A population-based study of DNA repair gene variants in relation to non-melanoma skin cancer as a marker of a cancer-prone phenotype

Ingo Ruczinski1,2, Timothy J.Jorgensen1,2, Yin Yao Shugart1, Yvette Berthier Schaad3,4, Bailey Kessing5, Judith Hoffman-Bolton4,6, Kathy J.Helzlsouer7, W.H.Linda Kao8, Lee Wheless9, Lesley Francis3, Rhoda M.Alan1, Paul T.Strickland4,5, Michael W.Smith1 and Anthony J.Alberg9

1Department of Biostatistics, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA 2Department of Radiation Medicine, Georgetown University School of Medicine, Washington, DC, USA 3Division of Intramural Research Program, National Institute of Mental Health, Bethesda, MD, USA 4Department of Epidemiology, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA 5Laboratory of Genomic Diversity, SAIC-Frederick, NCI-Frederick, Frederick, MD 6George W.Comstock Center for Public Health Research and Prevention, Washington County, MD, USA 7Mercy Medical Center, Baltimore, MD, USA 8Holings Cancer Center and Division of Epidemiology and Biostatistics, Department of Medicine, Medical University of South Carolina, Charleston, SC, USA 9Department of Dermatology, Boston University School of Medicine, Boston, MA, USA 10Department of Environmental Health Sciences, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA and 11Genetics and Genomics Group, Advanced Technology Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD, USA

For unknown reasons, non-melanoma skin cancer (NMSC) is associated with increased risk of other malignancies. Focusing solely on DNA repair or DNA repair-related genes, this study tested the hypothesis that DNA repair gene variants contribute to the increased cancer risk associated with a personal history of NMSC. From the parent CLUE II cohort study, established in 1989 in Washington County, MD, the study consisted of a cancer-free control group (n = 2296) compared with three mutually exclusive groups of cancer cases ascertained through 2007: (i) Other (non-NMSC) cancer only (n = 2349); (ii) NMSC only (n = 694) and (iii) NMSC plus other cancer (n = 577). The frequency of minor alleles in 759 DNA repair gene single nucleotide polymorphisms (SNPs) was compared in these four groups. Comparing those with both NMSC and other cancer versus those with no cancer, 10 SNPs had allelic trend P-values < 0.01. The two top-ranked SNPs were both within the thymine DNA glycosylase gene (TDG). One was a non-synonymous coding SNP (rs2888805) [per allele odds ratio (OR) 1.40, 95% confidence interval (CI) 1.16–1.70; P-value = 0.0006] and the other was an intronic SNP in high linkage disequilibrium with rs2888805 (rs4135150). None of the associations had a P-value < 0.05 × 10−5, the threshold for statistical significance after correcting for multiple comparisons. The results pinpoint DNA repair genes most likely to contribute to the NMSC cancer-prone phenotype. A promising lead is genetic variants in TDG, thymine DNA glycosylase gene; XP, xeroderma pigmentosum.

Abbreviations: BCC, basal cell carcinoma; LD, linkage disequilibrium; MAF, minor allele frequency; NER, nucleotide excision repair; NMSC, non-melanoma skin cancer; SNP, single nucleotide polymorphism; SCC, squamous cell carcinoma; TDG, thymine DNA glycosylase gene; XP, xeroderma pigmentosum.

†To whom correspondence should be addressed. Tel: +843 876 2349; Fax: +843 876 2344; Email: alberg@musc.edu

Introduction

Non-melanoma skin cancer (NMSC) is by far the most common human cancer (1). NMSC is usually treated by local excision and is rarely fatal, but a personal history of NMSC is associated with increased risk of other malignancies (2). In a meta-analysis of prospective studies with individual-level data, a prior NMSC diagnosis was associated with a 50% greater risk of developing another type of cancer (2). NMSC was associated with a broad spectrum of malignancies (2), suggesting that NMSC may be a marker of a cancer-prone phenotype. NMSC cases comprise two major histologic types, basal cell carcinoma (BCC) and squamous cell carcinoma (SCC), and the overall increased risk of other malignancies associated with NMSC has been consistently observed for both BCC and SCC. However, it is not known why a personal history of NMSC is associated with increased risk of non-cutaneous malignancies, and there has been no hypothesis-driven research on this topic. The fact that a personal history of NMSC affects overall cancer risk across a broad spectrum of malignancies potentially makes NMSC a cross-cutting marker of overall cancer risk that could potentially provide valuable clues to human carcinogenesis.

Major DNA repair defects cause cancer-prone phenotypes. For example, mutated DNA repair genes cause cancer-prone syndromes that involve multiple cancers, such as xeroderma pigmentosum (XP) (3,4) and Bloom’s syndrome (5,6). XP is an autosomal recessive condition caused by rare, high-penetrant mutations in nucleotide excision repair (NER) genes. XP directly demonstrates that defective DNA repair can be a potent cause of genetic predisposition to both NMSC and other malignancies. The DNA repair deficiency in XP is characterized by extreme sensitivity to ultraviolet radiation, causing >2000-fold increased skin cancer rates before age 20 years (4). XP patients also have a marked excess of internal malignancies (4). The fact that rare mutations in DNA repair genes cause major cancer susceptibility in XP patients suggests that even common, low-penetrant DNA repair gene variants may jointly affect susceptibility to NMSC and other malignancies in the general population (7,8). However, except for rare, autosomal recessive cancer syndromes caused by inherited mutations in DNA repair genes, the contribution of common allelic variants of DNA repair genes to cancer risk in the general population remains incompletely characterized (e.g. ref. 7).

DNA repair is accomplished through multiple interacting biochemical pathways entailing overlapping functions and cross-talk; collectively, these comprise >140 genes (Table I, column 1). All of these pathways have been implicated in carcinogenesis. The NER pathway is emphasized for NMSC because NER removes cyclobutane pyrimidine dimers and 6-4 photoproducts, carcinogenic DNA lesions unique to ultraviolet radiation exposure. Ultraviolet radiation-induced DNA lesions are not restricted to pyrimidine dimers but also entail oxidative lesions (8,9), demonstrating that even protection against skin cancer involves multiple DNA repair pathways. Thus, in the present study, all known human DNA repair genes were considered candidates for association with the NMSC-associated cancer-prone phenotype.

Consistent with other cohort studies, we previously reported that a prior NMSC diagnosis was associated with significantly elevated risk of other malignancies (10). Building on this finding, the present study focused solely on DNA repair and DNA repair-related genes and employed the same cohort to test the novel hypothesis that germ-line polymorphisms in DNA repair genes contribute to the cancer-prone phenotype characterized by the diagnosis of both NMSC and another type of cancer in the same individual.
NA repair gene variants and NMSC cancer-prone phenotype

Materials and Methods
This prospective, population-based study was embedded within the CLUE II (named after the campaign slogan of ‘Give us a clue to cancer and heart disease’) cohort. CLUE II is a community-based cohort study established in Washington County, MD in 1989 (11). In this study, the independent variables were single nucleotide polymorphisms (SNPs) in DNA repair genes. The dependent variable was the occurrence of cancer classified according to the joint outcome of NMSC and other cancer, with a cancer-free group to compare with those with cancer other than NMSC only, NMSC only, and both NMSC plus other cancer. The study of all non-NMSC cancers combined is consistent with the prior 10 year study that the excess risk of other malignancies observed in those with a personal history of NMSC is not confined to a few specific malignancies but rather seems to apply to an overall heightened cancer risk.

Study population
The CLUE II cohort was established when baseline data were collected in May through November 1989 from volunteers who were mostly residents of Washington County, MD. The campaign was designed to collect blood samples from as many adult residents as possible in the Washington County and surrounding (30 mile radius) area. Brief medical histories and blood pressures were taken, and 20 ml of blood was drawn into heparinized vacuum tubes. Specimens were refrigerated at once and processed within 24 h. Buffy coats were placed in storage at −70 °C. During the baseline data collection, participants completed a questionnaire that included information on age, race, sex, cigarette smoking, height, weight and years of schooling. This study was approved by the Institutional Review Boards of the Johns Hopkins University and Medical University of South Carolina.

From the entire cohort of 30,726 participants, 6589 were selected for inclusion in the present study. This included all those with a confirmed cancer diagnosis as of 30 September 2007 (plus 96 cases added 1 April 2008) and a cancer-free comparison group that was a 10% age-stratified random sample of adult CLUE II participants, plus 250 controls added for a lung cancer substudy. Once selected, NMSC or other cancer diagnoses that occurred by 31 December 2010 contributed to the categorization of study endpoints. Of the 6589 subjects selected for inclusion, genotyping could not be attempted for 312, and 263 with ≥5% genotyping failures were excluded. Among the remaining 6014 subjects, genetic heterogeneity was assessed using principal components. Comparison with the three HapMap2 populations identified 98 (1.6%) subjects of non-European ancestry, who were excluded due to concerns about population stratification and differential risks of NMSC. To account for any residual ancestral differences, the first three principal components were adjusted for in all analyses. The final study population of 5916 was classified into four categories: (i) cancer-free group (n = 2296) for comparison to those with a pathologically confirmed diagnosis of (ii) any cancer other than NMSC (‘other cancer only,’ n = 2349); (iii) NMSC with no other cancer diagnosis (‘NMSC only,’ n = 694) and (iv) NMSC plus another type of cancer (‘NMSC plus other cancer,’ n = 577).

Cancer diagnoses were ascertained through linkage to the Washington County Cancer Registry, augmented (except for NMSC) by linking to the Maryland Cancer Registry. Pathologically confirmed NMSC diagnoses were ascertained by the Washington County Cancer Registry, including NMSC cases diagnosed both before and after study baseline in 1989.

SNP Selection
The SNP selection algorithm was previously described (12). In brief, 118 human DNA-repair genes and DNA-repair-related genes (13) were assigned to one of the five central DNA repair pathways (Table I, rows 1–5) or to one of seven ancillary DNA repair-related pathways (Table I, rows 6–12) based on putative activities, functions, sequence homology or physical associations with other DNA repair proteins. These pathways were prioritized in the order listed in Table I based on evidence of association with NMSC. All known non-synonymous coding SNPs in genes from these 12 pathways were selected for genotyping, regardless of minor allele frequency (MAF). Tagging SNPs were selected using a general MAF cut-off of 0.05. This was relaxed to 0.01 for top priority pathway genes in the SNP selection process. Despite the inclusion of rare variants in the SNP selection, for the data analyses all SNPs with MAF < 0.05 were excluded, due to the high prevalence of monomorphic and quasi-monomorphic SNPs and the lack of precision associated with the low-MAF SNPs that were not monomorphic. The CEPH population (European ancestry) data from the HapMap (CEU) were used to estimate SNP frequencies and linkage disequilibrium, as the CLUE II cohort comprised largely northern European ancestry. R² was set to >0.80, although some SNPs could be in high linkage disequilibrium (LD) due to the approach of selecting functional and tagging SNPs. SNPs were excluded if they had a genotyping platform design score <0.60. HapMap data were accessed using Tagger™ software (14), with 1337 DNA repair gene SNPs identified that met these criteria.

Genomic DNA extraction and genotyping
Buffy coats with no previous DNA extraction were stored at −70°C from collection until DNA extraction. Genomic DNA was extracted from buffy coats using standard phenol/chloroform extraction procedures, followed by ethanol precipitation to uniform DNA concentrations (50 μg/mL) for genotyping. When previously extracted DNA was available, that DNA was resuspended in low-salt buffer to uniformly 50 μg/mL for genotyping. Genotyping was attempted for 1337 selected DNA repair SNPs using Illumina GoldenGate® arrays customized by the manufacturer. Among these, SNPs were excluded for the following reasons: 86 (6.4%) genotyping failure in ≥5% samples, 467 (34.9%) with MAF ≤ 0.05 (247 of these were monomorphic or quasi-monomorphic), and 25 (1.9%) deviated from Hardy–Weinberg equilibrium (HWE) at P < 0.001. This left 759 SNPs for analyses. For quality control, the genotyping results were compared for n = 24 SNPs genotyped in previous studies using different geno-

### Table I. DNA repair genes studied and the number of SNPs per gene according to DNA repair or DNA repair-related pathway.

<table>
<thead>
<tr>
<th>DNA repair pathway</th>
<th>Genes (no. of SNPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Nucleotide excision repair: 28 genes, 188 SNPs</td>
<td>CCNH (2), CDK7 (3), DBH1 (2), DDB2 (6), ERCC1 (4), ERCC2 (8), ERCC3 (7), ERCC4 (5), ERCC5 (8), ERCC6 (8), ERCC8 (5), GT2H1 (5), GT2H2 (1), GT2H4 (5), LIG1 (8), MMS19 (4), MNIAT1 (8), POLD1 (5), POLE (10), POLH (2), RAD23A (1), RAD23B (16), RPA1 (12), RPA2 (6), RPA3 (34), XAB2 (3), XPA (2), XPC (8)</td>
</tr>
<tr>
<td>2. Base excision repair: 14 genes, 56 SNPs</td>
<td>APEX1 (11), LIG3 (3), MBD4 (5), MPPG (2), MUTYH (4), NEIL1 (1), NEIL2 (14), NTHLI (2), OGG1 (3), PARP (2), SMUG1 (3), TDG (5), Ung (5), XRCC3 (9)</td>
</tr>
<tr>
<td>3. Non-homologous end joining: 7 genes, 78 SNPs</td>
<td>DCLRE1B (11), Lig4 (6), NEJHEH1 (6), PRKDC (9), XRCC4 (17), XRCC5 (27), XRCC6 (2)</td>
</tr>
<tr>
<td>4. Homologous recombinational repair: 21 genes, 139 SNPs</td>
<td>BRCA1 (9), BRCA2 (16), BRIP1 (12), DMC1 (OTOP3) (6), EME1 (4), EME2 (1), EME3 (1), MRE11A (8), MUS81 (2), NBN (11), PALB2 (FANCN) (6), RAD50 (3), RAD51 (9), RAD51C (5), RAD51L3 (5), RAD52 (9), RAD54B (5), RAD54L (3), SHFM1 (LOC402679) (5), XRCC2 (5), XRCC3 (7), ZNF350 (8)</td>
</tr>
<tr>
<td>5. Mismatch repair: 9 genes, 101 SNPs</td>
<td>MLH1 (4), MLH3 (4), MSH2 (25), MSH3 (25), MSH4 (6), MSH5 (6), MSH6 (3), PMS1 (19), PMS2 (9), PMSA (9), PFCAN (9), PANC2 (4), FANCE (3), FANCF (2), FANCV (2), FANCI (1) shared with POLG; see below POLG, FANCL (9), FANCN (6)</td>
</tr>
<tr>
<td>6. DNA–DNA crosslink repair: 9 unique genes, 1 shared gene, 47 unique SNPs</td>
<td>TDP1 (3)</td>
</tr>
<tr>
<td>7. DNA–protein crosslink repair: 1 gene, 3 SNPs</td>
<td>ABH2 (1), ABH3 (1), ABH3 (1), MGMT (21)</td>
</tr>
<tr>
<td>8. Direct reversal repair: 3 genes, 23 SNPs</td>
<td>PARP1 (9), PARP2 (1), PARP4 (14)</td>
</tr>
<tr>
<td>9. Purification family: 5 genes, 14 SNPs</td>
<td>ATR (8), CDKN1A (6), CDKN2A (11), CHEK1 (1), CHEK2 (10), HUS1 (1), PPB1 (1), PPB1R1L3 (gene overlaps with ERCC1) (4), RAD1 (3), RAD17 (5), RAD9A (1), TELO2 (5), TP53 (3)</td>
</tr>
<tr>
<td>10. DNA damage signal transduction: 12 genes, 60 SNPs</td>
<td>ATM (6), BLM (4), REQBL (2), WRN (12)</td>
</tr>
<tr>
<td>11. Hereditary disease genes with deficient DNA repair phenotypes: 4 genes, 24 SNPs</td>
<td>CHAF1A (8), DCLRE1B (2), H2AFX (2), NEIL3 (1), PCNA (4), POLB (3), POLG (6)</td>
</tr>
</tbody>
</table>

Some genes may be associated with more than one pathway; if so, the gene was arbitrarily assigned to the most recognized pathway.
typing platforms for those who overlapped with the present study population (per SNP number of participants with duplicate genotype results ranged from 3943 to 4757); the overall concordance was 98%.

Statistical analyses
For each SNP, an exact test was used to test whether genotypic frequencies in the cancer-free controls departed from HWE (15). Logistic regression was used to estimate odds ratios, confidence intervals and \( P \)-values for the associations between each SNP and cancer risk. Odds ratios were estimated using the cancer-free group as the referent category. The hypothesis tested was that minor alleles in DNA repair genes would be most prevalent in the NMSC plus other cancer group and least prevalent in the cancer-free comparison group. Given this hypothesis, the SNP screening strategy focused solely on the comparison of the NMSC plus other cancer group to the cancer-free comparison group. The additive genetic model (SNPs coded as having 0, 1 or 2 copies of the minor allele) was used to screen the associations with \( P \)-values < 0.01 in the NMSC plus other cancer category. For the SNPs screened with the additive genetic model \( P \)-value < 0.01 for the NMSC plus other cancer group, the associations between these SNPs and the risk of other cancer only and NMSC only were also evaluated. The associations between these SNPs and the NMSC cancer-prone phenotype were then further characterized in greater detail by assessing the dominant and recessive phenotypic models of inheritance.

Analyses were carried out in the statistical environment R (http://cran.r-project.org/). All statistical tests were two-sided. Assuming a \( P \)-value of 0.05 represents a statistically significant association for a single comparison, a Bonferroni multiple comparison correction for the 759 DNA repair gene SNPs (per SNP number of participants with duplicate genotype results ranged from 1694 to 1765) was 56% male, notably higher than the other study groups. The ages ranged from 57 to 62 years. The NMSC plus other cancer group was significantly younger (43 years) than the three groups that included cancer, whose average ages ranged from 57 to 62 years. The NMSC plus other cancer group was 56% male, notably higher than the other study groups.

Results
Baseline descriptive characteristics of the four study groups are summarized in Table II. The cancer-free group was significantly younger (43 years) than the three groups that included cancer, whose average ages ranged from 57 to 62 years. The NMSC plus other cancer group was 56% male, notably higher than the other study groups. The prevalence of ever-smokers was highest (56%) in the other cancer only group, and current smoking was less prevalent in study groups that included NMSC. Other than NMSC, the most common malignancies were breast, prostate, lung and colorectal cancer.

Table III lists results for the 10 SNPs with additive model \( P \)-values < 0.01 for the cancer-prone phenotype (‘both NMSC plus other cancer’) versus the ‘no cancer’ comparison. The smallest \( P \)-value was 6 × 10\(^{-4}\), greater than the Bonferroni threshold of 6.6 × 10\(^{-5}\); thus, none of the observed associations were statistically significant after correction for multiple comparisons. The 10 SNPs represented the following 6 pathways and 8 genes: 2 nucleotide excision repair (ERCC8, ERCC3), 2 homologous recombination repair (PALB2, MDM2) and 1 each from base excision repair (TDG), direct reversal repair (MGMT), DNA damage signal transduction (CHEK2) and mismatch repair (MSH6). Both TDG and PALB2 had two SNPs associated with the cancer-prone phenotype; each of these SNP pairs comprised one non-synonymous coding SNP and one intronic SNP that were in high LD and had identical associations with the NMSC cancer-prone phenotype. The results for all 759 SNPs are summarized in Supplementary Table I, available at Carcinogenesis Online.

The top ranked SNP (rs2888805) was a non-synonymous SNP in the coding region of the thymine DNA glycosylase gene (TDG) (per minor allele odds ratio for the ‘both’ category (OR\(_{\text{Both}}\)) 1.40, 95% confidence interval (CI) 1.16–1.70). The associations were null for the other cancer only (OR\(_{\text{Other cancer only}}\) 0.98; 95% CI 0.85–1.12) and NMSC only (OR\(_{\text{NMSC only}}\) 1.07; 95% CI 0.88–1.30) comparisons. This pattern of associations was mimicked by most of the other SNPs, with the risk association largely confined to the both NMSC plus other cancer category.

A measure of the internal consistency of this finding was that the second ranked SNP was an intronic TDG SNP (rs4135150) in high LD (\( r^2 = 0.99 \)) with rs2888805 and with virtually identical results,
indicating that both SNPs were equally informative about the strength of the allelic association because they are physically linked to each other. A total of six SNPs were successfully genotyped that tagged regions of TDG, rs2888805 and rs4135150 plus four more SNPs: rs4964435, rs6539116, rs4135113, and rs2583274. None of these other four were in high LD with one another (all \( r^2 < 0.10 \)), and none were associated with the NMSC cancer-prone phenotype except rs4964435, which was nominally significantly inversely associated with cancer risk (per minor allele OR\_both 0.81; 95\% CI 0.67–0.98, \( P \)-value 0.029).

This pattern of associations suggests that the TDG haplotype blocks tagged by these other four SNPs (rs4964435, rs6539116, rs4135113, rs2583274) are not likely to contribute to the functional changes in DNA repair gene variants and NMSC cancer-prone phenotype. SNPs were genotyped in known DNA repair genes in a population-based study to determine if germ-line polymorphisms contribute to joint susceptibility to both NMSC and other cancers. The results revealed several variant alleles that may contribute to a cancer-prone phenotype characterized by susceptibility to both NMSC and other malignancies.

An interesting observation was that carrying a non-synonymous coding SNP (rs2888805) in the base excision repair gene TDG was shown to be associated with 50% increased risk of developing NMSC plus another cancer. This polymorphic SNP leads to a methionine residue at position 367, or rarely leucine, rather than the more common valine (16). This residue does not fall within the known active sites for the protein or targets of acetylation, and its relevance to overall protein function is currently unknown. In the absence of direct experimental data, bioinformatics databases were queried to assess for the putative functionality and, consistent with the information above, TDG SNP rs2888805 was predicted to be benign in the PolyPhen-2, Condel and SIFT bioinformatics databases.

Table IV summarizes associations for the same 10 SNPs according to genotype (two degrees of freedom) and dominant and recessive modes of inheritance. Nine SNPs (all except rs797690) were more consistent with the dominant than recessive model based on the

Table III. DNA repair gene variants and NMSC cancer-prone phenotype

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Pathway</th>
<th>Type</th>
<th>MAF</th>
<th>( P )-HWE</th>
<th>Other cancer only</th>
<th>NMSC only</th>
<th>Both NMSC plus other cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2888805</td>
<td>TDG</td>
<td>BER</td>
<td>Nonsyn</td>
<td>0.10</td>
<td>0.25</td>
<td>0.98 (0.85–1.12)</td>
<td>1.07 (0.88–1.30)</td>
<td>1.40 (1.16–1.70)</td>
</tr>
<tr>
<td>rs4135150</td>
<td>TDG</td>
<td>BFR</td>
<td>Intron</td>
<td>0.10</td>
<td>0.38</td>
<td>0.98 (0.86–1.13)</td>
<td>1.08 (0.89–1.31)</td>
<td>1.40 (1.15–1.70)</td>
</tr>
<tr>
<td>rs1038144</td>
<td>ERCC8</td>
<td>NER</td>
<td>Intron</td>
<td>0.09</td>
<td>0.23</td>
<td>1.07 (0.93–1.24)</td>
<td>0.96 (0.77–1.19)</td>
<td>1.38 (1.12–1.71)</td>
</tr>
<tr>
<td>rs152451</td>
<td>PALB2</td>
<td>HR</td>
<td>Nonsyn</td>
<td>0.09</td>
<td>0.64</td>
<td>0.95 (0.83–1.10)</td>
<td>1.01 (0.82–1.25)</td>
<td>1.35 (1.10–1.66)</td>
</tr>
<tr>
<td>rs447529</td>
<td>PALB2</td>
<td>HR</td>
<td>Intron</td>
<td>0.09</td>
<td>0.64</td>
<td>0.95 (0.82–1.10)</td>
<td>1.01 (0.82–1.24)</td>
<td>1.35 (1.10–1.65)</td>
</tr>
<tr>
<td>rs11016857</td>
<td>MGMT</td>
<td>DR</td>
<td>Nonsyn</td>
<td>0.09</td>
<td>0.64</td>
<td>0.95 (0.82–1.10)</td>
<td>1.01 (0.82–1.24)</td>
<td>1.35 (1.10–1.65)</td>
</tr>
<tr>
<td>rs4150654</td>
<td>ERCC3</td>
<td>NER</td>
<td>Intron</td>
<td>0.39</td>
<td>0.44</td>
<td>1.01 (0.92–1.11)</td>
<td>0.89 (0.77–1.03)</td>
<td>0.82 (0.70–0.95)</td>
</tr>
<tr>
<td>rs1650024</td>
<td>DMC1</td>
<td>HR</td>
<td>Nonsyn</td>
<td>0.46</td>
<td>0.50</td>
<td>1.00 (0.92–1.09)</td>
<td>0.99 (0.88–1.12)</td>
<td>0.84 (0.73–0.96)</td>
</tr>
<tr>
<td>rs2347443</td>
<td>CHEK2</td>
<td>Transduc</td>
<td>Intron</td>
<td>0.16</td>
<td>0.92</td>
<td>1.09 (0.98–1.22)</td>
<td>1.01 (0.85–1.19)</td>
<td>1.25 (1.06–1.48)</td>
</tr>
<tr>
<td>rs797690</td>
<td>MSH6</td>
<td>MMR</td>
<td>5' UTR</td>
<td>0.26</td>
<td>0.43</td>
<td>1.10 (1.00–1.21)</td>
<td>1.13 (0.98–1.30)</td>
<td>1.21 (1.05–1.40)</td>
</tr>
</tbody>
</table>

Allelic trend test odds ratios (ORs) and 95\% CI are summarized for each SNP for comparisons of the following groups with the comparison group with no cancer: (i) other cancer only; (ii) NMSC only and (iii) both NMSC plus another cancer; all results adjusted for residual confounding by ancestry by including first three principal components.

\( ^1 \)BER, base excision repair; NHEJ, non-homologous end-joining; NER, nucleotide excision repair; DR, direct reversal; HR, homologous recombination; Transduc, DNA damage signal transduction; MMR, mismatch repair.

\( ^2 \)Nonsyn, non-synonymous coding; 5' UTR, 5' untranslated region.

\( ^3 \)MAF, minor allele frequency.

\( ^4 \)\( P \)-value for Hardy–Weinberg equilibrium.

Discussion

The present study was carried out to test the hypothesis that germ-line variants in DNA repair genes were associated with the NMSC cancer-prone phenotype. SNPs were genotyped in known DNA repair genes in a population-based study to determine if germ-line polymorphisms contribute to joint susceptibility to both NMSC and other cancers. The results revealed several variant alleles that may contribute to a cancer-prone phenotype characterized by susceptibility to both NMSC and other malignancies.
alleles (dominant model) and two versus less than two minor alleles (recessive model) of both NMSC plus other cancers compared with comparison group with no cancer, according to genotype and one or more versus zero minor alleles (dominant model) and two versus less than two minor alleles (recessive model).

2Adjusted for residual confounding by ancestry by including first three principal components.

3Numbers do not sum to 5916 due to genotyping failures.

2 (rs167715) were significantly inversely associated with one-year incidence, among vinyl-chloride exposed workers in China (25). In a DNA repair pathway study, no significant associations were observed for the eleven TDG tagging SNPs studied in relation to meningioma risk (26). Minor alleles of SNPs in TDG intron 1 (rs2374327) and intron 2 (rs167715) were significantly inversely associated with one-year transplant-related mortality after hematopoietic stem cell transplant (27).

Table IV. Odds ratios (OR) and 95% Confidence Intervals (CI) for the association between SNPs with additive model P-value < 0.01 and risk of both NMSC plus other cancers compared with comparison group with no cancer, according to genotype and one or more versus zero minor alleles (dominant model) and two versus less than two minor alleles (recessive model)

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>No. of Minor alleles</th>
<th>Genotype OR (95% CI)</th>
<th>Dominant model2,3 OR (95% CI) P-value</th>
<th>Recessive Model2,4 OR (95% CI) P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2888805</td>
<td>TDG</td>
<td>0 (432/1920)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
</tr>
<tr>
<td>rs435150</td>
<td>TDG</td>
<td>0 (433/1886)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
</tr>
<tr>
<td>rs1038144</td>
<td>ERCC8</td>
<td>0 (452/1924)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
</tr>
<tr>
<td>rs152451</td>
<td>PALB2</td>
<td>0 (444/1893)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
</tr>
<tr>
<td>rs447529</td>
<td>PALB2</td>
<td>0 (465/1892)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
</tr>
<tr>
<td>rs1101685</td>
<td>MGMT</td>
<td>0 (216/714)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
</tr>
<tr>
<td>rs1450454</td>
<td>ERCC3</td>
<td>0 (229/631)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
</tr>
<tr>
<td>rs11650024</td>
<td>DMC1</td>
<td>0 (194/640)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
</tr>
<tr>
<td>rs2347443</td>
<td>CHEK2</td>
<td>0 (379/1646)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
</tr>
<tr>
<td>rs797690</td>
<td>MSH6</td>
<td>0 (305/1312)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
<td>1.0 (referent)</td>
</tr>
</tbody>
</table>

Numbers do not sum to 5916 due to genotyping failures.

Adjusted for residual confounding by ancestry by including first three principal components.

Two minor alleles versus zero or one minor allele.

transplant-related mortality after hematopoietic stem cell transplant (27). Two previous reports specifically of rs2888805 were identified. In one, in 94 cases with a family history of colorectal cancer that excluded those with known colorectal cancer susceptibility alleles, rs2888805 was heterozygous in >10 patients; there was no control group (16). In the other study, rs2888805 was not significantly associated with lung cancer risk in a study of 96 cases and 94 controls (23). Thus, the limited available evidence on TDG polymorphisms in relation to cancer or biomarkers of cancer is equivocal, neither clearly supporting nor refuting possible associations with cancer risk in different contexts. The present study focused on a study population that is hypothesized to be enriched for genetic susceptibility to cancer, a circumstance that may enhance the likelihood of detecting associations that may otherwise go undetected in studies of a single cancer.

In addition to the two TDG SNPs, eight other SNPs were associated with the NMSC cancer-prone phenotype with additive model P-values < 0.01, raising the index of suspicion that these SNPs may be markers for variants that also contribute to the phenotype. These eight SNPs were located in six pathways and seven genes: two nucleotide excision repair (ERCC8, ERCC3), two homologous recombination repair (PALB2, DMC1) and one each from direct reversal repair (MGMT), DNA damage signal transduction (CHEK2) and mismatch repair (MSH6).

For only three of these eight SNPs were previously published reports identified, the two PALB2 SNPs (rs152451, rs447529) and the CHEK2 SNP (rs2347743). PALB2 (for Partner and Localizer of BRCA2) has been extensively studied in relation to hereditary breast cancer but is also associated with Fanconi anemia and pancreatic cancer (28). rs152451 is a non-synonymous coding SNP that may alter
the PALB2 protein, but queries of the PolyPhen-2, Condel and SWIFT bioinformatics databases indicated it was likely to be benign. Studies of familial or bilateral breast cancer documented that the prevalence of rs152451 heterozygotes ranged from 9% to 22%, consistently higher than that in the general population. An analysis of rs447529 in our study population, in a study in China, rs447529 was significantly associated with breast cancer risk (33). Specific CHEK2 mutations are known to increase susceptibility to breast cancer (34), and CHEK2 variants have been studied in relation to Li Fraumeni syndrome and a diverse suite of other malignancies with equivocal results (35). In a case–control study, the intronic SNP rs2347443 was not significantly associated with breast cancer risk (36).

No previous reports of the other five SNPs were ascertained, but a brief synopsis of the relevance of the genes is provided below. The two genes from the NER pathway were ERCC8 and ERCC3. ERCC8, also known as Cockayne syndrome A (CSA), is not only involved in NER but also in protecting DNA from damage via oxidative stress (37). Rare, high-penetrant mutations in ERCC8 are linked with Cockayne syndrome. ERCC3 encodes for XPB, which comprises an important element of NER as well as the general transcription factor IHI (TFIIH) complex, so it is critical not only to NER but also to gene transcription (38). ERCC3 mutations have been implicated in XP, Cockayne syndrome, and trichothiodystrophy (39). The DMCI gene contributes to meiosis-specific homologous recombination repair (40), with the only previous report of DMC1 variants in relation to cancer risk finding an association with cervical cancer (41). MSH6 is a mismatch repair gene, with mutations implicated in Lynch syndrome (42) and endometrial cancer (43). O6-methylguanine-DNA methyltransferase (MGMT) directly repairs alkyl adducts that arise in the O6 position of guanine, preventing G-C to A-T transition mutations (44). Germ-line MGMT polymorphisms could potentially affect carcinogenesis by directly diminishing MGMT-mediated DNA repair, either by reducing methyltransferase activity via a functional coding change or by increasing MGMT’s susceptibility to epigenetic silencing of its transcription.

In summary, this list includes an intriguing set of candidate SNPs that warrant consideration as potential genetic markers of susceptibility to the NMSC cancer-prone phenotype. By way of caution, the very nature of the candidate gene pathway approach ensured that these genes would have strong a priori biologic plausibility based on their key roles in carcinogenic pathways.

Associations observed in the NMSC plus other cancer group were not sustained in the NMSC only and other cancer only groups. The confinement of the risk associations to the NMSC plus other cancer group may reflect a greater likelihood of genetically driven, rather than environmentally driven, cancer susceptibility. A high-risk subgroup is expected to be enriched for genetic susceptibility markers associated with a predisposition to cancer. In contrast, environmental exposures likely play a more prominent role in driving carcinogenesis in the NMSC only and other cancer only groups, exposures that are more likely cancer-type specific (e.g. sun exposure for NMSC and smoking for lung cancer). Hence, in assessing risk for multiple cancers that lack common environmental risk factors (e.g. NMSC and lung cancer), genetic risk factors would be expected to predominate. This is analogous to other situations for which cancers are also more strongly genetically driven than the average-risk population, such as phenotypic markers defined by early age of cancer onset or a positive family history of cancer. Studying the relationship between NMSC and risk of other malignancies thus provides a fertile research model for identifying markers of potential susceptibility to human cancer.

The present study had multiple strengths, including addressing a novel question embedded within a well-established community-based cohort study. No previous study has tested a hypothesis to understand why a personal history of NMSC is associated with enhanced susceptibility to other cancers. The hypothesis that germ-line polymorphisms in DNA repair genes may contribute to this general cancer-prone phenotype was rigorously tested via thorough coverage of allelic variants in DNA repair pathway genes. Using the co-occurrence of NMSC and another cancer as a sentinel for a cancer-prone phenotype is novel and may enhance detection of common allelic variants that might otherwise escape notice as markers of cancer susceptibility. This is particularly true when one considers that in contrast to the complications usually encountered when studying multiple primary cancers, NMSC has the advantages that it is rarely fatal and therefore less susceptible to survival biases and, because most tumors are locally excised, its treatment is not associated with risk of other cancers.

Nevertheless, caution is warranted in drawing inferences. Even though a cohort of ~30 000 followed-up for almost two decades enabled the conclusion of a total of 3693 cancer cases, the primary study inferences centered on the both NMSC plus another cancer category, which was limited to 577 cancer cases. A resulting weakness was limited statistical power, a likely reason that no associations survived conservative significance thresholds imposed by corrections for multiple comparison testing. The sample size constraints of the both NMSC plus other cancer category highlight challenges of this line of inquiry and emphasize the need for replication of our findings. These same sample size constraints precluded meaningful subgroup analyses, such as separate analyses for BCC and SCC, but the totality of the current evidence suggests that the NMSC cancer-prone phenotype is consistently observed for both histologic types. A limitation of the genetic association study approach is that we have identified statistical associations, but confirmation of the specific SNPs that might be involved would require additional experiments not included in the present study. In this discovery phase research, the results identified the genetic regions of DNA repair genes where variants are most likely to be associated with the risk of NMSC plus another cancer. If the observed associations are genuine, the association of these SNPs with the NMSC cancer-prone phenotype allows us to infer that allelic variations in TDG and the other genes identified may affect cancer development but it does not allow us infer causality for the SNPs themselves, because they are simply markers for gene regions that contain many other sequence variations. Thus, a limitation of this study is that further research will be needed to identify and characterize the specific functional variants within these gene regions and the protein change in function that might modify cellular DNA repair capacity. In the absence of direct experimental data, bioinformatics databases were queried to assess for the putative functionality of the two non-synonymous SNPs. For both TDG SNP rs2888805 and PALB2 SNP rs152451, the scores for the PolyPhen-2, Condel and SIFT bioinformatics databases were consistent in indicating that these SNPs were expected to be benign.

In conclusion, the findings have generated new leads suggesting that common variants in DNA repair genes may affect the increased overall cancer risk experienced by individuals with a personal history of NMSC. The strongest association was for a non-synonymous coding SNP in TDG, suggesting that base excision repair, possibly via regulation of the epigenome, may be important in the NMSC cancer-prone phenotype.

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

Funding
National Cancer Institute (NCI) (R01 CA105069, HHSN26120080001E); Intramural Research Program of the NCI, Center for Cancer Research.

Acknowledgements
Cancer incidence data were provided by the Maryland Cancer Registry, Center for Cancer Surveillance and Control, Department of Health and Mental Hygiene, which is funded by the State of Maryland, the Maryland Cigarette Restitution
References


Received December 30, 2011; revised May 2, 2012; accepted May 8, 2012