A case-control study of the association of the polymorphisms and haplotypes of DNA ligase I with lung and upper-aerodigestive-tract cancers

Yuan-Chin Amy Lee1, Hal Morgenstern2,3, Sander Greenland4,5, Donald P. Tashkin6, Jeanette Papp6, Janet Sinsheimer6, Wei Cao7, Mia Hashibe7, Nai-Chieh Y. You1, Jenny T. Mao5, Wendy Cozen8, Thomas M. Mack8 and Zuo-Feng Zhang1*

1Department of Epidemiology, University of California at Los Angeles School of Public Health, Los Angeles, CA
2Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI
3Department of Environmental Health Sciences, University of Michigan School of Public Health, Ann Arbor, MI
4Department of Statistics, University of California at Los Angeles, Los Angeles, CA
5Division of Pulmonary and Critical Care Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA
6Department of Human Genetics, University of California at Los Angeles, Los Angeles, CA
7International Agency for Research on Cancer, Lyon, France
8Department of Preventive Medicine, Keck School of Medicine at University of Southern California, Los Angeles, CA

Tobacco smoking is a major risk factor for lung and upper-aerodigestive-tract (UADT) cancers. One possible mechanism for the associations may be through DNA damage pathways. DNA Ligase I (LIG1) is a DNA repair gene involved in both the nucleotide excision repair (NER) and the base excision repair (BER) pathways. We examined the association of four LIG1 polymorphisms with lung and UADT cancers, and their potential interactions with smoking in a case-control study in Los Angeles County. We performed genotyping using the SNPlex method from Applied Biosystems. Logistic regression analyses of 551 lung cancer cases, 489 UADT cancer cases and 948 controls showed the expected associations of tobacco smoking with lung and UADT cancers and new associations between the LIG1 haplotypes and these cancers. For lung cancer, when compared to the most common haplotype (rs20581-rs20580-rs20579-rs439132 = T-C-C-A), the adjusted odds ratio (OR) is 1.2 (95% confidence limits (CL) = 0.95, 1.5) for the CACA haplotype, 1.4 (1.0, 1.9) for the CATA haplotype and 1.8 (1.1, 2.8) for the CCCG haplotype, after controlling for age, gender, race/ethnicity, education and tobacco smoking. We observed weaker associations between the LIG1 haplotypes and UADT cancers. Our findings suggest the LIG1 haplotypes may affect the risk of lung and UADT cancers.

Key words: LIG1; polymorphisms; haplotypes; lung cancer; upper-aerodigestive-tract cancers

Lung cancer patients have the poorest survival rates among cancer patients. The 5-year survival rate of lung cancer is only 11% worldwide. It is also the most common cancer in the world and the third most common cancer in the United States. In the United States, lung cancer incidence ranks number three, while lung cancer mortality ranks number one in 2006. Although several risk factors for lung cancer such as tobacco smoking have been identified, little is known about the role of genetic susceptibility and its potential interactions with environmental factors. Studies have examined associations between lung cancer and the genetic polymorphisms of DNA repair genes; however, there are no conclusive results on potential interactions with environmental factors.

Upper-aerodigestive-tract (UADT) cancers of the oral cavity, pharynx, larynx and esophagus accounts for 10% of new cases and 11% of cancer deaths worldwide in 2002. In the United States, the estimated incidence and death rates are high in 2006. Oral/pharyngeal, laryngeal and esophageal cancers show 5-year survival rates of ~59, 64 and 16%, respectively, during 1995–2000. Tobacco and alcohol exposures have been well documented as risk factors for head and neck squamous cell carcinomas. However, results of previous studies on the potential associations between UADT cancers and the genetic polymorphisms of DNA repair genes and their possible interactions with environmental factors have been inconclusive.

DNA is vulnerable to injury due to endogenous and/or exogenous exposures such as exposure to carcinogens in tobacco smoke. DNA repair systems play an important role in maintaining the integrity of the genome. Deficiencies in this defense system are thought to contribute to the development of cancer. Based on a review of epidemiologic studies, Berwick and Vines suggested that reduced DNA repair capacity was associated with increased risk of cancers, with odds ratios in the range of 1.4–75, with the majority of values between 2 and 10. In addition, genetic variation in DNA repair genes may influence DNA repair capacity; consequently, DNA repair genetic polymorphisms may affect cancer susceptibility.

DNA ligase I (LIG1) is a 102-kDa nuclear enzyme, which is involved in joining Okazaki fragments during DNA replication and in sealing single-strand breaks. It plays a role in both the nucleotide excision repair (NER) and the long-patch base excision repair (BER) pathways. LIG1 is an important DNA repair pathway against tobacco-related carcinogenesis. A total of 31 SNPs for LIG1 were identified in SNP500 database (http://snp500cancer.nci.nih.gov/snplist.cfm). The polymorphic sites selected are shown in Figure 1. The T → C synonymous SNP on exon 25 (rs20581), the C → A synonymous SNP on exon 7 (rs20580), the C → T synonymous SNP at 5’ UTR on exon 2 (rs20579) and A → G SNP on intron 1 (rs439132) do not result in any alteration of amino acid and their functional variation is unknown. They were selected because the LIG1 gene plays a very important role in the BER pathway, and their rare allele frequencies are greater or equal to 5%. The D’ between rs20581 and rs20580 is 1.0; the D’ between rs20580 and rs20579 is 0.842; the D’ between rs20581 and rs20579 is 1.0. These polymorphisms may affect the transcription or translation of LIG1 protein for DNA repair and, in turn, may influence DNA repair capacity. The effect would thus take place after damage has occurred during carcinogenesis. These 4 SNPs may be associated with cancer risks separately or jointly. Therefore, the haplotypes of LIG1 may play a very important role.

We are aware of only one epidemiologic study that has examined the association between one of the LIG1 polymorphisms...
(rs20580) and lung cancer. That study, however, did not examine the associations of the other LIG1 SNPs, haplotypes and lung cancer, and the joint effects of the LIG1 SNPs with tobacco smoking. We know of no published studies that have explored the associations between these LIG1 SNPs, haplotypes and lung and UADT cancers. The objectives of the present study were to evaluate the associations between the DNA repair genetic polymorphisms and haplotypes of DNA Ligase I (LIG1) and lung and UADT cancer risks in the residential population of Los Angeles County, and to investigate possible interactions between the LIG1 SNPs and haplotypes and environmental factors, particularly tobacco smoking.

**Material and methods**

**Study design and study participants**

We used questionnaire data and buccal cell samples from a population-based case-control study, including 611 incident cases of lung cancer, 601 incident cases of UADT cancer (403 oral/pharyngeal, 90 laryngeal and 108 esophageal) and 1,040 population controls. Histologically confirmed newly diagnosed cancer cases were obtained by the rapid ascertainment system of the Cancer Surveillance Program for Los Angeles County, which is administered by the Keck School of Medicine and Norris Comprehensive Cancer Center at the University of Southern California. Over 95% of cancer reports were histologically verified; the remainder was verified by MRI, CT scan or other diagnostic methods. Controls did not have a history of oral/pharyngeal, laryngeal, esophageal or lung cancer and were identified using a formal algorithm that enumerated households in the neighborhoods of the cases. Controls were individually matched to cases by age (within 10-year categories), gender and neighborhood.

All participants met the following criteria: (i) were residents of Los Angeles County at the time of diagnosis for cases or at the time of recruitment for controls, (ii) were residents of Los Angeles County at the time of diagnosis for cases or at the time of recruitment for controls, (iii) were 18–62 years of age during the study period of 1999–2004 and (iii) spoke either English or Spanish, or had translators available at the time of interview. Additional details regarding recruitment and other exclusion criteria have been described previously. Each case and control participated in in-person interviews consisting of standardized questionnaires administered by specially trained interviewers. After the interview, buccal swabs were collected for subsequent DNA extraction.

Of the lung cancer and UADT cancer cases, respectively, 57 and 49% were interviewed within 3 months of diagnosis, 32 and 38% 3–6 months after diagnosis, and 11 and 12% more than 6 months after diagnosis. The response rates for those who were eligible to be in the study were 68, 39, 49, 40 and 34% for controls, lung cancer, oral/pharyngeal cancer, laryngeal cancer, and esophageal cancer cases, respectively. Furthermore, the response rates for those who completed interviews to provide the biospecimen were 89, 89, 68, 88 and 90% for controls, lung cancer, oral/pharyngeal cancer, laryngeal cancer and esophageal cancer cases, respectively. The protocol was approved by the Institutional Review Boards of the University of California, Los Angeles and the University of Southern California. Informed consent was obtained from every study subject.

**Buccal cell samples**

After participants were interviewed, interviewers collected ~25–50 ml of buccal cells with Scope™ mouthwash solution in tubes. All buccal cell samples were transported as soon as possible to the School of Public Health, University of California, Los Angeles where samples were stored at 4°C in the molecular epidemiology laboratory refrigerator. The buccal cell samples were centrifuged to separate the buccal cells and mouthwash solution. High molecular weight DNA was isolated by an altered phenol-chloroform method. DNA purity and concentration were determined by a spectrophotometric measurement of absorbance at 260/280 nm.

**Genotyping assays**

Genotypes of LIG1 were determined using the Applied Biosystems SNPlex™ assay. This method assays SNP genotypes in 48 SNP PCR multiplexes using an oligonucleotide ligation process. Each SNP allele was ligated to an allele-specific oligonucleotide, then hybridized with a Zipchute™ probe with a mobility modifier and a fluorescent label, allowing the probe to be separated and detected by capillary electrophoresis. Detection was performed on an Applied Biosystems 3730 DNA Analyzer, and data interpretation was performed with the Applied Biosystems Genemapper software v4.0, which uses a clustering algorithm with stringent proprietary quality checks to call genotypes. SNPlex genotyping was performed by the UCLA Genotyping Core. All automatic calls by the software were evaluated by Core technicians. Any SNPs with less than 90% of the sample auto-called by the software were either rescanned manually or discarded if clustering confidence was low. Samples were run with 2 positive controls of CEPH individual 1347-2 and 2 “no template”, negative controls per 96-well plate. Any run in which the positive controls failed to cluster was discarded. Only genotypes with a Genemapper Quality Score of greater than 95% were passed. On the DNA samples in this study, the call rate was 96%, the reproducibility rate was 97.8% and the concordance rate was 99.8%. Reproducibility was determined by comparing blinded replicates on plates and by rerunning entire plates. The reproducibility reported was calculated over 10,000 duplicate genotypes (3% of the entire sample), including not only blinded replicates on the same plate or the...
<table>
<thead>
<tr>
<th>Variables</th>
<th>Lung cancer cases ($n = 611$), $N$ (%)</th>
<th>Controls ($n = 1029$), $N$ (%)</th>
<th>UADT cancer cases ($n = 601$), $N$ (%)</th>
<th>Controls ($n = 1039$), $N$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With DNA (551)</td>
<td>Without DNA (60)</td>
<td>With DNA (489)</td>
<td>Without DNA (91)</td>
</tr>
<tr>
<td>Mean age (years), (age range)</td>
<td>52.2 (32–59)</td>
<td>52.3 (41–59)</td>
<td>50.6 (23–59)</td>
<td>49.9 (17–62)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.90</td>
<td>0.37</td>
<td>0.21</td>
<td>0.38</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>4 (0.73)</td>
<td>0 (0.0)</td>
<td>22 (5.0)</td>
<td>47 (5.0)</td>
</tr>
<tr>
<td>35–44</td>
<td>51 (9.3)</td>
<td>6 (10)</td>
<td>155 (17)</td>
<td>155 (16)</td>
</tr>
<tr>
<td>45–54</td>
<td>273 (50)</td>
<td>28 (47)</td>
<td>452 (48)</td>
<td>452 (48)</td>
</tr>
<tr>
<td>≥55</td>
<td>223 (40)</td>
<td>26 (43)</td>
<td>294 (31)</td>
<td>294 (31)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.88</td>
<td>0.79</td>
<td>0.85</td>
<td>0.64</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>320 (58)</td>
<td>39 (65)</td>
<td>583 (62)</td>
<td>588 (62)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>50 (9.1)</td>
<td>3 (5.0)</td>
<td>140 (15)</td>
<td>141 (15)</td>
</tr>
<tr>
<td>African American</td>
<td>90 (16)</td>
<td>6 (10)</td>
<td>79 (8.4)</td>
<td>79 (8.5)</td>
</tr>
<tr>
<td>Asian American</td>
<td>59 (11)</td>
<td>11 (18)</td>
<td>53 (5.7)</td>
<td>55 (5.8)</td>
</tr>
<tr>
<td>Other</td>
<td>31 (5.6)</td>
<td>1 (1.7)</td>
<td>83 (8.9)</td>
<td>84 (8.9)</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>p-value</td>
<td>0.13</td>
<td>p-value &lt; 0.0001</td>
<td>0.09</td>
<td>p-value &lt; 0.0001</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>278 (50)</td>
<td>25 (42)</td>
<td>575 (61)</td>
<td>582 (61)</td>
</tr>
<tr>
<td>Female</td>
<td>273 (50)</td>
<td>35 (58)</td>
<td>364 (39)</td>
<td>366 (39)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.20</td>
<td>p-value = 0.004</td>
<td>0.10</td>
<td>p-value = 0.002</td>
</tr>
<tr>
<td>Education Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤12 years</td>
<td>248 (45)</td>
<td>17 (28)</td>
<td>276 (29)</td>
<td>276 (29)</td>
</tr>
<tr>
<td>&gt;12 to ≤16 years</td>
<td>248 (45)</td>
<td>27 (45)</td>
<td>422 (45)</td>
<td>431 (46)</td>
</tr>
<tr>
<td>&gt;16 years</td>
<td>55 (10)</td>
<td>16 (27)</td>
<td>240 (26)</td>
<td>240 (25)</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0003</td>
<td>0.22</td>
<td>0.60</td>
<td>0.22</td>
</tr>
</tbody>
</table>
same processing, but also plates repeated on the entire assay, including all preprocessing steps, as much as 6 months apart. Concordance was determined by Taqman assay on ABI 7900. For concordance tests, we chose one SNP with high call rates, one average SNP and one SNP with low call rates. These SNPs were genotyped on one 96-well plate with high-quality DNA and high call rates and one plate with low-quality DNA and low call rates. The concordance rate of genotypes called successfully on both SNPlex and Taqman was not affected by SNP quality or DNA quality, although the samples with low call rates on SNPlex also had low call rates on Taqman. All genotypes were checked for Hardy-Weinberg equilibrium and divergence from published allele frequencies. Checks were performed across the entire sample and on individual genotyping runs to identify any possibly bad runs. Any markers, sample sets or genotyping runs that gave suspect results on HWE or on any of our other quality control validation tests were manually rescored.

Statistical analyses

Data analyses were performed with SAS 9 software. All P-values are 2-sided. To increase the precision and power over standard matched analyses, we compared each cancer group (lung and UADT) with all controls using unconditional logistic regression, including terms for the matching factors, age and sex, which allowed us to include cases with no matched controls. Odds-ratio estimates (ORs) and 95% confidence limits (CLs) were computed with and without adjustment for potential confounders including age, gender, race/ethnicity (white, Hispanic, African American, Asian/Pacific Islander and other), educational level, alcohol drinking and tobacco smoking. To minimize age confounding and to account for age matching, age was controlled in fine categories (<35, 35–36, 37–38, 39–40, 41–42, 43–44, 45–46, 47–48, 49–50, 51–52, 53–54, 55–56, 57–58 and 59–62). Controls who were more than 3 years younger than the youngest case or 3 years older than the oldest case were excluded from the analysis. Educational level (years of schooling) and tobacco smoking (pack-years) were controlled as continuous variables.

We assessed the results for the LIG1 polymorphisms by treating participants with the homozygote wild-type genotype as the referent. We performed the Pearson (χ²) test to evaluate Hardy-Weinberg equilibrium among the controls. The haplotypes were reconstructed from our population genoetype data, using PHASE v2.26,27 We treated the most frequent haplotype as the referent. The impact of tobacco smoking on the association of the LIG1 genotypes and haplotypes was also investigated by stratified analysis.

Results

Lung cancer cases ranged from 32 to 59 years of age, while oral/pharyngeal, laryngeal and esophageal cancer cases ranged from 17 to 59 years of age. Based on the age restriction mentioned previously, eleven controls were excluded from the lung-cancer analyses, and one control was excluded from the UADT-cancer analyses.

Among the lung cancer cases, there were 95 squamous cell carcinoma, 297 adenocarcinoma, 115 large cell lung cancer and 75 small cell lung cancer; among the UADT cancer cases, there were 497 squamous cell carcinomas. Table I shows the comparisons between interviewed cases and controls with and without DNA specimens. There were appreciable differences in race/ethnicity between those with and without DNA specimens among both UADT cancer cases and population controls; African Americans were less likely to provide DNA specimens. Women were less likely than men to provide DNA specimens in all case and control groups, and college-educated cases of lung cancer were less likely to provide DNA specimens than those with less education. The association of DNA sampling with these covariates implies that adjustment for them would be needed.

Table II shows the associations between tobacco smoking and lung and UADT cancer risk. Tobacco smoking (> 40 pack-years vs. 0 pack-years) was strongly associated with increased cancer risk with odds ratios of 22 (95% CL = 15, 33) for lung cancer and 4.3 (95% CL = 2.9, 6.4) for UADT cancers after adjusting for age, gender, ethnicity and education (plus alcohol drinking for UADT cancers only). The age at initiation of smoking (≤16-years old and >16-years old vs. never started) was also strongly associated with lung cancer (OR = 5.7, 95% CL = 4.3, 7.6 for ≤16-
<table>
<thead>
<tr>
<th>Genotype of LIG1</th>
<th>Cases</th>
<th>Controls</th>
<th>Crude OR</th>
<th>Adj. OR ¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rs20581 (Exon 25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>78</td>
<td>142</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>148</td>
<td>346</td>
<td>0.78 (0.56, 1.1)</td>
<td>0.92 (0.62, 1.4)</td>
</tr>
<tr>
<td>C/C</td>
<td>86</td>
<td>155</td>
<td>1.0 (0.69, 1.5)</td>
<td>1.2 (0.74, 1.8)</td>
</tr>
<tr>
<td>Stratified by smoking status in pack-years²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>12</td>
<td>62</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>23</td>
<td>164</td>
<td>0.73 (0.34, 1.5)</td>
<td>0.70 (0.32, 1.5)</td>
</tr>
<tr>
<td>C/C</td>
<td>21</td>
<td>71</td>
<td>1.5 (0.70, 3.4)</td>
<td>1.1 (0.49, 2.7)</td>
</tr>
<tr>
<td>&gt;0–20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>7</td>
<td>45</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>31</td>
<td>116</td>
<td>1.7 (0.71, 4.2)</td>
<td>1.6 (0.62, 3.9)</td>
</tr>
<tr>
<td>C/C</td>
<td>15</td>
<td>60</td>
<td>1.6 (0.61, 4.3)</td>
<td>1.5 (0.54, 4.1)</td>
</tr>
<tr>
<td>&gt;20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>59</td>
<td>35</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>94</td>
<td>66</td>
<td>0.85 (0.50, 1.4)</td>
<td>0.81 (0.47, 1.4)</td>
</tr>
<tr>
<td>C/C</td>
<td>50</td>
<td>23</td>
<td>1.3 (0.68, 2.5)</td>
<td>1.1 (0.58, 2.2)</td>
</tr>
<tr>
<td>Rs20580 (Exon 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>110</td>
<td>202</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/C</td>
<td>211</td>
<td>398</td>
<td>0.97 (0.73, 1.3)</td>
<td>1.0 (0.72, 1.4)</td>
</tr>
<tr>
<td>A/A</td>
<td>121</td>
<td>179</td>
<td>1.2 (0.90, 1.7)</td>
<td>1.2 (0.85, 1.8)</td>
</tr>
<tr>
<td>Stratified by smoking status in pack-years²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>20</td>
<td>97</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/C</td>
<td>36</td>
<td>183</td>
<td>0.95 (0.52, 1.7)</td>
<td>0.83 (0.44, 1.6)</td>
</tr>
<tr>
<td>A/A</td>
<td>24</td>
<td>80</td>
<td>1.5 (0.75, 2.8)</td>
<td>1.1 (0.55, 2.3)</td>
</tr>
<tr>
<td>&gt;0–20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>14</td>
<td>62</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/C</td>
<td>34</td>
<td>139</td>
<td>1.1 (0.54, 2.2)</td>
<td>1.1 (0.52, 2.2)</td>
</tr>
<tr>
<td>A/A</td>
<td>21</td>
<td>66</td>
<td>1.4 (0.66, 3.0)</td>
<td>1.5 (0.67, 3.2)</td>
</tr>
<tr>
<td>&gt;20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>76</td>
<td>43</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/C</td>
<td>141</td>
<td>76</td>
<td>1.1 (0.66, 1.7)</td>
<td>1.0 (0.64, 1.7)</td>
</tr>
<tr>
<td>A/A</td>
<td>76</td>
<td>33</td>
<td>1.3 (0.75, 2.3)</td>
<td>1.2 (0.71, 2.2)</td>
</tr>
<tr>
<td>Rs20579 (Exon 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>294</td>
<td>586</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>118</td>
<td>187</td>
<td>1.3 (0.96, 1.6)</td>
<td>1.3 (0.97, 1.8)</td>
</tr>
<tr>
<td>T/T</td>
<td>11</td>
<td>7</td>
<td>3.1 (1.2, 8.2)</td>
<td>2.1 (0.74, 5.9)</td>
</tr>
<tr>
<td>Stratified by smoking status in pack-years²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>58</td>
<td>276</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>15</td>
<td>88</td>
<td>0.81 (0.44, 1.5)</td>
<td>0.84 (0.44, 1.6)</td>
</tr>
<tr>
<td>T/T</td>
<td>1</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>&gt;0–20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>40</td>
<td>192</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>23</td>
<td>66</td>
<td>1.7 (0.93, 3.0)</td>
<td>1.7 (0.92, 3.1)</td>
</tr>
<tr>
<td>T/T</td>
<td>4</td>
<td>4</td>
<td>4.8 (1.2, 20)</td>
<td>3.8 (0.86, 17)</td>
</tr>
<tr>
<td>&gt;20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>196</td>
<td>118</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>80</td>
<td>33</td>
<td>1.5 (0.92, 2.3)</td>
<td>1.5 (0.91, 2.4)</td>
</tr>
<tr>
<td>T/T</td>
<td>6</td>
<td>3</td>
<td>1.2 (0.30, 4.9)</td>
<td>1.2 (0.30, 5.2)</td>
</tr>
<tr>
<td>Rs439132 (Intron 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>326</td>
<td>585</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/G</td>
<td>39</td>
<td>54</td>
<td>1.3 (0.84, 2.0)</td>
<td>1.3 (0.81, 2.3)</td>
</tr>
<tr>
<td>G/G</td>
<td>6</td>
<td>2</td>
<td>5.4 (1.1, 27)</td>
<td>5.3 (0.92, 30)</td>
</tr>
<tr>
<td>Stratified by smoking status in pack-years²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>51</td>
<td>284</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/G</td>
<td>10</td>
<td>23</td>
<td>2.4 (1.1, 5.4)</td>
<td>1.7 (0.74, 4.1)</td>
</tr>
<tr>
<td>G/G</td>
<td>0</td>
<td>1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>&gt;0–20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>47</td>
<td>184</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/G</td>
<td>5</td>
<td>24</td>
<td>0.82 (0.30, 2.3)</td>
<td>0.78 (0.27, 2.2)</td>
</tr>
<tr>
<td>G/G</td>
<td>2</td>
<td>1</td>
<td>7.8 (0.70, 88)</td>
<td>5.2 (0.44, 60)</td>
</tr>
<tr>
<td>&gt;20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>228</td>
<td>117</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/G</td>
<td>24</td>
<td>7</td>
<td>1.8 (0.74, 4.2)</td>
<td>1.5 (0.61, 3.8)</td>
</tr>
<tr>
<td>G/G</td>
<td>4</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

¹Adjusted for gender, age, education, ethnicity and tobacco smoking (pack-years).
²Adjusted for gender, age, education and ethnicity.
<table>
<thead>
<tr>
<th>Genotype of LIG1</th>
<th>Cases</th>
<th>Controls</th>
<th>Crude OR</th>
<th>Adj. OR*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rs20581 (Exon 25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>61</td>
<td>143</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>164</td>
<td>350</td>
<td>1.1 (0.77, 1.6)</td>
<td>1.2 (0.85, 1.8)</td>
</tr>
<tr>
<td>C/C</td>
<td>94</td>
<td>157</td>
<td>1.4 (0.95, 2.1)</td>
<td>1.5 (1.0, 2.3)</td>
</tr>
<tr>
<td>Stratified by smoking status in pack-years**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>23</td>
<td>62</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>58</td>
<td>167</td>
<td>0.94 (0.53, 1.6)</td>
<td>0.89 (0.50, 1.6)</td>
</tr>
<tr>
<td>C/C</td>
<td>25</td>
<td>72</td>
<td>0.94 (0.48, 1.8)</td>
<td>0.83 (0.42, 1.6)</td>
</tr>
<tr>
<td>&gt;0–20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>9</td>
<td>46</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>47</td>
<td>117</td>
<td>2.1 (0.93, 4.5)</td>
<td>1.7 (0.74, 4.8)</td>
</tr>
<tr>
<td>C/C</td>
<td>28</td>
<td>61</td>
<td>2.3 (1.0, 5.5)</td>
<td>2.3 (0.95, 5.4)</td>
</tr>
<tr>
<td>&gt;20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>29</td>
<td>35</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>59</td>
<td>66</td>
<td>1.1 (0.59, 2.0)</td>
<td>1.2 (0.62, 2.2)</td>
</tr>
<tr>
<td>C/C</td>
<td>41</td>
<td>23</td>
<td>2.2 (1.1, 4.4)</td>
<td>2.2 (1.0, 4.7)</td>
</tr>
<tr>
<td>Rs20580 (Exon 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>91</td>
<td>203</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/C</td>
<td>197</td>
<td>403</td>
<td>1.1 (0.81, 1.5)</td>
<td>1.1 (0.83, 1.6)</td>
</tr>
<tr>
<td>A/A</td>
<td>99</td>
<td>180</td>
<td>1.2 (0.87, 1.7)</td>
<td>1.2 (0.83, 1.7)</td>
</tr>
<tr>
<td>Stratified by smoking status in pack-years**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>34</td>
<td>98</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/C</td>
<td>64</td>
<td>187</td>
<td>0.99 (0.61, 1.6)</td>
<td>0.98 (0.60, 1.6)</td>
</tr>
<tr>
<td>A/A</td>
<td>24</td>
<td>80</td>
<td>0.86 (0.47, 1.6)</td>
<td>0.86 (0.47, 1.6)</td>
</tr>
<tr>
<td>&gt;0–20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>16</td>
<td>67</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/C</td>
<td>50</td>
<td>140</td>
<td>1.4 (0.73, 2.6)</td>
<td>1.3 (0.68, 2.5)</td>
</tr>
<tr>
<td>A/A</td>
<td>26</td>
<td>62</td>
<td>1.5 (0.74, 3.1)</td>
<td>1.5 (0.74, 3.2)</td>
</tr>
<tr>
<td>&gt;20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>41</td>
<td>43</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/C</td>
<td>83</td>
<td>76</td>
<td>1.1 (0.68, 1.9)</td>
<td>1.3 (0.73, 2.3)</td>
</tr>
<tr>
<td>A/A</td>
<td>49</td>
<td>33</td>
<td>1.6 (0.84, 2.9)</td>
<td>1.6 (0.84, 3.1)</td>
</tr>
<tr>
<td>Rs20579 (Exon 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>264</td>
<td>592</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>108</td>
<td>188</td>
<td>1.3 (0.98, 1.7)</td>
<td>1.3 (1.0, 1.8)</td>
</tr>
<tr>
<td>T/T</td>
<td>8</td>
<td>7</td>
<td>2.6 (0.92, 7.1)</td>
<td>2.0 (0.69, 5.8)</td>
</tr>
<tr>
<td>Stratified by smoking status in pack-years**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>80</td>
<td>281</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>35</td>
<td>88</td>
<td>1.4 (0.88, 2.2)</td>
<td>1.5 (0.91, 2.3)</td>
</tr>
<tr>
<td>T/T</td>
<td>3</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>&gt;0–20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>67</td>
<td>193</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>24</td>
<td>67</td>
<td>1.0 (0.60, 1.8)</td>
<td>1.1 (0.61, 1.9)</td>
</tr>
<tr>
<td>T/T</td>
<td>2</td>
<td>4</td>
<td>1.4 (0.26, 8.0)</td>
<td>1.2 (0.21, 6.9)</td>
</tr>
<tr>
<td>&gt;20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>117</td>
<td>118</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>C/T</td>
<td>49</td>
<td>33</td>
<td>1.5 (0.90, 2.5)</td>
<td>1.5 (0.88, 2.6)</td>
</tr>
<tr>
<td>T/T</td>
<td>3</td>
<td>3</td>
<td>1.0 (0.20, 5.1)</td>
<td>0.95 (0.18, 5.0)</td>
</tr>
<tr>
<td>Rs439132 (Intron 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>297</td>
<td>590</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/G</td>
<td>23</td>
<td>54</td>
<td>0.85 (0.51, 1.4)</td>
<td>0.78 (0.45, 1.4)</td>
</tr>
<tr>
<td>G/G</td>
<td>8</td>
<td>2</td>
<td>7.9 (1.7, 38)</td>
<td>5.9 (1.1, 31)</td>
</tr>
<tr>
<td>Stratified by smoking status in pack-years**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>100</td>
<td>288</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/G</td>
<td>5</td>
<td>23</td>
<td>0.63 (0.23, 1.7)</td>
<td>0.58 (0.21, 1.6)</td>
</tr>
<tr>
<td>G/G</td>
<td>1</td>
<td>1</td>
<td>2.9 (0.18, 46)</td>
<td>2.7 (0.16, 45)</td>
</tr>
<tr>
<td>&gt;0–20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>70</td>
<td>185</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/G</td>
<td>3</td>
<td>24</td>
<td>0.33 (0.096, 1.1)</td>
<td>0.32 (0.092, 1.1)</td>
</tr>
<tr>
<td>G/G</td>
<td>1</td>
<td>1</td>
<td>2.6 (0.16, 43)</td>
<td>3.3 (0.18, 60)</td>
</tr>
<tr>
<td>&gt;20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>127</td>
<td>117</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>A/G</td>
<td>15</td>
<td>7</td>
<td>2.0 (0.78, 5.0)</td>
<td>2.0 (0.74, 5.7)</td>
</tr>
<tr>
<td>G/G</td>
<td>6</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*Adjusted for gender, age, education, ethnicity and tobacco smoking (pack-years).**Adjusted for gender, age, education and ethnicity.
smoking, there was little or no association between the age, sex, race/ethnicity, education and pack-years of tobacco use for the genotyping frequencies among the controls. Adjusting for smoking, the OR was 1.3 (95% CL 0.97, 1.8) for the CATA haplotype, 1.2 (1.0, 1.9) for the CCAG haplotype compared to the most common haplotype (T-C-C-A), the adjusted OR was 1.2 (95% CL 0.95, 1.5) for the CACA haplotype and 1.4 (1.0, 1.9) for the CATA haplotype. In addition, the association with the CATA haplotype was more evident among moderate smokers (>0–20 pack-years) with the adjusted ORs of 1.4 (0.99, 2.1) and 6.3 (2.4, 17), respectively (Table VI). The test for interaction between haplotypes and pack-years yielded a p value of 0.99.

For UADT cancers, when compared to the most common haplotype (rs20581-rs20580-rs20579-rs439132 = T-C-C-A), the adjusted OR was 1.2 (95% CL 0.95, 1.5) for the CACA haplotype, 1.4 (1.0, 1.9) for the CATA haplotype and 1.8 (1.1, 2.8) for the CCCG haplotype. In addition, the association with the CATA haplotype was more evident among moderate smokers (>0–20 pack-years) with the adjusted ORs of 1.4 (0.99, 2.1) and 6.3 (2.4, 17), respectively (Table VI). The test for interaction yielded a p value of 0.08.

### Discussion

DNA ligase I (LIG1), located on chromosome 19, is a 102-kDa nuclear enzyme that appears to be involved in joining Okazaki fragments during DNA replication and in sealing single-strand breaks in the base excision repair pathway. Bentley et al. suggested that LIG1 null cells have normal DNA repair capacity, but also have increased genome instability, possibly affecting cancer susceptibility. It is important to identify the functional significance of the LIG1 polymorphism to interpret its potential effects on cancer development in the population. Bentley et al. hypothesized that the function of LIG1 may be compensated by another ligase gene. This hypothesis warrants further investigation.
DNA repair plays an essential role by protecting cells' chromosomes from environmental exposure, such as tobacco smoke. Among the different repair pathways, base excision repair removes small lesions, DNA single strand breaks and nonbulky adducts. An observed association between increased risk of UADT cancer and higher levels of BPEE-induced DNA adducts in peripheral lymphocytes may reflect a combination of carcinogen metabolism and DNA repair. An increasing number of publications imply the association between reduced DNA repair capacity and altered cancer risk and the association between DNA repair genetic variations and DNA repair function and the consequently modified cancer risk.

Only weak associations with the haplotypes of LIG1 for lung cancer were observed in this study. For a few of the associations, the data showed more pronounced patterns among moderate and/or heavy smokers, consistent with our expectation to see stronger genotype associations among moderate/heavy smokers. In addition, interactions between smoking and the polymorphisms on the development of the above-mentioned cancers were implied, based on the stratified analyses. However, we had limited power to detect the extent of the potential interactions.

A previous study on the LIG1 polymorphism on exon 7 (rs20580) and lung cancer detected no association between the LIG1 polymorphism and the risk of lung cancer in a case-control study that included non-Hispanic whites (OR = 1.0, 95% CL 0.8–1.4 for the AC vs. CC genotype; OR = 1.03, 95% CL 0.6–1.6 for the AA vs. CC genotype) after adjusting for age, sex, smoking, alcohol use and family history of cancer. In our study, we also investigated ethnic groups other than Caucasians. However, we did not have enough subjects to detect interactions with ethnicity, upon examining the combined associations of these 4 LIG1 polymorphisms and tobacco smoking with lung and UADT cancers.

Nonetheless, when we considered all 4 of the SNPs together as haplotypes, we observed associations between the LIG1 haplotypes and the cancer risks. Our findings showed the possibility that these 4 LIG1 polymorphisms and haplotypes may play a role in the susceptibility to the cancers of interest. Tobacco smoking results in the formation of adducts on DNA. The main pathway for repairing this type of DNA damage is NER. Since LIG1 is involved in the NER and BER pathways, there is potential for interaction between these SNPs and tobacco smoking. The associations of the LIG1 polymorphisms and haplotypes with lung and head and neck cancers among moderate/heavy smokers may suggest carcinogenesis mechanisms involving exposure to tobacco smoke.

Another possible explanation for the observed association is that the examined SNPs could be in linkage disequilibrium with other functional SNPs that were not genotyped. Functional studies on the LIG1 gene are warranted to test these hypotheses. Given the limited precision for interaction assessment in single studies, further studies of these effects are needed and study samples will have to be carefully pooled to obtain an accurate picture of gene-environment and gene–gene interactions. Furthermore, although diplotypes represent the chromosome pairs for a particular individual, and thus are more informative to assessing individual risk, we do not have enough power to assess diplotypes due to the limited sample size. Studies with larger sample sizes are warranted for further investigation on diplotypes.

A major strength of this study was that the study design was population-based. The controls were selected from the same population that gave rise to the cases based on a neighborhood-match algorithm; therefore, it reduced an important potential source of selection bias.

Despite its strengths, this study also has a number of limitations. There might be possible selection bias introduced by nonparticipation of eligible persons, due to premature death, refusal or inability to contact. Lung cancer cases with poor prognoses may result in a rapid progression to an advanced stage, precluding interviews and participation in the study and resulting in possible selection bias. Nonetheless, the active surveillance procedure, which was used to identify cases shortly after diagnosis, reduces this problem. Recall bias may occur due to the nature in which information is collected in a case-control study. Compared to controls, cases might differentially recall exposures of tobacco smoking or alcohol consumption because of their cancer diagnosis. Reporting bias might occur if certain exposures are under reported or over reported, especially if behaviors or habits, such as tobacco smoking, carry a social stigma in the population. On the other hand, we expect minimal misclassification of the DNA repair genetic polymorphisms. Standard laboratory protocols were used to genotype the markers, while the laboratory personnel were blinded to the case-control status of the samples.

Additionally, our sample size, although modest, provided limited power to detect associations in subgroups, especially for the rare variants. We also assessed the false positive reporting probabilities (FPRPs) according to Wacholder et al. With a prior of 25% for UADT cancer, the FPRP is 27% for Rs20581 (Exon 25) C/C vs. T/T; 24% for Rs20581 (Exon 25) C/C vs. T/T among heavy smokers; and 18% for Rs439132 (Intron 1) G/G vs. A/A. Thus, the probability of false positives among the reported associations is high.

In summary, we observed weak-to-moderate associations of the LIG1 polymorphisms (on exon 2 and on intron 1) with lung cancer and of the LIG1 polymorphisms (on exon 25, on exon 2 and on intron 1) with UADT cancers. Despite limitations of possible bias and imprecise estimation for rare genotypes and haplotypes, our findings suggest that LIG1 polymorphisms and haplotypes may affect the risk of lung and UADT cancers. Although observing such effects may help us better understand or overreport, especially if behaviors or habits, such as tobacco smoking, increases lung cancer risk, further research is needed to examine gene-smoking and gene–gene interactions in large populations.

Acknowledgements

The authors thank Mr. John Dermand for supervising the administration of questionnaires and the collection of buccal smears. The authors thank Ms. Taylor McAfee for extracting DNA from buccal cells, and Dr. Yan Cui for managing the data files. The authors are also indebted to the study participants for their dedication and commitment.

References

10. Huang WY, Olshan AF, Schwartz SM, Berndt SI, Chen C, Llaca V, Chanock SJ, Fraumeni JF, Jr, Hayes RB. Selected genetic polymorphisms in MGMT, XRC1C, XPD, and XRC5 and risk of head and
1638 LEE ET AL.
