

DNA repair gene polymorphisms and tobacco smoking in the risk for colorectal adenomas

Ying Gao*, Richard B. Hayes, Wen-Yi Huang,
Neil E. Caporaso, Laurie Burdette, Meredith Yeager,
Stephen J. Chanock and Sonja I. Berndt

Division of Cancer Epidemiology and Genetics, National Cancer Institute,
National Institutes of Health, Bethesda, MD 20892-7236, USA

*To whom correspondence should be addressed. Genetic Epidemiology
Branch, Division of Cancer Epidemiology and Genetics, National Cancer
Institute, National Institutes of Health, 6120 Executive Boulevard, Building
EPS/Room 7110, Bethesda, MD 20892-7236, USA. Tel: +1 301 496 9249;
Fax: +1 301 402 4489;
Email: gaoying@mail.nih.gov

DNA damage is thought to play a critical role in the development of colorectal adenoma. Variation in DNA repair genes may alter their capacity to correct endogenous and exogenous DNA damage. We explored the association between common single-nucleotide polymorphisms (SNPs) in DNA repair genes and adenoma risk with a case-control study nested in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. A total of 1338 left sided, advanced colorectal adenoma cases and 1503 matched controls free of left-sided polyps were included in the study. Using DNA extracted from blood, 3144 tag SNPs in 149 DNA repair genes were successfully genotyped. Among Caucasians, 30 SNPs were associated with adenoma risk at $P < 0.01$, with four SNPs remaining significant after gene-based adjustment for multiple testing. The most significant finding was for a non-synonymous SNP (rs9350) in Exonuclease-1 (*EXO1*) [odds ratio (OR) = 1.30, 95% confidence interval (CI) = 1.11–1.51, $P = 0.001$], which was predicted to be damaging using bioinformatics methods. However, the association was limited to smokers with a strong risk for current smokers (OR = 2.15, 95% CI = 1.27–3.65) and an intermediate risk for former smokers (OR = 1.45, 95% CI = 1.14–1.82) and no association for never smokers (OR = 0.98, 95% CI = 0.76–1.25) ($P_{\text{interaction}} = 0.002$). Among the top findings, an SNP (rs17503908) in ataxia telangiectasia mutated (*ATM*) was inversely related to adenoma risk (OR = 0.75, 95% CI = 0.63–0.91). The association was restricted to never smokers (OR = 0.55, 95% CI = 0.40–0.76) with no increased risk observed among smokers (OR = 0.89, 95% CI = 0.70–1.13) ($P_{\text{interaction}} = 0.006$). This large comprehensive study, which evaluated all presently known DNA repair genes, suggests that polymorphisms in *EXO1* and *ATM* may be associated with risk for advanced colorectal adenoma with the associations modified by tobacco-smoking status.

Introduction

Colorectal cancer is the third most common cancer in the USA for both men and women (1). Epidemiological studies have shown that non-steroidal anti-inflammatory drugs, exogenous hormones and select dietary factors are risk factors of colorectal neoplasia (2,3). Genetic factors also contribute to risk with the heritability of colorectal cancer estimated to be 35% from a large twin study (4). Genome-wide association studies have identified at least 14 loci associated with the risk of colorectal cancer (5,6); however, additional loci are predicted to exist (7). Although there have been no genome-wide association studies exclusively of colorectal adenoma, a known precursor to

Abbreviations: ATM, ataxia telangiectasia mutated; CI, confidence interval; EXO1, Exonuclease-1; HNPCC, hereditary non-polyposis colorectal cancer; MMR, mismatch repair; OR, odds ratio; PLCO, Prostate, Lung, Colorectal and Ovarian; SNP, single-nucleotide polymorphism.

colorectal cancer, studying genetic susceptibility to colorectal adenoma may give insight into the etiology of colorectal cancer.

Smoking has been consistently associated with an increased risk of colorectal adenoma (8,9). A recent meta-analysis found the risk estimate of adenoma for ever smokers to be 1.82 [95% confidence interval (CI) = 1.65–2.00] (8). Generally, the risk was stronger for current [odds ratio (OR) = 2.14, 95% CI = 1.86–2.46] as opposed to former smokers (OR = 1.47, 95% CI = 1.29–1.67) (8). Carcinogens generated from tobacco smoking interact with DNA to form DNA adducts, which can interfere with cell replication and if not repaired correctly, can cause somatic mutations leading eventually to cancer. Emerging evidence has also shown that tobacco smoking may interact with genetic factors, predisposing certain individuals to greater polyp susceptibility (10). However, the studies focused on limited number of candidate genes to date have only provided limited evidence for gene-environmental interactions (10).

Damage caused by smoking and other environmental exposures activates several different DNA repair pathways, including base excision repair, mismatch repair (MMR), nucleotide excision repair and double-strand break repair pathways (11). Rare germ line mutations in MMR have been shown to lead to hereditary non-polyposis colorectal cancer (HNPCC) (12) and mutations in the base excision repair gene, *MUTYH*, have been linked to a familial polyposis syndrome (13). Some significant associations have also been reported between common DNA repair gene polymorphisms and colorectal neoplasia risk (14–17). In particular, a common variant in the *MLH1* gene region has been linked to the risk of colorectal cancers with microsatellite instability (18–21). However, most studies examined only small sets of single-nucleotide polymorphisms (SNPs) (14–17), and the associations remain to be confirmed. As a complex disease, multiple genetic variants with minor to moderate effects probably contribute to colorectal adenoma development, which makes scanning a large number of genes simultaneously rather than a small set of individual SNPs attractive.

Studies of genetic susceptibility to colorectal adenoma may give insight into the etiology of colorectal carcinogenesis. To examine the relationship between polymorphisms in DNA repair genes and colorectal adenoma risk as well as potential effect modification by tobacco smoking, we conducted a nested case-control study within the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. As most DNA repair genes have not been well characterized with regard to disease susceptibility, we decided to undertake a comprehensive approach including all DNA repair genes known to our knowledge to date (11). For each gene, we selected tag SNPs to thoroughly capture the common genetic variation in the region.

Methods

The PLCO Cancer Screening Trial

The PLCO is a clinical trial, designed to assess the efficacy of screening tests to reduce death from cancers of prostate, lung, colon and rectum and ovary. As described previously (22,23), 154 938 cancer-free men and women aged 55–74 were recruited from 10 sites in the USA between 1993 and 2001. Participants were randomly assigned to the control group or the intervention group (screening arm), where they underwent a 60 cm flexible sigmoidoscopy examination at study entry (T0) and year 3 (T3) or year 5 (T5) of the study. Those found to have a suspicious lesion were referred to their personal physician for subsequent diagnostic follow-up. Among the 64 658 men and women who underwent sigmoidoscopy at T0 in the PLCO Cancer Screening Trial, 8.8% were found to have adenoma (24). Cases of colorectal adenoma were pathologically verified according to medical records. Information on demographics, personal and family medical history and lifestyle factors (e.g. smoking and dietary intake) were collected by standard questionnaire at baseline. This trial was approved by the institutional review boards of the 10 screening centers and the National

Cancer Institute in Bethesda, MD, USA, and all participants provided informed consent.

Study population

A nested case-control study was conducted among the PLCO participants randomized to the screening arm, who consented to participate in etiologic studies of cancer and related diseases, completed a risk factor questionnaire, provided a blood specimen and had no previous history of inflammatory bowel disease, colorectal polyps, Gardner's syndrome, familial polyposis or cancer other than basal or squamous cell skin cancer. For this study, cases were participants found to have advanced colorectal adenoma (≥ 1 cm in size, containing villous/tubulovillous characteristics, or had severe dysplasia) of the distal colon or rectum at the T0 exam. Carcinoma *in situ* was classified as severe dysplasia. Controls were participants, who had successful sigmoidoscopy at T0 and were negative for polyps in the distal colon and rectum. Controls were frequency matched to cases on self-reported ethnicity (non-Hispanic Caucasian, non-Hispanic Black, Asian, Hispanic, Pacific Islander, American Indian or Alaskan Native or Unknown), gender and for a subset on age (55–59, 60–64, 65–69 and 70–74 years). Over 90% of the subjects were non-Hispanic Caucasians.

Genotyping

A total of 3338 tag SNPs from 149 genes (supplementary Table 1 is available at *Carcinogenesis* Online) involved in DNA repair pathways (11) were selected to comprehensively interrogate genetic variations across the candidate genes. Tag SNPs were selected for each gene, including the region 20 kb upstream and 10 kb downstream of the gene, using the CEU, JPT, CHB and YRI HapMap populations and the Carlson method (25) as implemented in Tagzilla with a r^2 threshold of 0.8 and minor allele frequency $\geq 5\%$. SNPs with known or putative functional significance (i.e. non-synonymous, promoter, intron-exon splice sites) were also included whenever possible. The SNPs were genotyped on a custom iSelect panel utilizing Illumina's GoldenGate platform.

Whole blood or buffy coat DNA was extracted with QIAamp DNA Blood Midi or Maxi Kits. For this study, sufficient DNA was available from 1342 cases and 1507 controls for genotyping. For quality control purposes, replicate samples from 195 individuals (~7% of the population) were interspersed randomly within the plates. Genotyping was conducted at the National Cancer Institute Core Genotyping Facility, NIH. A total of 1338 cases and 1503 controls were successfully genotyped with over 90% of the genotypes for each subject having valid calls, and the overall concordance rate was >99% for replicated samples. After excluding the SNPs with call rate <90%, minor allele frequency <1%, or Hardy-Weinberg Equilibrium P -value $< 1 \times 10^{-6}$ among Caucasian controls, 3144 SNPs of 3401 selected (92%) remained for analysis. For each gene, the percentage of tagSNPs passing our quality control criteria varied from 67 to 100% with an average of 93%.

Statistical analysis

The initial analyses were conducted using Plink, a whole genome association analysis toolset (26). Logistic regression was used to estimate the OR and 95% CI for the association between each SNP and colorectal adenoma risk assuming a log-additive model for the genotype, adjusting for age (55–59, 60–64, 65–69, 70–74 years), gender and ethnicity (non-Hispanic Caucasian, non-Hispanic Black and other). Another set of analysis was conducted among non-Hispanic Caucasians only.

For SNPs with main association $P < 0.01$, we tested if the SNP associations differed by smoking status (ever versus never) using the Breslow-Day test of homogeneity. The SNPs with evidence of heterogeneity ($P < 0.05$ between ever versus never smoker) among Caucasians were further tested for interaction with smoking status (never, former and current smoking) using a likelihood ratio test. The P -value for trend was generated by treating the smoking status (never = 0, former = 1 and current = 2) as a continuous variable in the interaction model, and the P -value for interaction was generated by treating the smoking status as categorical variable in the interaction model. Similarly, we also conducted analyses to examine the effect modification by smoking duration (<24 years, ≥ 24 years) and years since quitting smoking (<20 years, ≥ 20 years). These additional analyses were conducted with SAS 9.1.

Pairwise linkage disequilibrium measures (D' and r^2) were inferred from the Caucasian controls using the program Haploview (27). Haplotypes among Caucasians were estimated using an expectation-maximization algorithm for SNPs within the gene Exonuclease-1 (*EXO1*), which carried the SNP with lowest P -value in the current study, and risks for individual haplotypes were calculated assuming a log-additive model and using the generalized linear model implemented in R Haplostats package, adjusted for age and gender (28). For consistency with the SNP results, we used the haplotype containing the T allele at rs9350 as the reference haplotype.

To evaluate the potential for false-positive findings due to multiple testing, we adjusted the P values using a Bonferroni correction for the total number of (a) tag SNPs for each individual gene (gene based) as well as (b) all the SNPs tested in the current analysis, using the R multtest package.

Results

A total of 1338 colorectal adenoma cases and 1503 frequency-matched controls were included in the current analyses (Table I). Over 90% of the study subjects were Caucasian and <10% were African-American or other ethnicity. Compared with controls, cases were more probably to be current smokers, have a family history of colorectal cancer and be less educated.

Of the 3144 SNPs analyzed, 129 SNPs (supplementary Table 2 is available at *Carcinogenesis* Online) were associated with adenoma risk among all subjects and 127 SNPs with risk among Caucasians only at $P < 0.05$. The SNPs associated with colorectal adenoma risk at $P \leq 0.01$ level among Caucasians and are shown in Table II. After adjusting for multiple testing for all the SNPs tested in the analysis, none of these 30 SNPs remained statistically significant. However, six SNPs remained associated with adenoma risk among Caucasians at $P \leq 0.05$ level after a gene-based multiple testing correction: *EXO1* rs9350, *FANCC* rs400727, *ERCC1* rs10412761, *DCLRE1A* rs2301180, *RAD54B* rs3762053 and *POLE* rs11614717. Further adjustment for smoking status did not substantially alter the results in Table II (results not shown). The most statistically significant SNP associated with risk was *EXO1* rs9350 with heterozygotes displaying a 1.95-fold risk (95% CI = 1.05–3.62) and CC homozygotes displaying a 2.39-fold risk (95% CI = 1.30–4.38) comparing with TT homozygotes.

Given the importance of smoking as a risk factor for adenoma, we examined the heterogeneity in risk by smoking status (ever versus

Table I. Demographic characteristics of cases and controls, PLCO Cancer Screening Trial baseline exam 1993–2001

| | Controls (%) $N = 1503$ | Cases (%) $N = 1338$ |
|--|-------------------------|----------------------|
| Race | | |
| White | 1370 (91) | 1241 (93) |
| Black | 66 (4) | 50 (4) |
| Other | 67 (5) | 47 (3) |
| Gender | | |
| Male | 950 (63) | 853 (64) |
| Female | 553 (37) | 485 (36) |
| Age | | |
| 55–59 | 498 (33) | 390 (29) |
| 60–64 | 460 (31) | 419 (31) |
| 65–69 | 344 (23) | 333 (25) |
| 70–75 | 201 (13) | 196 (15) |
| Smoking status | | |
| Never | 650 (43) | 451 (34) |
| Former cigarette smoker | 679 (45) | 627 (47) |
| Current cigarette smoker | 105 (7) | 197 (15) |
| Cigar or pipe smoker | 69 (5) | 62 (5) |
| Family history of colorectal cancer | | |
| Yes | 141 (9) | 162 (12) |
| No | 1362 (91) | 1176 (88) |
| Body mass index (BMI) | | |
| <25 | 451 (30) | 369 (28) |
| 25–29.9 | 680 (46) | 594 (45) |
| ≥ 30 | 354 (24) | 364 (27) |
| Education | | |
| <12 years | 439 (29) | 470 (35) |
| Post high school | 489 (33) | 479 (36) |
| College graduate or postgraduate | 574 (38) | 389 (29) |
| Regular non-steroidal anti-inflammatory drug use | | |
| Yes | 935 (62) | 776 (58) |
| No | 565 (38) | 559 (42) |

Table II. OR of colorectal adenoma for polymorphisms in DNA repair genes with $P < 0.01$, PLCO Cancer Screening Trial baseline exam 1993–2001

| SNP | Gene | Risk allele | RAF ^c | Caucasian subjects ^a | | | All subjects ^b | | |
|----------------|----------------|-------------|------------------|---------------------------------|-----------|----------|---------------------------|-----------|----------|
| | | | | OR | 95% CI | <i>P</i> | OR | 95% CI | <i>P</i> |
| rs9350 T→C | <i>EXO1</i> | C | 0.82 | 1.30 | 1.11–1.51 | 0.001 | 1.24 | 1.07–1.44 | 0.003 |
| rs400727 T→G | <i>FANCC</i> | G | 0.91 | 1.44 | 1.15–1.80 | 0.002 | 1.40 | 1.14–1.72 | 0.001 |
| rs3762053 G→T | <i>RAD54B</i> | G | 0.16 | 1.25 | 1.09–1.44 | 0.002 | 1.23 | 1.07–1.41 | 0.004 |
| rs11614717 T→G | <i>POLE</i> | T | 0.14 | 1.27 | 1.09–1.49 | 0.002 | 1.24 | 1.07–1.44 | 0.005 |
| rs4703564 T→C | <i>XRCC4</i> | C | 0.89 | 1.42 | 1.13–1.78 | 0.002 | 1.31 | 1.09–1.59 | 0.005 |
| rs242448 A→T | <i>FANCC</i> | T | 0.92 | 1.42 | 1.13–1.78 | 0.002 | 1.41 | 1.14–1.75 | 0.002 |
| rs4658535 A→G | <i>EXO1</i> | G | 0.79 | 1.26 | 1.08–1.45 | 0.002 | 1.25 | 1.09–1.42 | 0.002 |
| rs1699499 C→T | <i>FANCC</i> | T | 0.97 | 1.70 | 1.20–2.40 | 0.003 | 1.67 | 1.19–2.35 | 0.003 |
| rs17503908 G→T | <i>ATM</i> | G | 0.08 | 1.34 | 1.11–1.62 | 0.003 | 1.33 | 1.10–1.60 | 0.003 |
| rs10412761 G→A | <i>ERCC1</i> | A | 0.57 | 1.19 | 1.06–1.33 | 0.003 | 1.19 | 1.07–1.33 | 0.001 |
| rs3885676 C→T | <i>XRCC4</i> | T | 0.87 | 1.36 | 1.11–1.66 | 0.003 | 1.31 | 1.10–1.56 | 0.003 |
| rs4465523 A→G | <i>APEX1</i> | A | 0.33 | 1.18 | 1.05–1.32 | 0.006 | 1.20 | 1.08–1.34 | 0.001 |
| rs17277375 T→C | <i>APEX1</i> | T | 0.32 | 1.18 | 1.05–1.33 | 0.006 | 1.19 | 1.06–1.33 | 0.003 |
| rs2301180 C→T | <i>DCLRE1A</i> | T | 0.71 | 1.18 | 1.05–1.34 | 0.006 | 1.16 | 1.03–1.30 | 0.01 |
| rs6864054 G→A | <i>XRCC4</i> | A | 0.94 | 1.44 | 1.11–1.87 | 0.006 | 1.33 | 1.05–1.70 | 0.02 |
| rs5744990 A→G | <i>POLE</i> | A | 0.13 | 1.24 | 1.06–1.45 | 0.006 | 1.23 | 1.06–1.43 | 0.007 |
| rs6573333 T→G | <i>MNAT1</i> | G | 0.98 | 2.36 | 1.26–4.40 | 0.007 | 1.18 | 0.80–1.74 | 0.39 |
| rs3102854 G→A | <i>RAD54B</i> | G | 0.24 | 1.19 | 1.05–1.35 | 0.007 | 1.18 | 1.04–1.33 | 0.009 |
| rs7829886 A→T | <i>RAD54B</i> | T | 0.73 | 1.19 | 1.05–1.35 | 0.008 | 1.16 | 1.03–1.31 | 0.02 |
| rs12050102 T→G | <i>APEX1</i> | T | 0.33 | 1.17 | 1.04–1.32 | 0.008 | 1.19 | 1.06–1.33 | 0.002 |
| rs4150628 G→A | <i>GTF2H1</i> | G | 0.44 | 1.16 | 1.04–1.23 | 0.008 | 1.14 | 1.03–1.27 | 0.02 |
| rs228606 T→G | <i>ATM</i> | T | 0.39 | 1.16 | 1.04–1.30 | 0.008 | 1.15 | 1.04–1.28 | 0.009 |
| rs3212986 A→C | <i>ERCC1</i> | C | 0.74 | 1.19 | 1.04–1.35 | 0.008 | 1.19 | 1.05–1.34 | 0.006 |
| rs10838192 C→T | <i>ALKBH3</i> | C | 0.21 | 1.19 | 1.05–1.36 | 0.009 | 1.19 | 1.04–1.35 | 0.008 |
| rs828918 G→C | <i>XRCC5</i> | G | 0.13 | 1.24 | 1.06–1.45 | 0.009 | 1.19 | 1.02–1.38 | 0.03 |
| rs3131382 T→C | <i>MSH5</i> | C | 0.93 | 1.36 | 1.08–1.71 | 0.009 | 1.37 | 1.09–1.72 | 0.007 |
| rs1150793 G→A | <i>MSH5</i> | G | 0.06 | 1.34 | 1.08–1.67 | 0.009 | 1.26 | 1.02–1.55 | 0.03 |
| rs707915 A→T | <i>MSH5</i> | A | 0.06 | 1.34 | 1.08–1.67 | 0.009 | 1.26 | 1.02–1.55 | 0.03 |
| rs11615 G→A | <i>ERCC1</i> | A | 0.58 | 1.16 | 1.04–1.31 | 0.009 | 1.15 | 1.02–1.28 | 0.02 |
| rs5744903 T→C | <i>POLE</i> | T | 0.06 | 1.32 | 1.07–1.62 | 0.01 | 1.28 | 1.04–1.57 | 0.02 |

^aOR per risk allele assuming a log-additive model, adjusted for age and sex.^bOR per risk allele assuming a log-additive model, adjusted for age, sex and race.^cRAF, risk allele frequency among Caucasian controls.**Table III.** OR of colorectal adenoma for select DNA repair SNPs stratified by smoking status among Caucasians^a, PLCO Cancer Screening Trial baseline exam 1993–2001

| Gene | Risk allele | Never | | Former | | Current | | <i>P</i> trend ^b | <i>P</i> interaction ^c | |
|------------|-------------|-------|------------------|--------|------------------|---------|------------------|-----------------------------|-----------------------------------|-------|
| | | OR | <i>P</i> | OR | <i>P</i> | OR | <i>P</i> | | | |
| rs9350 | <i>EXO1</i> | C | 0.98 (0.76–1.25) | 0.86 | 1.45 (1.14–1.82) | 0.002 | 2.15 (1.27–3.65) | 0.004 | 0.002 | 0.006 |
| rs4658535 | <i>EXO1</i> | G | 0.98 (0.77–1.23) | 0.84 | 1.34 (1.08–1.68) | 0.008 | 2.13 (1.30–3.49) | 0.003 | 0.002 | 0.006 |
| rs17503908 | <i>ATM</i> | T | 0.55 (0.40–0.76) | 0.0003 | 0.84 (0.64–1.11) | 0.22 | 1.25 (0.71–2.22) | 0.44 | 0.008 | 0.02 |

^aSNPs selected here were those with significant main effect (adjusted for age, sex) ($P < 0.01$) and gene–smoking (ever versus never smoking) interaction ($P < 0.05$).^b*P*-value for trend was calculated assuming smoking status (never = 0, former = 1 and current = 2) as continuous variable in the interaction model.^c*P*-value for interaction was calculated assuming smoking status (never = 0, former = 1 and current = 2) as categorical variable in the interaction model.

never). Of the SNPs with a $P < 0.01$ for their main association, three SNPs displayed significant heterogeneity in risk by smoking status ($P < 0.05$) and were selected for further analyses stratified by cigarette smoking status (never, former and current) (Table III). The risk of colorectal adenoma at *EXO1* rs9350 was significantly modified by smoking status ($P_{\text{interaction}} = 0.006$) with a 2-fold increased risk among current smokers (OR = 2.15; 95% CI: 1.27–3.65), a modest increased risk among former smokers (OR = 1.45; 95% CI: 1.14–1.82), and no association among never smokers (OR = 0.98; 95% CI: 0.76–1.25) ($P_{\text{trend}} = 0.002$). A stronger increased risk was also observed among individuals with longer smoking duration (≥ 24 years: OR = 1.59; 95% CI: 1.20–2.10) compared with shorter duration (< 24 years: OR = 1.48; 95% CI: 1.09–1.99) ($P_{\text{interaction}} = 0.02$), as well as individuals who quit more recently (< 20 years: OR = 1.64; 95% CI: 1.24–2.18) ($P_{\text{interaction}} = 0.02$) compared with those who quit a longer time ago (≥ 20 years: OR = 1.42; 95% CI: 1.0–1.99). Similarly, another SNP in *EXO1*, rs4658535, also showed a monotonically increasing pattern of risk from

never, former, to current smokers ($P_{\text{trend}} = 0.002$). These two SNPs in *EXO1* were strongly correlated ($r^2 = 0.82$), making it difficult to differentiate the associations of one from the other statistically. When both SNPs were put in the same model, neither of them remained significantly associated with adenoma risk ($P > 0.05$ for both) due to the high correlation. Combining former and current smokers, the C allele at rs9350 and G allele at rs4658535 were associated with increased risk among ever smokers with ORs of 1.54 (95% CI = 1.26–1.89) and 1.46 (95% CI = 1.21–1.77) for the two SNPs, respectively.

The risk of adenoma at rs17503908 in ataxia telangiectasia mutated (*ATM*) was also modified by smoking status ($P_{\text{interaction}} = 0.02$) with the T allele displaying a decreased risk for adenoma only among never smokers (OR = 0.55, 95% CI = 0.40–0.76) (Table III) but not among ever smokers (OR = 0.89, 95% CI = 0.70–1.13). Notably, rs17503908 showed an intermediate risk for adenoma also among former smokers (OR = 0.84, 95% CI = 0.64–1.11) compared with never smokers (OR = 0.55, 95% CI = 0.40–0.76) and current

smokers (OR = 1.25, 95% CI = 0.71–2.22) ($P_{\text{trend}} = 0.008$). Similar patterns were also observed when the results were stratified by smoking duration and time since quitting (data not shown).

Given that the associations of two SNPs in *EXO1* appeared to be modified by smoking status, we explored this region in greater detail. A total of 25 SNPs were genotyped in *EXO1*. Seven SNPs (rs9350, rs1635488, rs4408133, rs1635484, rs4150018, rs4150027 and rs4658535), including the two SNPs significantly associated with risk, were in strong linkage disequilibrium as measured by D' (Figure 1) and mildly to strongly correlated (r^2 range: 0.23–0.84). The five most frequent haplotypes comprised of these seven SNPs were analyzed in association with adenoma risk by smoking status (never and ever smoking, but not never, former and current smoking status due to limited power) (supplementary Table 3 is available at *Carcinogenesis* Online). Compared with the haplotype containing the T allele at rs9350 (rs9350-rs1635488-rs4408133-rs1635484-rs4150018-rs4150027-rs4658535: TCGCGTA), three haplotypes, each containing the risk alleles at rs9350 and rs4658535, were significantly associated with an increased risk of adenoma with risk estimates ranging from 1.23 to 1.30. The associations were stronger among smokers (40–60% increased risk) compared with never smokers for which none of the haplotypes were associated with risk. The test for the haplotype–smoking interaction was marginally significant ($P = 0.06$).

Discussion

DNA repair has long been implicated in colorectal cancer with the discovery that germ line mutations in MMR genes lead to HNPCC (29) and mutations in the base excision repair gene, *MUTYH*, lead to a familial polyposis syndrome. However, the etiological role of common genetic variation in DNA repair genes in colorectal adenoma and cancer has not been comprehensively studied in the context of epide-

miological studies. Although some common SNPs in DNA repair genes have been reported to be associated with colorectal cancer and/or adenoma (18–21,30,31), with the exception of the *MHL1*-93G > A variant with microsatellite instable tumors (18–21), most associations have not been replicated (30,31). Furthermore, data on effect modifications by important environmental factors are sparse.

In our study, >3000 SNPs from 153 DNA repair genes were evaluated simultaneously among 2841 study subjects, which is the largest and most comprehensive study for colorectal adenoma risk focusing on DNA repair genes to date. Among the SNPs associated with risk, we found that genetic polymorphisms in *EXO1* and *ATM* significantly modified the effect of cigarette smoking on risk, predisposing smokers to greater adenoma susceptibility. When stratified by genotype, smoking was only significantly associated with increased adenoma risk among individuals homozygous for the risk allele at *EXO1* rs9350 (OR = 1.80, 95% CI = 1.47–2.19 for ever smokers with the CC genotype) and *ATM* rs17503908 (OR = 1.69, 95% CI = 1.40–2.03 for ever smokers with the GG genotype).

EXO1, located at chromosome 1q42–q43, encodes a protein with 5' → 3' and 3' → 5' double-stranded DNA exonuclease activity. It also exhibits some endonuclease activity correcting 5'-overhanging flap structures. *EXO1* is involved in DNA MMR, recombination, replication and telomere stabilization (32). *EXO1*-mutant cells showed increased microsatellite instability and incomplete MMR capability (33). Mice with *EXO1* knockout were found to have lower survival rates and higher mutation rates as well as higher susceptibility to lymphomas (33). *EXO1* has been implicated in hereditary HNPCC due to its role in DNA MMR; however, studies investigating rare germ line variants in *EXO1* have not shown consistent findings as reviewed by Liberti *et al.* (34).

We observed an increased colorectal adenoma risk among individuals carrying a C allele at rs9350 and a G allele at rs4658535 in *EXO1*.

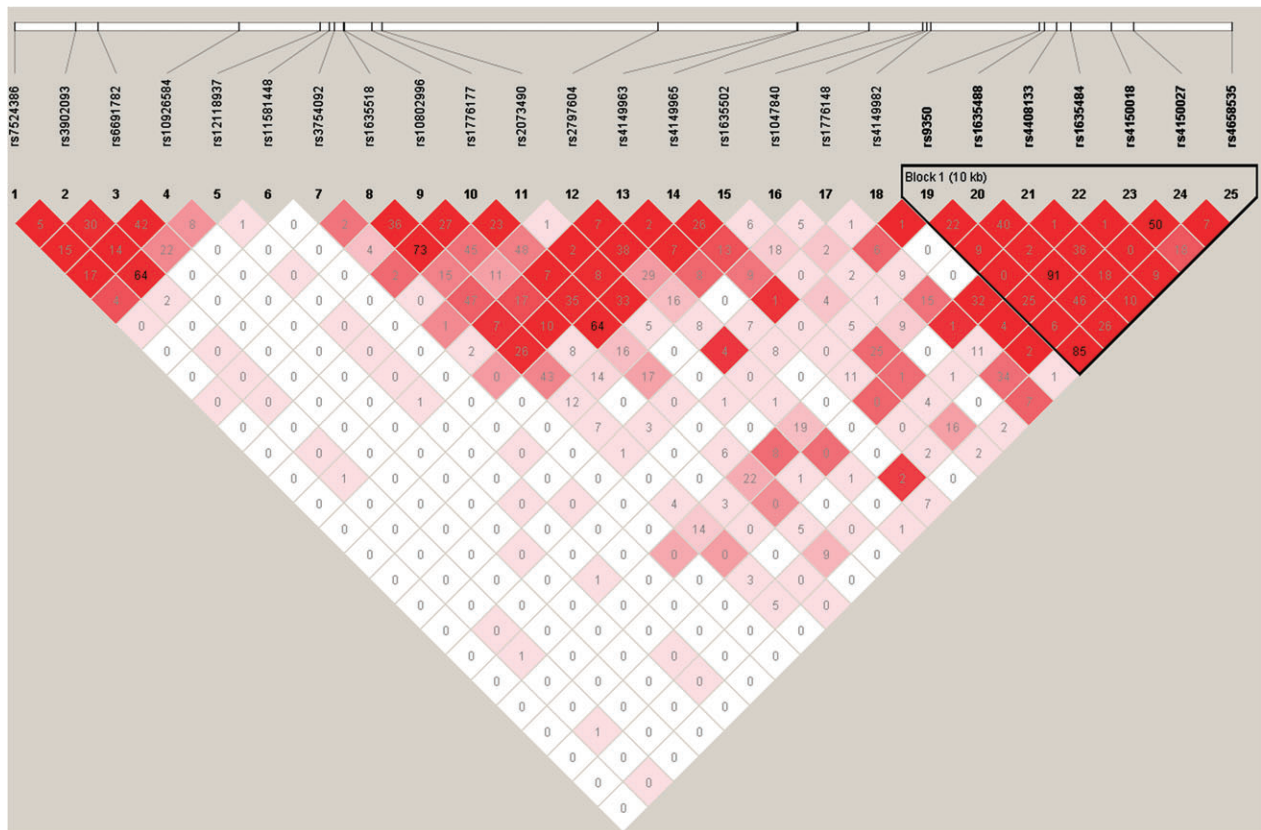


Fig. 1. Linkage disequilibrium plot of *EXO1* among Caucasian controls from the PLCO Cancer Screening Trial baseline exam 1993–2001. The colors represent the extent of pairwise linkage disequilibrium as measured by D' with red indicating $D' = 1.0$ and shades of pink to white indicating $D' < 1.0$. The numeric values are the pairwise r^2 values.

The SNPs were highly correlated ($r^2 = 0.82$), making it impossible to differentiate the associations of each statistically. Using the PolyPhen database, we found that rs9350 was predicted to be 'probably damaging' (position-specific independent counts score difference = 2.17) with the C to T substitution causing a non-synonymous amino acid change by replacing proline with leucine (35), suggesting that rs9350 may be a causal variant. The substitution was also predicted to be 'deleterious' using the SIFT database. Several studies have examined the association between common polymorphisms in *EXO1* and the risk of lung, oral, brain and colorectal cancer (36–42); however, data are limited and inconclusive due to the small sample size and differences in the SNPs genotyped in the studies (36–42). No studies have examined the association between this polymorphism and adenoma. Consistent with our findings, two case–control studies of colorectal cancer found a decreased cancer risk for individuals carrying T allele at *EXO1* rs9350 compared with C allele (40,43). No association was observed with rs9350 in two studies of lung cancer (37,38), one study of oral cancer (39) and one study of breast cancer (44), suggesting that the association of C allele at rs9350 may be organ/tissue specific. In addition, the previous studies in other cancers (37–39) were relatively small ($N \leq 680$ cases each) and conducted in Asian populations, where differences in environmental exposures may modify the association of rs9350 and cancer risk compared with Caucasians.

Smoking is an important risk factor for adenoma with current smokers having an 1.8-fold increased risk (95% CI = 1.5–2.1) in the full PLCO cohort (45). We observed a stronger association between adenoma risk and rs9350 in *EXO1* among smokers compared with non-smokers and hypothesize that the C allele at *EXO1* rs9350 may increase risk among smokers by reducing the protein's capacity or efficiency to repair the damage caused by smoking exposure. Tsai *et al.* also reported an increased risk of oral cancer among smokers who carried the A allele at rs1047840 ($r^2 = 0.18$ with rs9350 in our study) in *EXO1* but not among non-smokers (39). The haplotype results further confirmed the strong association between the *EXO1* region encompassing rs9350 and adenoma risk.

We observed a significant decreased risk among never smokers for the T allele at rs17503908 in *ATM*. *ATM*, located at chromosome 11q22.3, encodes a cell cycle checkpoint kinase which regulates many downstream proteins, including the tumor suppressor proteins p53 and BRCA1, the checkpoint kinase CHK2, checkpoint proteins RAD17 and RAD9 and the DNA repair protein NBS1. *ATM* is thought to be a master controller of the cell cycle checkpoint-signaling pathways and functions to repair DNA damage and maintain genome stability. Persons with ataxia telangiectasia, an autosomal recessive disease caused by rare missense or truncating mutations in *ATM*, have an increased sensitivity to ionizing radiation and an increased risk of cancer (46). Heterozygous carriers of these rare *ATM* mutations have an increased risk of several cancers including colorectal cancer (47).

Several lines of evidence have suggested the etiological role for *ATM* in colorectal carcinogenesis (47,48). Polymorphisms at rs1800056 and rs1800057 in *ATM* have been associated with colorectal cancer risk (49) and although the results were not replicated in a follow-up study (50), rs1801516 has been associated with disease penetrance among HNPCC carriers (51). In our study, the T allele at *ATM* rs1801516 was also marginally associated with adenoma risk ($P = 0.011$ for Caucasians and $P = 0.016$ for all subjects), but no association was observed with rs1800056 ($P = 0.22$). To date, no study has reported an association between *ATM* rs17503908 and colorectal adenoma, which is located in an intronic region of *ATM*. We observed an inverse association with the T allele at rs17503908. Differences in sensitivity to DNA damage or higher expression levels of *ATM* could reduce risk for neoplastic transformation or subsequent proliferation by activating p53 (52). In stratified analyses, this inverse association was restricted to never smokers. It is possible that smokers do not benefit from carrying this allele due to an antagonistic effect between the SNP and smoking exposure. An *in vitro* study observed that smoking exposure activated *ATM* in human pulmonary adenocarcinoma cells through phosphorylation in a dose-dependent manner (53). Similarly, ben-

zo[a]pyrene diol epoxide, a polycyclic aromatic hydrocarbon found in tobacco smoke, has been shown to bind to *ATM* (54) and induce *ATM* expression in esophageal cancer cell lines (55), suggesting that *ATM* plays an active role in responding to tobacco smoke exposure. We speculate that the kinase encoded by the *ATM* may be saturated by smoking exposure, which may prevent its protective effect.

Our study has several advantages and limitations. First, it is the largest study to date to investigate a broad range of DNA repair gene polymorphisms for colorectal adenoma risk. Although still somewhat underpowered to examine gene–environment interactions for the SNPs of moderate association, our findings provide promising leads for replication in future pooled analyses. Future analyses exploring these interactions with regard to colorectal cancer may lead to additional insight into colorectal neoplasm progression. Our study included only advanced adenoma cases and so the results may not be generalizable to non-advanced adenomatous polyps; however, advanced adenomas are more to progress to colorectal cancer and therefore clinically more relevant. Moreover, our study only included left-sided adenomas in the distal colon and rectum. Thus, the results of our study may not be generalizable to adenomas observed in the proximal colon, which may be more probably to occur as the result of deficiencies in MMR. In addition, our study population came from a cancer prevention screening trial in which participants were generally more probably to be Caucasian, more educated, less probably to smoke and more physically active than the general population (56). Thus, our results may not be broadly generalizable to the entire population or to other ethnicities. However, since our case–control study was nested within a randomized population-based colorectal cancer screening trial, this reduces the potential for selection bias often inherited in clinic-based case–control studies of adenoma, where persons may undergo endoscopy for reasons other than routine screening, such as gastrointestinal symptoms, blood in their stools, diagnostic follow-up or because they have a family history of colorectal cancer.

In summary, in this large comprehensive study of DNA repair gene polymorphisms and colorectal adenoma risk, we found that an SNP in *EXO1* predicted to deleteriously alter function was associated with increased adenoma risk. The association was restricted to ever smokers and stronger in current smokers than former smokers. Although additional studies are needed to confirm our findings, this intriguing result suggests that genetic variation in *EXO1* may modify susceptibility to colorectal adenoma, particularly among smokers.

Supplementary material

Supplementary Tables 1–3 can be found at <http://carcin.oxfordjournals.org/>

Funding

Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health (NIH).

Acknowledgements

The authors thank Drs Christine Berg and Philip Prorok, Division of Cancer Prevention, NCI, the screening center investigators and staff of the PLCO Cancer Screening Trial, Mr Thomas Riley and staff at Information Management Services and Ms Barbara O'Brien and staff at Westat for their contributions to the PLCO Cancer Screening Trial. Finally, we acknowledge the study participants for donating their time and making this study possible.

Author contributions: Y.G., S.B., R.H. and W.-Y.H. designed the study. Y.G. and S.B. also analyzed data and wrote the manuscript. L.B., M.Y. and S.C. were instrumental in the genotyping for this project. All authors read, gave comments and approved the final version of the manuscript. All authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of Interest Statement: None declared.

References

- Jemal,A. *et al.* (2009) Cancer statistics, 2009. *CA Cancer J. Clin.*, **59**, 225–249.
- Terry,M.B. *et al.* (2002) Risk factors for advanced colorectal adenomas: a pooled analysis. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 622–629.
- Huxley,R.R. *et al.* (2009) The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. *Int. J. Cancer*, **125**, 171–180.
- Lichtenstein,P. *et al.* (2005) Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.*, **343**, 78–85.
- Tenesa,A. *et al.* (2009) New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat. Rev. Genet.*, **10**, 353–358.
- Houlston,R.S. *et al.* (2010) Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.*, **42**, 973–977.
- Park,J.H. *et al.* (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.*, **42**, 570–575.
- Botteri,E. *et al.* (2008) Cigarette smoking and adenomatous polyps: a meta-analysis. *Gastroenterology*, **134**, 388–395.
- Giovannucci,E. *et al.* (1996) Tobacco, colorectal cancer, and adenomas: a review of the evidence. *J. Natl Cancer Inst.*, **88**, 1717–1730.
- Raimondi,S. *et al.* (2009) Gene-smoking interaction on colorectal adenoma and cancer risk: review and meta-analysis. *Mutat. Res.*, **670**, 6–14.
- Wood,R.D. *et al.* (2005) Human DNA repair genes, 2005. *Mutat. Res.*, **577**, 275–283.
- Kinzler,K.W. *et al.* (1996) Lessons from hereditary colorectal cancer. *Cell*, **87**, 159–170.
- Nielsen,M. *et al.* (2009) Analysis of MUTYH genotypes and colorectal phenotypes in patients With MUTYH-associated polyposis. *Gastroenterology*, **136**, 471–476.
- Berndt,S.I. *et al.* (2007) Genetic variation in base excision repair genes and the prevalence of advanced colorectal adenoma. *Cancer Res.*, **67**, 1395–1404.
- Stern,M.C. *et al.* (2006) XRCC1, XRCC3, and XPD polymorphisms as modifiers of the effect of smoking and alcohol on colorectal adenoma risk. *Cancer Epidemiol. Biomarkers Prev.*, **15**, 2384–2390.
- Skjelbred,C.F. *et al.* (2006) Polymorphisms of the XRCC1, XRCC3 and XPD genes and risk of colorectal adenoma and carcinoma, in a Norwegian cohort: a case control study. *BMC Cancer*, **6**, 67.
- Bigler,J. *et al.* (2005) DNA repair polymorphisms and risk of colorectal adenomatous or hyperplastic polyps. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 2501–2508.
- Mrkonjic,M. *et al.* (2010) Specific variants in the MLH1 gene region may drive DNA methylation, loss of protein expression, and MSI-H colorectal cancer. *PLoS One*, **5**, e13314.
- Raptis,S. *et al.* (2007) MLH1 -93G>A promoter polymorphism and the risk of microsatellite-unstable colorectal cancer. *J. Natl Cancer Inst.*, **99**, 463–474.
- Samowitz,W.S. *et al.* (2008) The MLH1 -93 G>A promoter polymorphism and genetic and epigenetic alterations in colon cancer. *Genes Chromosomes Cancer*, **47**, 835–844.
- Campbell,P.T. *et al.* (2009) Mismatch repair polymorphisms and risk of colon cancer, tumour microsatellite instability and interactions with lifestyle factors. *Gut*, **58**, 661–667.
- Hayes,R.B. *et al.* (2005) Methods for etiologic and early marker investigations in the PLCO trial. *Mutat. Res.*, **592**, 147–154.
- Prorok,P.C. *et al.* (2000) Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control. Clin. Trials*, **21**, 273S–309S.
- Weissfeld,J.L. *et al.* (2005) Flexible sigmoidoscopy in the PLCO cancer screening trial: results from the baseline screening examination of a randomized trial. *J. Natl Cancer Inst.*, **97**, 989–997.
- Carlson,C.S. *et al.* (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.*, **74**, 106–120.
- Purcell,S. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
- Barrett,J.C. *et al.* (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
- Schaid,D.J. *et al.* (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am. J. Hum. Genet.*, **70**, 425–434.
- Lynch,H.T. *et al.* (2009) Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin. Genet.*, **76**, 1–18.
- de la Chapelle,A. (2004) Genetic predisposition to colorectal cancer. *Nat. Rev. Cancer*, **4**, 769–780.
- Foulkes,W.D. (2008) Inherited susceptibility to common cancers. *N. Engl. J. Med.*, **359**, 2143–2153.
- Tran,P.T. *et al.* (2004) EXO1-A multi-tasking eukaryotic nuclease. *DNA Repair (Amst.)*, **3**, 1549–1559.
- Wei,K. *et al.* (2003) Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility. *Genes Dev.*, **17**, 603–614.
- Liberti,S.E. *et al.* (2004) Is hEXO1 a cancer predisposing gene? *Mol. Cancer Res.*, **2**, 427–432.
- Ramensky,V. *et al.* (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.*, **30**, 3894–3900.
- Chang,J.S. *et al.* (2008) Pathway analysis of single-nucleotide polymorphisms potentially associated with glioblastoma multiforme susceptibility using random forests. *Cancer Epidemiol. Biomarkers Prev.*, **17**, 1368–1373.
- Hsu,N.Y. *et al.* (2009) Lung cancer susceptibility and genetic polymorphisms of Exo1 gene in Taiwan. *Anticancer Res.*, **29**, 725–730.
- Jin,G. *et al.* (2008) Potentially functional polymorphisms of EXO1 and risk of lung cancer in a Chinese population: a case-control analysis. *Lung Cancer*, **60**, 340–346.
- Tsai,M.H. *et al.* (2009) Interaction of Exo1 genotypes and smoking habit in oral cancer in Taiwan. *Oral Oncol.*, **45**, e90–e94.
- Yamamoto,H. *et al.* (2005) Single nucleotide polymorphisms in the EXO1 gene and risk of colorectal cancer in a Japanese population. *Carcinogenesis*, **26**, 411–416.
- Yoshiya,G. *et al.* (2008) Influence of cancer-related gene polymorphisms on clinicopathological features in colorectal cancer. *J. Gastroenterol. Hepatol.*, **23**, 948–953.
- Zienolddiny,S. *et al.* (2006) Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. *Carcinogenesis*, **27**, 560–567.
- Haghighi,M.M. *et al.* (2010) Impact of EXO1 polymorphism in susceptibility to colorectal cancer. *Genet. Test. Mol. Biomarkers*, **14**, 649–652.
- Wang,H.C. *et al.* (2009) Association of genetic polymorphisms of EXO1 gene with risk of breast cancer in Taiwan. *Anticancer Res.*, **29**, 3897–3901.
- Ji,B.T. *et al.* (2006) Tobacco smoking and colorectal hyperplastic and adenomatous polyps. *Cancer Epidemiol. Biomarkers Prev.*, **15**, 897–901.
- Morrell,D. *et al.* (1986) Mortality and cancer incidence in 263 patients with ataxia-telangiectasia. *J. Natl Cancer Inst.*, **77**, 89–92.
- Thompson,D. *et al.* (2005) Cancer risks and mortality in heterozygous ATM mutation carriers. *J. Natl Cancer Inst.*, **97**, 813–822.
- Kweekel,D.M. *et al.* (2009) Explorative study to identify novel candidate genes related to oxaliplatin efficacy and toxicity using a DNA repair array. *Br. J. Cancer.*, **101**, 357–362.
- Webb,E.L. *et al.* (2006) Search for low penetrance alleles for colorectal cancer through a scan of 1467 non-synonymous SNPs in 2575 cases and 2707 controls with validation by kin-cohort analysis of 14 704 first-degree relatives. *Hum. Mol. Genet.*, **15**, 3263–3271.
- Jones,J.S. *et al.* (2005) ATM polymorphism and hereditary nonpolyposis colorectal cancer (HNPCC) age of onset (United States). *Cancer Causes Control*, **16**, 749–753.
- Maillet,P. *et al.* (2000) A polymorphism in the ATM gene modulates the penetrance of hereditary non-polyposis colorectal cancer. *Int. J. Cancer*, **88**, 928–931.
- Concannon,P. *et al.* (2008) Variants in the ATM gene associated with a reduced risk of contralateral breast cancer. *Cancer Res.*, **68**, 6486–6491.
- Jorgensen,E.D. *et al.* (2010) DNA damage response induced by exposure of human lung adenocarcinoma cells to smoke from tobacco- and nicotine-free cigarettes. *Cell Cycle*, **9**, 2170–2176.
- Liang,Z. *et al.* (2003) Identification of benzo(a)pyrene diol epoxide-binding DNA fragments using DNA immunoprecipitation technique. *Cancer Res.*, **63**, 1470–1474.
- Jiang,Y. *et al.* (2007) Ataxia-telangiectasia mutated expression is associated with tobacco smoke exposure in esophageal cancer tissues and benzo(a)pyrene diol epoxide in cell lines. *Int. J. Cancer*, **120**, 91–95.
- Pinsky,P.F. *et al.* (2007) Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. *Am. J. Epidemiol.*, **165**, 874–881.

Received October 27, 2010; revised March 17, 2011; accepted March 25, 2011