LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-201871

A prospective study of XRCC1 haplotypes and their interaction with plasma carotenoids on breast cancer risk

J. Han , S. E. Hankinson , I. De Vivo , D. Spiegelman , R. M. Tamimi , H. W. Mohrenweiser , G. A. Colditz , D. J. Hunter

January 16, 2004

Cancer Research

A Prospective Study of XRCC1 Haplotypes and Their Interaction with Plasma Carotenoids on Breast Cancer Risk

Jiali Han,[1,3] Susan E. Hankinson,[2,4] Immaculata De Vivo,[2,3,4] Donna Spiegelman,[2] Rulla M. Tamimi,[2] Harvey W. Mohrenweiser,[5] Graham A. Colditz,[2,3,4] and David J. Hunter[1,2,3,4]

Departments of [1]Nutrition and [2]Epidemiology and the [3]Harvard Center for Cancer Prevention, Harvard School of Public Health, Boston, Massachusetts; [4]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts; and [5]Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California

ABSTRACT

The XRCC1 protein is involved in the base excision repair pathway through interactions with other proteins. Polymorphisms in the *XRCC1* gene may lead to variation in repair proficiency and confer inherited predisposition to cancer. We prospectively assessed the associations between polymorphisms and haplotypes in *XRCC1* and breast cancer risk in a nested case-control study within the Nurses' Health Study (incident cases, *n* 1004; controls, *n* 1385). We further investigated gene-environment interactions between the *XRCC1* variations and plasma carotenoids on breast cancer risk. We genotyped four haplotype-tagging single nucleotide polymorphisms (Arg^{194}Trp, C26602T, Arg^{399}Gln, and Gln^{632}Gln) in the*XRCC1* gene. Five common haplotypes accounted for 99% of the chromosomes in the present study population of mostly Caucasian women. We observed a marginally significant reduction in the risk of breast cancer among ^{194}Trp carriers. As compared with no-carriers, women with at least one ^{194}Trp allele had a multivariate odds ratio of 0.79 (95% of the confidence interval, 0.60 –1.04). The inferred haplotype harboring the ^{194}Trp allele was more common in controls than in cases (6.6 *versus* 5.3%, *P* 0.07). We observed that the Arg^{194}Trp modified the inverse associations of plasma -carotene level (*P*, ordinal test for interaction 0.02) and plasma -carotene level (*P*, ordinal test for interaction 0.003) with breast cancer risk. No suggestion of an interaction was observed between the Arg^{194}Trp and cigarette smoking. Our results suggest an inverse association between *XRCC1* ^{194}Trp allele and breast cancer risk. The findings of the effect modification of the Arg^{194}Trp on the relations of plasma -and -carotene levels with breast cancer risk suggest a potential protective effect of carotenoids in breast carcinogenesis by preventing oxidative DNA damage.

INTRODUCTION

It has been suggested that reduced DNA repair capacity may be a susceptibility factor predisposing women to breast cancer (1– 6). Genetic variations in DNA repair genes may lead to interindividual variation in DNA repair capacity and modify the associations between exogenous and endogenous carcinogens and breast cancer risk.

The XRCC1 protein is involved in the BER[6] pathway. BER is responsible for repair of a wide variety of nonbulky exogenous and endogenous base damage and single strand breaks (7, 8). Although XRCC1 has no known enzymatic activity, there are three distinct domains that are sites of interaction with DNA polymerase (amino acids 1–183; Refs. 9, 10), poly(ADP-ribose) polymerase (BRCT-I, amino acids 384–476), and DNA ligase III (BRCT-II, amino acids 573–592; Refs. 11–13). More recently, XRCC1 was also reported to interact with polynucleotide kinase (14) and APE1 (15). This suggests that XRCC1 may act as a nucleating factor in BER by bringing different components together at the site of action to promote the efficiency of the repair machinery. A number of SNPs in *XRCC1* have been identified (16, 17). These polymorphisms may alter BER proficiency and, in turn, confer genetic predisposition to breast cancer. We prospectively assessed whether candidate polymorphisms and haplotypes in the *XRCC1* gene are associated with breast cancer risk in a nested case-control study within the Nurses' Health Study.

Reactive oxygen species and free radicals can cause oxidative DNA damage, which is mainly repaired by BER. Carotenoids can neutralize reactive oxygen species and reduce oxidative DNA damage because of their antioxidant properties (18, 19). Among carotenoids, higher plasma -carotene or -carotene level was significantly associated with a decreased risk of breast cancer in this nested case-control study.[7] We further investigated the gene-environment interaction hypotheses that *XRCC1* genetic variations may modify the associations of plasma carotenoid levels with breast cancer risk.

MATERIALS AND METHODS

Study Population. The Nurses' Health Study was established in 1976 when 121,700 female registered nurses between the ages of 30 and 55 completed a self-administered questionnaire on their medical histories and baseline health-related exposures. Updated information has been obtained by questionnaires every 2 years. Incident breast cancers were identified by self-report and confirmed by medical record review. Between 1989 and 1990, blood samples were collected from 32,826 of the cohort members. Subsequent follow-up has been greater than 98% for this subcohort.

Eligible cases in this study consisted of women with pathologically confirmed incident breast cancer from the subcohort who gave a blood specimen. Cases with a diagnosis anytime after blood collection up to June 1, 1998, with no previously diagnosed cancer except for nonmelanoma skin cancer were included. One or two controls were randomly selected among women who gave a blood sample and were free of diagnosed cancer (excluding nonmelanoma skin cancer) up to and including the questionnaire cycle in which the case was diagnosed. Controls were matched to cases on year of birth, menopausal status, postmenopausal hormone use at blood collection, month of blood return, time of day of blood collection, and fasting status at blood draw; menopause was defined as described previously (20). The nested case-control study consists of 1004 incident breast cancer cases and 1385 matched controls. The study protocol was approved by the Committee on Use of Human Subjects of the Brigham and Women's Hospital (Boston, MA).

Exposure Data. Information regarding breast cancer risk factors was obtained from the 1976 baseline questionnaire, subsequent biennial questionnaires, and a questionnaire completed at the time of blood sampling. Menopausal status and use of postmenopausal hormones were assessed at blood draw and updated until date of diagnosis for cases and the equivalent date for matched controls. First-degree family history of breast cancer was asked in 1982 and updated in subsequent questionnaires. Information regarding cigarette smoking was asked in the 1976 questionnaire and updated in subsequent questionnaires (21).

[7] R. M. Tamimi, S. E. Hankinson, H. Campos, D. Spiegelman, S. Zhang, G. A. Colditz, W. C. Willett, and D. J. Hunter. Plasma antioxidants and risk of breast cancer, submitted for publication.

Haplotype Analysis. We selected haplotype-tagging SNPs for the *XRCC1* gene, using data derived from the resequencing of the exons and adjacent intronic and noncoding regions of the gene in a multiple ethnicity group of 90 samples from the NIH DNA Polymorphism Discovery Resource available from the Coriell Institute for Medical Research (22). The PCR products for direct sequencing included the splice sites and the 5 and 3 regions of the gene (16, 17). The 90 samples are from United States residents representing the ethnic diversity of the United States population, but the ethnicity of individual samples is unknown. The 90 individuals in this subset are from the following ethnic groups: 23 European Americans; 23 African Americans; 11 Mexican Americans; 11 Native Americans; and 22 Asian Americans. We performed haplotype estimation using the Partition-Ligation EM Algorithm of Qin *et al.* (23). On the basis of seventeen common (1% allele frequency) SNPs found in the *XRCC1* gene (Table 1), five common haplotypes were inferred with frequency above a threshold of 1.3% (0.05 23/90 0.013), which was set to ascertain potential common (5%) Caucasian-specific haplotypes from the mixed population. The five common haplotypes accounted for 87% of the alleles of this mixed population group. Only 4 of the 17 SNPs are necessary to reconstruct the five common haplotypes (Table 1). We thus genotyped four haplotype-tagging SNPs [C26304T(Arg^{194}Trp), C26602T, G28152A(Arg^{399}Gln), and G36189A(Gln^{632}Gln)] in the present case-control study of mostly Caucasian women.

Laboratory Assays. To maximize power, we included all available plasma samples, including samples left unpaired because of either subsequent exclusions (of either the case or controls) or laboratory issues. Plasma samples of 994 cases and 995 controls were assessed using reversed-phase high-perfor-mance liquid chromatography methods described by El-Sohemy *et al.* (24) in the Micronutrient Analysis Laboratory in the Department of Nutrition at the Harvard School of Public Health to determine the concentrations of -caro-tene, -carotene, -cryptoxanthin, lycopene, lutein/zeaxanthin, -tocopherol, and -tocopherol. All case control pairs were assayed together; the samples were ordered randomly and labeled within each pair. Plasma quality control samples were interspersed to assess laboratory precision. Coefficient of variation for each antioxidant, weighted by the proportion of samples on a batch-and pool-specific basis, was 10%. Laboratory personnel were blinded to case control status and identity of replicate samples.[7]

Genotyping assays were performed by Pyrosequencing (Pyrosequencing, Inc., Uppsala, Sweden; Ref. 25), RFLP, and the 5 nuclease assay (TaqMan) using the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). The primers, probes, enzymes, and conditions for genotyping assays are available upon request. Genotyping was performed by laboratory personnel blinded to case-control status, and blinded quality control samples were inserted to validate genotyping procedures; concordance for the blinded samples was 100%.

Statistical Analysis. We used a ^{2} test to assess whether the *XRCC1* genotypes were in Hardy-Weinberg equilibrium and to determine the *P* for difference in haplotype frequencies between cases and controls. Conditional logistic regression was used to calculate ORs and 95% CIs to assess the risk of breast cancer for *XRCC1* genotypes among all women. Unconditional logistic regression was used in analyses stratified by genotype. Plasma levels of antioxidants were categorized into quartiles based on the batch-specific cut points of control subjects. Weighted median values for each quartile were based on batch-specific medians of controls weighted by the proportion of subjects in each batch. In multivariate analysis, in addition to the matching variables, we adjusted for the following breast cancer risk factors: BMI (kg/m^2) at age 18 years, weight gain since age 18, age at menarche, age at menopause, parity/age at first birth, first-degree family history of breast cancer, personal history of benign breast disease, and duration of postmenopausal hormone use. Alcohol consumption was based on the 1990 dietary questionnaire; the 1986 questionnaire was used for individuals who did not provide this information on the 1990 questionnaire.

Tests for trend were conducted by assigning the weighted median values for quartiles of plasma

antioxidant levels and the median values for strata of smoking duration among controls to both cases and controls as continuous variables. To test statistical significance of interactions between the environmental exposures and the *XRCC1* genotype, we first used a LRT to compare nested models that included terms for all combinations of the *XRCC1* genotype and levels of environmental exposure to the models with indicator variables for the main effects only (nominal LRT). We also modeled *XRCC1* genotypes as ordinal variables and environmental exposures as continuous variables as described for trend tests to assess the statistical significance of interactions by LRT test of a single interaction term (ordinal LRT). All *P*s were two-sided.

RESULTS

Descriptive Characteristics of Cases and Controls. The mean age at blood draw of cases was 57.3 years and that of controls was
57.8 years. There were 20.1% premenopausal, 70.7% postmenopausal women, and 9.2% women with uncertain menopausal status at blood draw. Cases and controls had similar mean BMI at blood draw (25.4 *versus* 25.5 kg/m^2) and weight gain since age 18 years (11.8 *versus* 11.4 kg). Compared with controls, cases had similar ages at menarche (12.5 *versus* 12.6 years), first birth (24.8 *versus* 24.8 years), and menopause (48.2 *versus* 48.0 years). The proportion of women with a first-degree family history of breast cancer was higher among the cases (20.9 *versus* 14.8%). Cases were more likely to have a personal history of benign breast disease (65.0 *versus* 49.6%), a longer duration of postmenopausal hormone use (38.0 *versus* 26.9% current users for 5 years), and a longer duration of smoking (13.1 *versus* 12.4 years). Cases and controls had similar median values of plasma -carotene (0.10 *versus* 0.11 mol/liters) and -carotene (0.44 *versus* 0.45 mol/ liters) concentrations.

XRCC1 and Breast Cancer Risk. The five common haplotypes inferred from the four haplotype-tagging SNPs accounted for 99% of the chromosomes in the present study population of mostly Caucasian women (Table 2; self-reported major ethnicity was similar between cases and controls in this study population; cases *versus* controls: Southern European, 17.5 *versus* 17.8%; Scandinavian, 8.6 *versus* 7.6%; and other Caucasian, 63.1 *versus* 61.3%; Asians, Hispanics, and African Americans comprised 1% of the total cases and controls). Five rare haplotypes were also inferred. The haplotype harboring the variant allele ^{194}Trp was more common in controls than in cases (6.6 *versus* 5.3%, *P* 0.07). There was no notable difference in frequency distribution in cases and controls for other haplotypes. Among these five common haplotypes, one haplotype consisted of the common allele at each polymorphic site; each of the other four haplotypes carried only one variant allele of the four sites exclusively. In other words, none of these four SNPs were in linkage disequilibrium. Individuals with the variant allele of one polymorphism carried the haplotype that only harbored this variant but not others. Therefore, the carriage of a polymorphism can be viewed as the carriage of the corresponding haplotype in the evaluation of the main effect of polymorphic sites and gene-environment interactions. Additionally, it is noteworthy that there is another nonsynonymous polymorphism Arg^{280}His, which is in 100% genotype concordance with C26602T among the 90 individuals in the NIH DNA Polymorphism Discovery Resource, according to the *XRCC1* resequencing data of National Institute of Environmental Health Sciences Environmental Genome Project at the University of Washington.[8]

The distribution of all *XRCC1* genotypes among the controls was in Hardy-Weinberg equilibrium. We observed a marginally significant reduction in the risk of breast cancer among ^{194}Trp carriers. As compared with noncarriers, women with at least one ^{194}Trp allele had a decreased risk of breast cancer (multivariate OR, 0.79; 95% CI, 0.60–1.04; Table 3). No overall associations between C26602T, Arg^{399}Gln, and Gln^{632}Gln and breast cancer risk were found. For the two nonsynonymous polymorphisms Arg^{194}Trp and Arg^{399}Gln, carriage of either variant allele was not associated with altered breast cancer risk compared with being homozygous wild-type for both alleles (data not shown).

Plasma -/-Carotene and Arg^{194}Trp on Breast Cancer Risk. We assessed whether the associations between the plasma carotenoid levels and breast cancer risk differed by Arg^{194}Trp. The ORs for breast

cancer risk by Arg^{194}Trp and plasma -carotene and -caro-tene levels are shown in Table 4. We observed a significant 60% reduction in risk (multivariate OR, 0.40; 95% CI, 0.21–0.77) for ^{194}Trp carriers in the highest -carotene quartile *versus* noncarriers in the lowest quartile. The ^{194}Trp carriers in the third quartile also had a significantly decreased risk compared with noncarriers in lowest quartile (OR, 0.53; 95% CI, 0.30–0.97). The inverse association between plasma -carotene level and breast cancer risk was significantly different between ^{194}Trp carriers and noncarriers (*P*, test for ordinal interaction 0.02). Incorporating the cross-classified interaction terms also significantly increased the goodness fit of the model (*P*, test for nominal interaction 0.03). This pattern was similar for plasma -carotene level, with a 68% reduction in risk (multivariate OR, 0.32; 95% CI, 0.16–0.61) in the top quartile among ^{194}Trp carriers compared with noncarriers in the bottom quartile. Compared with non-carriers, the inverse association between -carotene level and breast cancer risk was significantly stronger among ^{194}Trp carriers (*P*, test for trend 0.003 *versus* 0.07; *P*, test for ordinal interaction 0.003). The cross-classified interaction between plasma -carotene level and *XRCC1* 194 genotype on breast cancer risk was statistically significant (*P*, test for nominal interaction 0.02). Of note, in these analyses of interactions between *XRCC1* Arg^{194}Trp and plasma -or -carotene level, the multivariate ORs did not change materially, and the tests of interaction remained significant after controlling for plasma -carotene (in -carotene analyses), -carotene (in -carotene analyses), -cryptoxanthin, lycopene, -tocopherol, -tocopherol, lu-tein/zeaxanthin, and folate, one at a time or all simultaneously. No statistically significant interactions were observed between the Arg^{194}Trp genotype and other plasma antioxidants such as -cryp-toxanthin, lycopene, -tocopherol, -tocopherol, and lutein/zeaxan-thin on breast cancer risk (data not shown).

Duration of Smoking and Arg^{194}Trp on Breast Cancer Risk. The positive association of duration of smoking with breast cancer risk was stronger among ^{194}Trp noncarriers than carriers (*P*, test for trend 0.06 *versus* 0.40; Table 5). Among noncarriers, the multivariate OR was 1.31 (95% CI, 0.98–1.74) for women who smoked 35 years compared with never-smokers. However, there was no evidence of a statistically significant interaction between duration of smoking and the Arg^{194}Trp on breast cancer risk (*P*, test for nominal interaction 0.76; test for ordinal interaction 0.90).

We did not observe significant interactions between the other three SNPs (C26602T, Arg^{399}Gln, and Gln^{632}Gln) and the above exposures on breast cancer risk in this study.

DISCUSSION

In this prospective nested case-control study, we genotyped four haplotype-tagging SNPs in the *XRCC1* gene. We observed that *XRCC1* ^{194}Trp carriers were at marginally significant decreased risk of breast cancer; the inferred haplotype harboring the ^{194}Trp allele was nonsignificantly less frequent in cases than in controls. We observed that the Arg^{194}Trp modified the associations of plasma -and -carotene levels with breast cancer risk. The prospective design blood sample collection before case diagnosis, the relatively large number of incident cases, and the high follow-up rates strengthen the validity of this study.

We inferred five common haplotypes from seventeen SNPs (1% allele frequency; Table 1) based on data from the resequencing of the exons and adjacent intronic and noncoding regions of the *XRCC1* gene on 90 samples from the NIH DNA Polymorphism Discovery Resource (17). We used a threshold of 1.3% (0.05 23/90 0.013) to ascertain alleles that occurred at 5% prevalence among the 23 Caucasians in the 90 individuals in the sample set used for resequencing. This is a conservative procedure because (*a*) it is likely that there are few common SNPs that are unique to Caucasians and (*b*) because of the population admixture it is estimated that 40% of the chromosomes in these 90 people are of Caucasian origin (22). Furthermore, two recent studies of haplotype variation in different ethnic groups (26, 27) reported substantial conservation of haplotypes among ethnicities with fewest population-specific common

haplotypes in Caucasians. This suggests that use of the NIH DNA Polymorphism Discovery Resource to infer haplotype in our study of mainly Caucasian individuals is unlikely to have missed defining common haplotypes in Caucasians. We genotyped four haplotype-tagging SNPs that were adequate to reconstruct the five common haplotypes. Only the same five haplotypes inferred from the genotyping of these four SNPs in this study were common (5%).

One previously published study assessed two coding region polymorphisms of *XRCC1* and breast cancer risk in whites. Duell *et al.* (28) reported data on the associations of Arg^{194}Trp and Arg^{399}Gln with breast cancer risk in a population-based case-control study in North Carolina. These authors observed that ^{194}Trp carriers in whites were at a nonsignificantly decreased risk of breast cancer (OR, 0.7; 95% CI, 0.4–1.3; 251 cases and 234 controls), and ^{399}Gln carriers in whites had unaltered breast cancer risk (OR, 1.0; 95% CI, 0.8–1.4; 386 cases and 381 controls). These authors also observed no notable difference in the effect of smoking duration according to *XRCC1* Arg^{399}Gln or Arg^{194}Trp genotype in white women. Our data, with a much larger sample size, support these findings, with a marginally significant inverse association of Arg^{194}Trp with breast cancer risk (multivariate OR, 0.79; 95% CI, 0.60–1.04), and no significant interaction with cigarette smoking. Our study is also consistent with most of the published studies reporting inverse associations between the *XRCC1* ^{194}Trp allele and cancer risk at other sites (29).

Because of its location in the region of the BRCT-I interaction domain of XRCC1 with poly(ADP-ribose) polymerase, the Arg^{399}Gln polymorphism has recently drawn considerable attention. Associations have been reported between the ^{399}Gln allele and higher DNA adduct levels (30–32), higher sister chromatid exchange frequency (30), and radiation-induced G2-phase delay (33). However, in an *in vitro* transfection experiment, the 399 wild-type and variant alleles equally complemented both the single-strand break repair defect and the sensitivity to methyl methanesulfonate in *XRCC1*-deficient EM9 cells, suggesting that the ^{399}Gln variant retained a substantial level of function (34). The marginally significant inverse association of the *XRCC1* ^{194}Trp with breast cancer risk in this study may suggest a protective role for Arg^{194}Trp in the development of breast cancer, potentially by increasing DNA repair capacity. The transition from a positively charged Arg to a hydrophobic Trp within a mammalian-conserved region may alter XRCC1 function. However, no association has been found between Arg^{194}Trp and altered levels of DNA damage biomarkers (32, 33). Our data showed that the two *XRCC1* polymorphisms ^{194}Trp and ^{399}Gln were not on the same common haplotype. Given the potential functional relevance of the two non-synonymous SNPs, we compared the association of carriage of either variant allele to those who were homozygous wild-type at both sites. No notable opposite or synergistic effects of these two variants were observed.

Carotenoids can prevent oxidative DNA damage (35–38). Oxidative DNA damage caused by low plasma levels of carotenoids may lead to increased dependency upon BER activity for maintaining genomic integrity. Thus, variation of BER capacity caused by *XRCC1* genetic polymorphisms may modify the associations of plasma carotene levels with breast cancer risk. Our data showed that high plasma levels of carotene were particularly beneficial among ^{194}Trp carriers. The lowest risk was seen among ^{194}Trp carriers in the highest -or -carotene quartile. Furthermore, the inverse association of the ^{194}Trp allele and breast cancer risk was abolished in the setting of low carotene levels. This suggests that different DNA repair capacity caused by the genetic polymorphism Arg^{194}Trp may be more apparent in the situation of low DNA damage level and overwhelmed by excessive DNA damage. Among carotenoids, higher plasma -carotene or -carotene level was significantly associated with a decreased risk of breast cancer in this nested case-control study.[7] The finding of effect modification of *XRCC1*

genotype on the relation between -carotene and -carotene and breast cancer risk suggests that the protective effect of carotenes on breast cancer risk is at least partially because of prevention of oxidative DNA damage. In addition, in the analysis of the interaction between the *XRCC1* Arg[194]Trp polymorphism and plasma -carotene or -carotene, the multivariate ORs did not change materially and the interactions remained significant after controlling for other plasma antioxidants one at a time or all simultaneously, suggesting that -carotene and -carotene may be the most relevant carotenoids in prevention of oxidative DNA damage in breast tissues.

Oxidative stress caused by cigarette smoking can induce oxidative DNA damage. However, we did not observe evidence of statistically significant interaction between smoking duration and any polymorphism. We lacked power to detect a subtle effect of smoking on oxidative DNA damage in breast tissues repaired by the BER, and smoking-related DNA adducts repaired by nucleotide excision repair may be more relevant to breast carcinogenesis (39). In addition, alcohol consumption is also believed to contribute to oxidative stress. In this nested case-control study, we did not observe the interactions between *XRCC1* genotypes and alcohol consumption on breast cancer risk.

In summary, we provided evidence that women with the *XRCC1* [194]Trp allele and the associated haplotype had a marginally significantly decreased risk of breast cancer. The data also indicate an interaction between the Arg[194]Trp and plasma carotene level on breast cancer risk, which suggests the potential protective role of carotenes in breast carcinogenesis by preventing oxidative DNA damage. Characterization of this potential functional polymorphism and others that may be in linkage disequilibrium on the same haplotype is warranted to explore the biological significance of this variant.

ACKNOWLEDGMENTS

We thank Dr. Hardeep Ranu, Robert O'Brien, Pati Soule, and Alicia Whittington for their laboratory assistance, Rong Chen for her programming support, and Dr. Hannia Campos for plasma antioxidant assays. We also thank the participants in the Nurses' Health Study for their dedication and commitment.

Received 6/2/03; revised 7/22/03; accepted 8/18/03.

Grant support: NIH Grants CA65725, CA87969, and CA49449. The work (H. W. M.) at Lawrence Livermore National Laboratory was performed under the auspices of the United States Department of Energy by the University of California, Lawrence Livermore National Laboratory under Grant W-7405-Eng-48 with additional support by an Interagency Agreement with National Institute of Environmental Health Sciences Grant Y1-ES-8054-05 and National Cancer Institute Grant 1 U-1 CA 83180-03.

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.

Requests for reprints: Jiali Han, Harvard Center for Cancer Prevention, Harvard School of Public Health, Room 105, Building II, 665 Huntington Ave., Boston, MA 02115. E-mail: jhan@hsph.harvard.edu.

[6] The abbreviations used are: BER, base excision repair; OR, odds ratio; CI, confidence interval; XRCC1, X-ray repair cross complementing gene 1; BRCT, BRCA1 COOH terminus; BMI, body mass index; LRT, likelihood ratio test; SNP, single nucleotide polymorphism.

REFERENCES

1. Berwick, M., Matullo, G., and Vineis, P. Studies of DNA repair and human cancer: an update. *In:* S. H. Wilson, W. A. Suk (eds.) Biomarkers of Environmentally Associated Disease: Technologies, Concepts and Perspectives, pp. 84–105. Boca Raton, FL: Lewis Publishers, 2002.
2. Berwick, M., and Vineis, P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. J. Natl. Cancer Inst. (Bethesda), *92:* 874–897, 2000.
3. Roberts, S. A., Spreadborough, A. R., Bulman, B., Barber, J. B., Evans, D. G., and Scott, D. Heritability of cellular radiosensitivity: a marker of low-penetrance predisposition genes in breast cancer? Am. J. Hum. Genet., *65:* 784–794, 1999.
4. Parshad, R., Price, F. M., Bohr, V. A., Cowans, K. H., Zujewski, J. A., and Sanford, K. K. Deficient DNA repair capacity, a predisposing factor in breast cancer. Br. J. Cancer, *74:* 1–5, 1996.
5. Helzlsouer, K. J., Harris, E. L., Parshad, R., Perry, H. R., Price, F. M., and Sanford, K. K. DNA repair proficiency: potential susceptibility factor for breast cancer. J. Natl. Cancer Inst. (Bethesda), *88:* 754–755, 1996.
6. Helzlsouer, K. J., Harris, E. L., Parshad, R., Fogel, S., Bigbee, W. L., and Sanford, K. K. Familial clustering of breast cancer: possible interaction between DNA repair proficiency and radiation exposure in the development of breast cancer. Int. J. Cancer, *64:* 14–17, 1995.
7. Nilsen, H., and Krokan, H. E. Base excision repair in a network of defence and tolerance. Carcinogenesis (Lond.), *22:* 987–998, 2001.
8. Memisoglu, A., and Samson, L. Base excision repair in yeast and mammals. Mutat. Res., *451:* 39–51, 2000.
9. Kubota, Y., Nash, R. A., Klungland, A., Schar, P., Barnes, D. E., and Lindahl, T. Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase and the XRCC1 protein. EMBO J., *15:* 6662–6670, 1996.
10. Marintchev, A., Robertson, A., Dimitriadis, E. K., Prasad, R., Wilson, S. H., and Mullen, G. P. Domain specific interaction in the XRCC1-DNA polymerase complex. Nucleic Acids Res., *28:* 2049–2059, 2000.
11. Thompson, L. H., and West, M. G. XRCC1 keeps DNA from getting stranded. Mutat. Res., *459:* 1–18, 2000.
12. Callebaut, I., and Mornon, J. P. From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. FEBS Lett., *400:* 25–30, 1997.
13. Zhang, X., Morera, S., Bates, P. A., Whitehead, P. C., Coffer, A. I., Hainbucher, K., Nash, R. A., Sternberg, M. J., Lindahl, T., and Freemont, P. S. Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. EMBO J., *17:* 6404–6411, 1998.
14. Whitehouse, C. J., Taylor, R. M., Thistlethwaite, A., Zhang, H., Karimi-Busheri, F., Lasko, D. D., Weinfeld, M., and Caldecott, K. W. XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. Cell, *104:* 107–117, 2001.
15. Vidal, A. E., Boiteux, S., Hickson, I. D., and Radicella, J. P. XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. EMBO J., *20:* 6530–6539, 2001.
16. Shen, M. R., Jones, I. M., and Mohrenweiser, H. Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. Cancer Res., *58:* 604–608, 1998.
17. Mohrenweiser, H. W., Xi, T., Vazquez-Matias, J., and Jones, I. M. Identification of 127 amino acid substitution variants in screening 37 DNA repair genes in humans. Cancer Epidemiol. Biomark. Prev., *11:* 1054–1064, 2002.
18. McCall, M. R., and Frei, B. Can antioxidant vitamins materially reduce oxidative damage in humans? Free Radic. Biol. Med., *26:* 1034–1053, 1999.
19. Frei, B. Reactive oxygen species and antioxidant vitamins: mechanisms of action. Am. J. Med., *97:* 5S–13S; discussion 22S–28S, 1994.
20. Haiman, C. A., Brown, M., Hankinson, S. E., Spiegelman, D., Colditz, G. A., Willett, W. C., Kantoff, P. W., and Hunter, D. J. The androgen receptor CAG repeat polymorphism and risk of breast cancer in the

Nurses' Health Study. Cancer Res., *62:* 1045–1049, 2002.
21. Egan, K. M., Stampfer, M. J., Hunter, D., Hankinson, S., Rosner, B. A., Holmes, M., Willett, W. C., and Colditz, G. A. Active and passive smoking in breast cancer: prospective results from the Nurses' Health Study. Epidemiology, *13:* 138–145, 2002.
22. Collins, F. S., Brooks, L. D., and Chakravarti, A. A DNA polymorphism discovery resource for research on human genetic variation. Genome Res., *8:* 1229–1231, 1998.
23. Qin, Z. S., Niu, T., and Liu, J. S. Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. Am. J. Hum. Genet., *71:* 1242–1247, 2002.
24. El-Sohemy, A., Baylin, A., Kabagambe, E., Ascherio, A., Spiegelman, D., and Campos, H. Individual carotenoid concentrations in adipose tissue and plasma as biomarkers of dietary intake. Am. J. Clin. Nutr., *76:* 172–179, 2002.
25. Ronaghi, M. Pyrosequencing sheds light on DNA sequencing. Genome Res., *11:* 3–11, 2001.
26. Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., and Altshuler, D. The structure of haplotype blocks in the human genome. Science (Wash. DC), *296:* 2225–2259, 2002.
27. Stephens, J. C., Schneider, J. A., Tanguay, D. A., Choi, J., Acharya, T., Stanley, S. E., Jiang, R., Messer, C. J., Chew, A., Han, J. H., Duan, J., Carr, J. L., Lee, M. S., Koshy, B., Kumar, A. M., Zhang, G., Newell, W. R., Windemuth, A., Xu, C., Kalbfleisch, T. S., Shaner, S. L., Arnold, K., Schulz, V., Drysdale, C. M., Nandabalan, K., Judson,
R. S., Ruano, G., and Vovis, G. F. Haplotype variation and linkage disequilibrium in 313 human genes. Science (Wash. DC), *293:* 489–493, 2001.
28. Duell, E. J., Millikan, R. C., Pittman, G. S., Winkel, S., Lunn, R. M., Tse, C. K., Eaton, A., Mohrenweiser, H. W., Newman, B., and Bell, D. A. Polymorphisms in the DNA repair gene XRCC1 and breast cancer. Cancer Epidemiol. Biomark. Prev., *10:* 217–222, 2001.
29. Goode, E. L., Ulrich, C. M., and Potter, J. D. Polymorphisms in DNA repair genes and associations with cancer risk. Cancer Epidemiol. Biomark. Prev., *11:* 1513–1530, 2002.
30. Duell, E. J., Wiencke, J. K., Cheng, T. J., Varkonyi, A., Zuo, Z. F., Ashok, T. D., Mark, E. J., Wain, J. C., Christiani, D. C., and Kelsey, K. T. Polymorphisms in the DNA repair genes XRCC1 and ERCC2 and biomarkers of DNA damage in human blood mononuclear cells. Carcinogenesis (Lond.), *21:* 965–971, 2000.
31. Matullo, G., Palli, D., Peluso, M., Guarrera, S., Carturan, S., Celentano, E., Krogh, V., Munnia, A., Tumino, R., Polidoro, S., Piazza, A., and Vineis, P. XRCC1, XRCC3, XPD gene polymorphisms, smoking and (32)P-DNA adducts in a sample of healthy subjects. Carcinogenesis (Lond.), *22:* 1437–1445, 2001.
32. Lunn, R. M., Langlois, R. G., Hsieh, L. L., Thompson, C. L., and Bell, D. A. XRCC1 polymorphisms: effects on aflatoxin B1-DNA adducts and glycophorin A variant frequency. Cancer Res., *59:* 2557–2261, 1999.
33. Hu, J. J., Smith, T. R., Miller, M. S., Mohrenweiser, H. W., Golden, A., and Case, L. D. Amino acid substitution variants of *APE1* and *XRCC1* genes associated with ionizing radiation sensitivity. Carcinogenesis (Lond.), *22:* 917–922, 2001.
34. Taylor, R. M., Thistlethwaite, A., and Caldecott, K. W. Central role for the XRCC1 BRCT I domain in mammalian DNA single-strand break repair. Mol. Cell. Biol., *22:* 2556–2263, 2002.
35. Duthie, S. J., Ma, A., Ross, M. A., and Collins, A. R. Antioxidant supplementation decreases oxidative DNA damage in human lymphocytes. Cancer Res., *56:* 1291– 1295, 1996.
36. Pool-Zobel, B. L., Bub, A., Muller, H., Wollowski, I., and Rechkemmer, G. Consumption of vegetables reduces genetic damage in humans: first results of a human intervention trial with carotenoid-rich foods. Carcinogenesis (Lond.), *18:* 1847–1850, 1997.
37. Collins, A. R., Olmedilla, B., Southon, S., Granado, F., and Duthie, S. J. Serum carotenoids and oxidative DNA damage in human lymphocytes. Carcinogenesis (Lond.), *19:* 2159–2162, 1998.
38. Torbergsen, A. C., and Collins, A. R. Recovery of human lymphocytes from oxidative DNA damage; the apparent enhancement of DNA repair by carotenoids is probably simply an antioxidant effect. Eur. J. Nutr., *39:* 80–85, 2000.

39. Terry, P. D., and Rohan, T. E. Cigarette smoking and the risk of breast cancer in women: a review of the literature. Cancer Epidemiol. Biomark. Prev., *11:* 953–971, 2002.

Table 1 *XRCC1 common SNPs and inferred haplotypes*[a]

SNP ID	Nucleotide position	Genotype	Amino acid position	Amino acid substitution	Allele frequency	Common haplotypes	
						SNP ID 1–17	Haplotype frequency (%)
1	18776	T/C	72	Val/Ala	0.028		
2	24737	C/A	Intronic		0.280		
3	26304	C/T	194	Arg/Trp	0.115	0 0 **00** 0000000 **0** 0000 **0**	25.7
4	26602	C/T	Intronic		0.103	0 1 **00** 0000001 **1** 0000 **0**	18.7
5	26651	A/G	206	Pro/Pro	0.268	0 0 **10** 0000000 **0** 0000 **0**	10.5
6	26759	C/T	Intronic		0.012	0 0 **01** 0000000 **0** 0000 **0**	9.2
7	26760	T/A	Intronic		0.012	0 0 **00** 1000100 **0** 0100 **1**	22.7
8	27772	C/T	Intronic		0.019		
9	27826	T/C	Intronic		0.254		
10	27853	C/G	Intronic		0.018		
11	27980	A/G	Intronic		0.259		
12	28152	G/A	399	Arg/Gln	0.238		
13	32734	G/A	Intronic		0.014		
14	33543	G/A	Intronic		0.253		
15	33582	G/A	Intronic		0.016		
16	35914	A/C	576	Tyr/Ser	0.011		
17	36189	G/A	632	Gln/Gln	0.170		

[a] Seventeen SNPs in *XRCC1* gene with 1% allele frequency were selected based on the resequencing data of the subset of 90 samples from the NIH DNA Polymorphism Discovery Resource. On the basis of the selected 17 SNPs, five common haplotypes were inferred by the Partition-Ligation EM algorithm. These five common haplotypes accounted for 87% of the alleles of the mixed population group. Genotyping four bolded SNPs (3, 4, 12, 17) reconstructed the five common haplotypes. “0” stands for more common allele; “1” stands for less common allele.

Table 2 *Frequencies of inferred haplotypes in cases and controls*[a]

Haplotypes Allele Frequency						
C26304T (Arg194Trp)	C26602T	G28152A (Arg399Gln)	G36189A (Gln632Gln)	Cases (*n*,%) *n* 1896	Controls (*n*,%) *n* 2587	*P*
0	0	0	0	222 (11.7)	299 (11.6)	0.91*b*
0	0	0	1	766 (40.4)	1063 (41.1)	0.66*b*
0	0	1	0	689 (36.3)	913 (35.3)	0.49*b*
0	1	0	0	106 (5.6)	120 (4.6)	0.17*b*
1	0	0	0	100 (5.3)	171 (6.6)	0.07*b*
0	0	1	1	8 (0.4)	17 (0.7)	0.40*b*
1	0	0	1	5 (0.3)	4 (0.2)	0.51*c*

[a] Haplotype frequencies from our observed genotypes were estimated by the Partition-Ligation EM algorithm. "0" stands for more common allele; "1" stands for less common allele. *P* (^{2}, *df* 6) 0.41. Three additional rare haplotypes were predicted in five alleles in controls only; one haplotype was predicted in one allele, two in two.

[b] *P* is based on ^{2} test.

[c] *P* is based on Fisher's exact test.

Table 3 *Polymorphisms in XRCC1 and breast cancer risk*

	Wild type	Heterozygote		Homozygous variant
C26304T (Arg194Trp)*a*				
Cases/controls*b*	890/1190	105/176		3/3
Matched OR*c*	1.00		0.81 (0.62–1.05)	
Multivariate OR*d*	1.00		0.79 (0.60–1.04)	
C26602T*a*				
Cases/controls*b*	874/1229	101/137		6/2
Matched OR*c*	1.00		1.09 (0.83–1.43)	
Multivariate OR*d*	1.00		1.09 (0.81–1.45)	
G28152A (Arg399Gln)				
Cases/controls*b*	391/545	460/616		135/176
Matched OR*c*	1.00	1.06 (0.88–1.27)		1.06 (0.81–1.39)
Multivariate OR*d*	1.00	1.07 (0.88–1.30)		1.03 (0.77–1.37)
G36189A (Gln632Gln)				
Cases/controls*b*	340/444	480/678		160/232
Matched OR*c*	1.00	0.90 (0.75–1.09)		0.89 (0.69–1.13)
Multivariate OR*d*	1.00	0.91 (0.75–1.11)		0.90 (0.69–1.16)

[a] Heterozygote and homozygous variant are combined to compare with wild type.

[b] The number of participants does not sum to total women because of missing data on genotype.

[c] Conditional logistic regression adjusted for the matching variables: age; menopausal status; postmenopausal hormone use; date of blood draw; time of blood draw; and fastingstatus.

[d] Conditional logistic regression adjusted for the matching variables, BMI at age 18 years (continuous), weight gain since age 18 years (5, 5–19.9, 20 kg), age at menarche (12, 12, 13, 13 years), age at menopause (45, 45 to 50, 50 to 55, or 55 years), parity/age at first birth (nulliparous, 1–2 children/age at first birth 24 years, 1–2 children/age at first birth 24 years, 3 children/age at first birth 24 years, 3 children/age at first birth 24 years), first-degree family history of breast cancer (yes/no), history of benign breast disease (yes/no), and duration of postmenopausal hormone use (never use, past use, current use 5 years, current use 5 years).

Table 4 *Breast cancer risk by plasma carotene levels and XRCC1 Arg*^{194}*Trp*

	1 (Lowest)	2	3	4 (Highest)	*P* for trend
Plasma -carotene level					
194Trp noncarriers					
Case/controls*a*	229/217	233/214	246/203	173/210	
Multivariate OR1*b*	1.00	1.03 (0.79–1.35)	1.15 (0.89–1.50)	0.78 (0.59–1.03)	0.07
Multivariate OR2*c*	1.00	0.97 (0.74–1.28)	1.12 (0.84–1.48)	0.68 (0.50–0.92)	0.008
194Trp carriers					
Case/controls*a*	30/21	35/30	21/35	18/32	
Multivariate OR1*b*	1.36 (0.76–2.46)	1.12 (0.67–1.90)	0.57 (0.32–1.01)	0.53 (0.29–0.98)	0.01
Multivariate OR2*c*	1.30 (0.70–2.40)	1.19 (0.69–2.05)	0.53 (0.30–0.97)	0.40 (0.21–0.77)	0.03
P 0.03 by the nominal test for interaction and 0.02 by the ordinal test in the multivariate model 2.					
Plasma -carotene level					
194Trp noncarriers					
Case/controls*a*	215/217	244/215	240/214	182/198	
Multivariate OR1*b*	1.00	1.14 (0.87–1.48)	1.14 (0.87–1.48)	0.93 (0.70–1.23)	0.38
Multivariate OR2*c*	1.00	1.15 (0.87–1.52)	1.11 (0.83–1.47)	0.84 (0.62–1.14)	0.07
194Trp carriers					
Case/controls*a*	32/22	28/28	29/28	15/40	
Multivariate OR1*b*	1.48 (0.83–2.64)	1.03 (0.59–1.80)	1.04 (0.60–1.81)	0.37 (0.20–0.70)	0.001
Multivariate OR2*c*	1.38 (0.76–2.51)	1.13 (0.63–2.01)	1.01 (0.57–1.81)	0.32 (0.16–0.61)	0.003
P 0.02 by the nominal test for interaction and 0.003 by the ordinal test in the multivariate model 2.					

[a] The number of participants does not sum to total women because of missing data on genotype.

[b] Unconditional logistic regression adjusted for the matching variables: age; menopausal status; postmenopausal hormone use; date of blood draw; time of blood draw; and fasting status.

[c] Unconditional logistic regression adjusted for the matching variables, BMI at age 18 years (continuous), weight gain since age 18 years (5, 5–19.9, 20 kg), age at menarche (12, 12, 13, 13 years), age at menopause (45, 45 to 50, 50 to 55, or 55 years), parity/age at first birth (nulliparous, 1–2 children/age at first birth 24 years, 1–2 children/age at first birth 24 years, 3 children/age at first birth 24 years, 3 children/age at first birth 24 years), first-degree family history of breast cancer (yes/no), history of benign breast disease (yes/no), alcohol intake (0, 0to 5, 5to 15, 15 to 30, 30 g/day), during smoking (0, 0to 20, 20 to 35, 35 years), and duration of postmenopausal hormone use (never use, past use, current use 5 years, current use 5 years).

Table 5 *Breast cancer risk by duration of smoking and XRCC1 Arg*^{194}*Trp*

	Duration of smoking				
	Never	0–20	20–35	35	*P* for trend
194Trp noncarriers					
Case/controls*a*	405/584	190/232	165/208	130/166	
Multivariate OR1*b*	1.00	1.13 (0.90–1.43)	1.14 (0.89–1.46)	1.25 (0.96–1.64)	0.08
Multivariate OR2*c*	1.00	1.17 (0.92–1.49)	1.18 (0.92 1.53)	1.31 (0.98–1.74)	0.06
194Trp carriers					
Case/controls*a*	51/83	23/45	20/29	14/22	
Multivariate OR1*b*	0.88 (0.60–1.27)	0.73 (0.43–1.24)	1.13 (0.63–2.04)	1.09 (0.54–2.18)	0.35
Multivariate OR2*c*	0.88 (0.60–1.29)	0.74 (0.43–1.27)	1.11 (0.60–2.03)	0.99 (0.48–2.00)	0.40

P 0.76 by the nominal test for interaction and 0.90 by the ordinal test in the multivariate model 2.

[a] The number of participants does not sum to total women because of missing data on genotype.

[b] Unconditional logistic regression adjusted for the matching variables: age; menopausal status; postmenopausal hormone use; date of blood draw; time of blood draw; and fasting status.

[c] Unconditional logistic regression adjusted for the matching variables, BMI at age 18 years (continuous), weight gain since age 18 years (5, 5–19.9, 20 kg), age at menarche (12, 12, 13, 13 years), age at menopause (45, 45 to 50, 50 to 55, or 55 years), parity/age at first birth (nulliparous, 1–2 children/age at first birth 24 years, 1–2 children/age at first birth 24 years, 3 children/age at first birth 24 years, 3 children/age at first birth 24 years), first-degree family history of breast cancer (yes/no), history of benign breast disease (yes/no), alcohol intake (0, 0to 5, 5to 15, 15 to 30, 30 g/day), and duration of postmenopausal hormone use (never use, past use, current use 5 years, current use 5 years).