolic activation of quinones and anti-tumor drugs. *Biochem. Pharmacol.*, **72**, 366–376.
11. Harada, S., Fujii, C., Hayashi, A. and Ohkoshi, N. (2001) An association between idiopathic Parkinson's disease and polymorphisms of phase II detoxification enzymes: glutathione S-transferase M1 and quinone oxidoreductase 1 and 2. *Biochem. Biophys. Res. Commun.*, **288**, 887–892.
12. Iskander, K., Paquet, M., Brayton, C. and Jaiswal, A.K. (2004) Deficiency of NRH:quinone oxidoreductase 2 increases susceptibility to 7,12-dimethylbenzva)anthracene and benzo(a)pyrene-induced skin carcinogenesis. *Cancer Res.*, **64**, 5925–5928.
13. Gaikwad, N.W., Yang, L., Rogan, E.G. and Cavalieri, E.L. (2009) Evidence for NQO2-mediated reduction of the carcinogenic estrogen ortho-quinones. *Free Radic. Biol. Med.*, **46**, 253–262.
14. Gong, X., Kole, L., Iskander, K. and Jaiswal, A.K. (2007) NRH:quinone oxidoreductase 2 and NAD(P)H:quinone oxidoreductase 1 protect tumor suppressor p53 against 20s proteasomal degradation leading to stabilization and activation of p53. *Cancer Res.*, **67**, 5380–5388.
15. Liu, X., Holstege, H., van der Gulden, H., Treur-Mulder, M., Zevenhoven, J., Velds, A., Kerkhoven, R.M., van Vliet, M.H., Wessels, L.F., Peterse, J.L. *et al.* (2007) Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer. *Proc. Natl Acad. Sci. USA*, **104**, 12111–12116.
16. Orphanos, V., McGown, G., Hey, Y., Boyle, J.M. and Santibanez-Koref, M. (1995) Proximal 6q, a region showing allele loss in primary breast cancer. *Br. J. Cancer*, **71**, 290–293.
17. Piao, Z., Lee, K.S., Kim, H., Perucho, M. and Malkhosyan, S. (2001) Identification of novel deletion regions on chromosome arms 2q and 6p in breast carcinomas by amplotype analysis. *Genes Chromosomes Cancer*, **30**, 113–122.
18. Lerebours, F., Bertheau, P., Bieche, I., Driouch, K., De The, H., Hacene, K., Espie, M., Marty, M. and Lidereau, R. (2002) Evidence of chromosome regions and gene involvement in inflammatory breast cancer. *Int. J. Cancer*, **102**, 618–622.
19. Wang, W., Le, W.D., Pan, T., Stringer, J.L. and Jaiswal, A.K. (2008) Association of NRH:quinone oxidoreductase 2 gene promoter polymorphism with higher gene expression and increased susceptibility to Parkinson's disease. *J. Gerontol. A Biol. Sci. Med. Sci.*, **63**, 127–134.
20. Gyorffy, B., Kocsis, I. and Vasarhelyi, B. (2004) Missed calculations and new conclusions: re-calculation of genotype distribution data published in Journal of Investigative Dermatology, 1998–2003. *J. Invest. Dermatol.*, **122**, 644–646.
21. Deng, H.W., Chen, W.M. and Recker, R.R. (2001) Population admixture: detection by Hardy-Weinberg test and its quantitative effects on linkage-disequilibrium methods for localizing genes underlying complex traits. *Genetics*, **157**, 885–897.
22. Wittke-Thompson, J.K., Pluznikov, A. and Cox, N.J. (2005) Rational inferences about departures from Hardy–Weinberg equilibrium. *Am. J. Hum. Genet.*, **76**, 967–986.
23. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.
24. Yin, W.J., Lu, J.S., Di, G.H., Lin, Y.P., Zhou, L.H., Liu, G.Y., Wu, J., Shen, K.W., Han, Q.X., Shen, Z.Z. *et al.* (2008) Clinicopathological features of the triple-negative tumors in Chinese breast cancer patients. *Breast Cancer Res. Treat.* In press.

25. Scheet, P. and Stephens, M. (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.*, **78**, 629–644.

26. Wang, W. and Jaiswal, A.K. (2004) Sp3 repression of polymorphic human NRH:quinone oxidoreductase 2 gene promoter. *Free Radic. Biol. Med.*, **37**, 1231–1243.

27. Strassburg, A., Strassburg, C.P., Manns, M.P. and Tukey, R.H. (2002) Differential gene expression of NAD(P)H:quinone oxidoreductase and NRH:quinone oxidoreductase in human hepatocellular and biliary tissue. *Mol. Pharmacol.*, **61**, 320–325.

28. Walsh, T. and King, M.C. (2007) Ten genes for inherited breast cancer. *Cancer Cell*, **11**, 103–105.

29. Johnson, N., Fletcher, O., Palles, C., Rudd, M., Webb, E., Sellick, G., dos Santos Silva, I., McCormack, V., Gibson, L., Fraser, A. *et al.* (2007) Counting potentially functional variants in BRCA1, BRCA2 and ATM predicts breast cancer susceptibility. *Hum. Mol. Genet.*, **16**, 1051–1057.

30. Chen, P., Liang, J., Wang, Z., Zhou, X., Chen, L., Li, M., Xie, D., Hu, Z., Shen, H. and Wang, H. (2008) Association of common PALB2 polymorphisms with breast cancer risk: a case–control study. *Clin. Cancer Res.*, **14**, 5931–5937.

31. Song, H., Ramus, S.J., Kjaer, S.K., Hogdall, E., Dicioccio, R.A., Whittemore, A.S., McGuire, V., Hogdall, C., Jacobs, I.J., Easton, D.F. *et al.* (2007) Tagging single nucleotide polymorphisms in the BRIP1 gene and susceptibility to breast and ovarian cancer. *PLoS ONE*, **2**, e268.

32. Peto, J. (2002) Breast cancer susceptibility-A new look at an old model. *Cancer Cell*, **1**, 411–412.

33. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R. *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447**, 1087–1093.

34. Easton, D.F. and Eeles, R.A. (2008) Genome-wide association studies in cancer. *Hum. Mol. Genet.*, **17**, R109–R115.

35. Li, W.F., Hu, Z., Rao, N.Y., Song, C.G., Zhang, B., Cao, M.Z., Su, F.X., Wang, Y.S., He, P.Q., Di, G.H. *et al.* (2008) The prevalence of BRCA1 and BRCA2 germline mutations in high-risk breast cancer patients of Chinese Han nationality: two recurrent mutations were identified. *Breast Cancer Res. Treat.*, **110**, 99–109.

36. Cao, A.Y., Huang, J., Hu, Z., Li, W.F., Ma, Z.L., Tang, L.L., Zhang, B., Su, F.X., Zhou, J., Di, G.H. *et al.* (2008) Mutation analysis of BRIP1/BACH1 in BRCA1/BRCA2 negative Chinese women with early onset breast cancer or affected relatives. *Breast Cancer Res. Treat.* In press.

37. Cao, A.Y., Huang, J., Hu, Z., Li, W.F., Ma, Z.L., Tang, L.L., Zhang, B., Su, F.X., Zhou, J., Di, G.H. *et al.* (2009) The prevalence of PALB2 germline mutations in BRCA1/BRCA2 negative Chinese women with early onset breast cancer or affected relatives. *Breast Cancer Res. Treat.*, **114**, 457–462.

38. Sensi, E., Tancredi, M., Aretini, P., Cipollini, G., Naccarato, A.G., Viacava, P., Bevilacqua, G. and Caligo, M.A. (2003) p53 inactivation is a rare event in familial breast tumors negative for BRCA1 and BRCA2 mutations. *Breast Cancer Res. Treat.*, **82**, 1–9.

39. Frebourg, T., Barbier, N., Yan, Y.X., Garber, J.E., Dreyfus, M., Fraumeni, J. Jr, Li, F.P. and Friend, S.H. (1995) Germ-line p53 mutations in 15 families with Li-Fraumeni syndrome. *Am. J. Hum. Genet.*, **56**, 608–615.

40. Walsh, T., Casadei, S., Coats, K.H., Swisher, E., Stray, S.M., Higgins, J., Roach, K.C., Mandell, J., Lee, M.K., Ciernikova, S. *et al.* (2006) Spectrum of mutations in BRCA1, BRCA2, CHEK2 and TP53 in families at high risk of breast cancer. *JAMA*, **295**, 1379–1388.

41. Coles, C., Condie, A., Chetty, U., Steel, C.M., Evans, H.J. and Prosser, J. (1992) p53 mutations in breast cancer. *Cancer Res.*, **52**, 5291–5298.

42. Hartmann, A., Blaszyk, H., Saitoh, S., Tsushima, K., Tamura, Y., Cunningham, J.M., McGovern, R.M., Schroeder, J.J., Sommer, S.S. and Kovach, J.S. (1996) High frequency of p53 gene mutations in primary breast cancers in Japanese women, a low-incidence population. *Br. J. Cancer*, **73**, 896–901.

43. Visakorpi, T., Kallioniemi, O.P., Heikkinen, A., Koivula, T. and Isola, J. (1992) Small subgroup of aggressive, highly proliferative prostatic carcinomas defined by p53 accumulation. *J. Natl. Cancer Inst.*, **84**, 883–887.

44. Alsner, J., Jensen, V., Kyndi, M., Offersen, B.V., Vu, P., Borresen-Dale, A.L. and Overgaard, J. (2008) A comparison between p53 accumulation determined by immunohistochemistry and TP53 mutations as prognostic variables in tumours from breast cancer patients. *Acta Oncol.*, **47**, 600–607.

45. Ghayor, C., Chadjichristos, C., Herrouin, J.F., Ala-Kokko, L., Suske, G., Pujol, J.P. and Galera, P. (2001) Sp3 represses the Sp1-mediated transactivation of the human COL2A1 gene in primary and de-differentiated chondrocytes. *J. Biol. Chem.*, **276**, 36881–36895.

46. Santen, R.J. and Mansel, R. (2005) Benign breast disorders. *N. Engl. J. Med.*, **353**, 275–285.

47. Yu, K.D., Di, G.H., Wu, J., Lu, J.S., Shen, K.W., Shen, Z.Z. and Shao, Z.M. (2007) Development and trends of surgical modalities for breast cancer in china: a review of 16-year data. *Ann. Surg. Oncol.*, **14**, 2502–2509.

48. Song, C.G., Hu, Z., Wu, J., Luo, J.M., Shen, Z.Z., Huang, W. and Shao, Z.M. (2006) The prevalence of BRCA1 and BRCA2 mutations in eastern Chinese women with breast cancer. *J. Cancer Res. Clin. Oncol.*, **132**, 617–626.

49. Bell, P.A., Chaturvedi, S., Gelfand, C.A., Huang, C.Y., Kochersperger, M., Kopla, R., Modica, F., Pohl, M., Varde, S., Zhao, R. *et al.* (2002) SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques*, **32**(Suppl.), 70–72, 74, 76–77.

50. Moshynska, O., Moshynskyy, I., Misra, V. and Saxena, A. (2005) G125A single-nucleotide polymorphism in the human BAX promoter affects gene expression. *Oncogene*, **24**, 2042–2049.

51. Aldea, C., Alvarez, C.P., Folgueira, L., Delgado, R. and Otero, J.R. (2002) Rapid detection of herpes simplex virus DNA in genital ulcers by real-time PCR using SYBR green I dye as the detection signal. *J. Clin. Microbiol.*, **40**, 1060–1062.