A Functional Variant at 19q13.3, rs967591G>A, Is Associated with Shorter Survival of Early-Stage Lung Cancer

Hyo-Sung Jeon1,2, Guang Jin6, Hyo-Gyoung Kang2, Yi Young Choi2, Won Kee Lee3, Jin Eun Choi2, Eun Young Bae2, Seung Soo Yoo5, Shin Yup Lee6, Eung Bae Lee5, Young Tae Kim7, Jaehee Lee6, Seung-Ick Cha4, Chang Ho Kim4, Sanghoon Jheon7, In San Kim2, and Jae Yong Park1,2,4

Abstract

Purpose: This study was conducted to investigate the associations between single-nucleotide polymorphisms (SNP) in 19q13.3 and survival of patients with early-stage non–small cell lung cancer (NSCLC), and to define the causative functional SNP of the association.

Experimental Design: A two-stage study design was used to evaluate five SNPs in relation to survival outcomes in 328 patients and then to validate the results in an independent patient population (n = 483). Luciferase assay and real-time PCR were conducted to examine functional relevance of a potentially functional SNP.

Results: Of the five SNPs, three SNPs (rs105165C>T, rs967591G>A, and rs735482A>C) were significantly associated with survival outcomes in a stage I study. The rs967591A allele had significantly higher activity of the CD3EAP promoter compared with the rs967591G allele (P = 0.002), but the SNP did not have an effect on the activity of PPP1R13L promoter. The rs967591G>A was associated with the level of CD3EAP mRNA expression in lung tissues (P = 0.01). The rs967591G>A exhibited consistent associations in a stage II study. In combined analysis, the rs967591 AA genotype exhibited a worse overall survival (adjusted HR = 1.69; 95% confidence interval 1.29–2.20; P = 0.0001).

Conclusion: The rs967591G>A affects CD3EAP expression and thus influences survival in early-stage NSCLC. The analysis of the rs967591G>A polymorphism can help identify patients at high risk of a poor disease outcome. Clin Cancer Res; 19(15); 4185–95. © 2013 AACR.

Introduction

Lung cancer is the most common cause of cancer-related death worldwide. More than 80% of all lung cancers are non–small cell lung cancer (NSCLC), with an average 5-year survival rate of 15% (1). The tumor-node-metastasis (TNM) staging system is the best index for determining prognosis of NSCLC (2). However, patients with the same stage of disease display marked variability in survival, indicating the heterogeneity of prognosis within the same population and the inadequacy of the TNM staging system to account fully for this heterogeneity. Recent advances in the molecular biology of NSCLC have led to intensive research to identify molecular markers to predict prognosis for the individual patient. This may help to categorize patients in subgroups to achieve better therapeutic treatment, survival, and quality of life (3, 4).

The chromosomal region 19q13.3 harbors several genes involved in DNA repair, apoptosis, and cell proliferation. Excision repair cross-complementing group 1 (ERCC1) is involved in nucleotide excision repair pathway that eliminates bulky DNA adducts caused by carcinogens in tobacco smoke and platinum-based chemotherapeutic agents (5, 6). Protein phosphatase 1 regulatory (inhibitor) subunit 13 like [PPP1R13L; alias, inhibitor of apoptosis-stimulating protein of p53 (IASPP) or RelA-associated inhibitor (RAI)] was identified as an inhibitor of the p65/RelA subunit of the NF-kB, and shown to modulate NF-kB–related apoptosis (7, 8). The PPP1R13L protein also modulates p53-dependent and p53-independent apoptosis (9, 10). The CD3e molecule, epsilon-associated protein [CD3EAP; alias, antisense to ERCC1 (ASE1)] encodes a nucleoprotein and is positioned in an antisense orientation to, and overlaps with, ERCC1. Although the biologic function of the CD3EAP is unclear, the protein localizes in the fibrillar centers of the nucleolus and...
may be a member of the RNA polymerase I transcription complex that synthesize rRNA precursors, thus implicating CD3EAP in cell proliferation (11).

It has been reported that haplotypes of 3 single-nucleotide polymorphisms (SNP), rs1970764C (PPP1R13L IVS8-1435), rs967591G>A (-8358 from PPP1R13L), or -21 from CD3EAP, and rs11615G>A (ERCC1 N118N), are associated with the risk of basal cell carcinoma, breast cancer, and lung cancer (12–14) suggesting that genetic variation(s) in the 19q13.3 region may affect the susceptibility to cancer. In addition, SNPs in the 19q13.3 region, particularly rs735482A>C (ERCC1 ‘931 (the nucleotide 3’ of the translation termination codon denoted by ‘)’, or CD3EAP K259T] and rs11615G>A have been studied in relation to clinical outcomes of patients treated with platinum-based regimens (15–17) because the low/negative expression of ERCC1 in tissue is associated with higher objective response to the platinum-based chemotherapy and better survival in patients with cancer (18). However, the results of published studies regarding the predictive role of the ERCC1 SNPs in patients with cancer are inconsistent (17, 19), and the functional SNP(s) in the 19q13.3 region remains unknown.

In the present study, we evaluated the effects of SNPs in the 19q13.3 region on the survival of patients with surgically resected early-stage NSCLC. In addition, we examined the functional relevance of the SNPs to discern which was the causative functional SNP for the associations.

Materials and Methods

Study design and patients

A two-stage study design was used to evaluate SNPs in the 19q13.3 region in relation to prognosis of NSCLC and then to validate promising associations in a second independent patient population. Stage I of the study included 328 patients with pathologic stages I, II, or IIIA (microinvasive N2) NSCLC who underwent curative surgical resection at the Kyungpook National University Hospital (Daegu, Korea) between September 1998 and December 2006. The details of this study population are described elsewhere (20). The stage II study population included 483 patients with pathologic stage I, II, or IIIA who underwent surgical resection at Seoul National University Hospital between December 1997 and October 2008. This study was approved by the Institutional Review Boards of the Kyungpook National University Hospital and Seoul National University Hospital (Seoul, Korea).

SNPs and genotyping

To select all the potentially functional variants of PPP1R13L, CD3EAP, and ERCC1 genes, we used the public database (http://www.ncbi.nlm.nih.gov/SNP) to search for candidate variants in the promoter region, all exons including intron–exon boundaries and the 3’-untranslated region (3’-UTR). Five SNPs [rs1005165C (–7747 from PPP1R13L; or –905 from CD3EAP), rs967591G>A, rs735482A>C, rs3212986G>T (ERCC1 ‘197; or CD3EAP Q504K), and rs11615G>A] were selected and evaluated in stage I of the study. The nucleotide substitutions and SNP identification numbers of the SNPs are shown in Fig. 1. These SNPs were genotyped using a PCR restriction length polymorphism assay.

Promoter-luciferase constructs and luciferase assay

We investigated whether rs967591 modulates the activity of the promoter of CD3EAP or PPP1R13L genes using a luciferase assay. The promoter fragments of CD3EAP and PPP1R13L, including the rs967591, were synthesized by PCR. Each PCR primer was shown in Supplementary Table S1. The PCR products were cloned into the NheI/BglII site of the pGL3-basic plasmid (Promega). The correct sequence of all the clones was verified by DNA sequencing. The NSCLC cell lines, H1299 and A549, were transfected with each report construct, and pRL-SV40 vector (Promega) using Effectene transfection reagent (Qiagen). The cells were collected 48 hours after transfection and the cell lysates were prepared according to Promega’s instruction manual. The luciferase activity was measured using a Lumat LB953 luminometer (EG & G Berthhold), and the results were normalized by using the activity of Renilla luciferase. All experiments were carried out in triplicate.

To test whether pharmacologic manipulation of estrogen receptor (ER) signaling would influence the results of the luciferase assay in an allele-specific manner, we conducted promoter assay with/without estrogen or estrogen inhibitor. At 48 hours after transfection with CD3EAP promoter plasmid for luciferase assay, the H1299 cells were treated with ethanol (EtOH-vehicle control), 100 nmol/L 17b-estradiol (E2; Sigma-Aldrich), or 100 nmol/L antiestrogen 4-hydroxytamoxifen (4-OHT; Sigma-Aldrich) for 1 hour. The luciferase activity was measured in each cell lysate.

Translational Relevance

There is a critical need for biomarkers for predicting prognosis after lung cancer surgery; however, reliable biomarkers useful in the clinical setting are still scarce. The 19q13.3 region harbors several genes involved in DNA repair, apoptosis, and cell proliferation, including ERCC1, PPP1R13L, and CD3EAP (also known as antisense to ERCC1, ASE). We investigated whether single-nucleotide polymorphisms (SNP) in the 19q13.3 are prognostic factors in early-stage non–small cell lung cancer (NSCLC). The rs967591G>A in the CD3EAP gene was found to be an independent prognostic marker for patients with surgically resected early-stage NSCLC. Functionally, the SNP was associated with CD3EAP expression. In addition to the pathologic stage, the rs967591G>A could be used to identify patients at high risk of relapse after surgery, thereby helping to select subgroups of patients for adjuvant therapy to potentially improve survival.
RNA preparation and quantitative reverse transcription-PCR

CD3EAP, PPP1R13L, and ERCC1 mRNA expression was examined by quantitative reverse transcription-PCR (qRT-PCR). Total RNA from tumor and paired nonmalignant lung tissues (n = 118 for CD3EAP and ERCC1, and 102 for PPP1R13L) was isolated using TRIzol (Invitrogen). Real-time PCR with SYBR Green detection was conducted using a LightCycler 480 (Roche Applied Science) with QuantiFast SYBR Green PCR Master Mix (Qiagen). The real-time PCR primers for CD3EAP, PPP1R13L, ERCC1, and β-actin gene were listed in Supplementary Table S1. Each sample was run in duplicate. The relative CD3EAP, PPP1R13L, and ERCC1 mRNA expression were normalized with β-actin expression and then calculated by the 2−ΔΔCT method (21).

CD3EAP overexpression

The full-length CD3EAP gene was amplified by RT-PCR with primers (Supplementary Table S1) and the PCR product was inserted into the p3XFlag-pCMV10 vector (Sigma-Aldrich). To identify that CD3EAP represses ERCC1 expression, we transfected p3Xflag-pCMV10–containing CD3EAP to H1299 cells. At 48 hours after transfection, expression of ERCC1 and CD3EAP were measured by real-time PCR and Western blotting.

Analysis of copy number and methylation status of CD3EAP

To investigate the mechanism of CD3EAP overexpression in tumor tissues, we determined copy number and CD3EAP methylation status in 44 NSCLCs. Copy number of the CD3EAP in lung tumor tissues was determined by real-time quantitative PCR (qPCR) using SYBR Green PCR Master Mix (Qiagen). Briefly, we used LINE-1 gene as a reference gene for copy number analysis. Relative copy number of CD3EAP gene was determined by comparing the ratio of CD3EAP with LINE-1 with the ratio of that in eight normal human blood genomic DNA, which was used as a diploid control. Methylation status of the CD3EAP promoter was analyzed by methylation-specific sequencing in both nonmalignant lung tissues and tumor tissues. The bisulfite-modified CD3EAP promoter was amplified with specific primers (Supplementary Table S1) and methylation status was determined by sequencing.

Statistical analyses

Differences in the distribution of genotypes according to the clinicopathologic factors of the patients were compared using χ² tests. The linkage disequilibrium (LD) status among SNPs was measured by using the program HaploView. The haplotype frequencies were estimated on the basis of a...
Bayesian algorithm using the phase program (22). LD blocks were inferred from the definition proposed by Gabriel and colleagues (23). Overall survival (OS) was measured from the day of surgery until the date of death or to the date of the last follow-up. The association of OS with genotypes was investigated using the Kaplan–Meier method and assessed using the log-rank test. HR and 95% confidence intervals (CI) were estimated using multivariate Cox proportional hazards models, with adjustment for age, gender, smoking status, pathologic stage, and adjuvant therapy. The issue of multiple tests was controlled using Bonferroni correction. All statistical testing was conducted with Statistical Analysis System for Windows, version 9.2 (SAS Institute).

Results

Patient characteristics and clinical predictors

The clinical and pathologic characteristics of the patients of the stage I and stage II studies and the association with OS are shown in Table 1. The pathologic stage was significantly associated with OS [log-rank $P (P_{\text{log-rank}}) < 0.001$] in both the stage I and stage II studies. Age and gender were also associated with OS in the stage II study ($P_{\text{log-rank}}$ for OS = 0.001 and 0.01, respectively). According to the estimated HRs (data not shown), the associations of survival with age and gender were in the same direction in the two stages, although there was a slightly larger difference of survival between subgroups by age and gender in the stage II. Therefore, the significant association only in stage II may be due to the statistical power of larger stage II population.

Genotype frequencies and effect on OS

The genotype distributions of all the five SNPs evaluated were in Hardy–Weinberg equilibrium. None of the five SNPs were significantly associated with patient- or tumor-related factors, such as age, gender, smoking status,
histologic subtype, pathologic stage, and adjuvant therapy (data not shown).

Of the five SNPs studied, three SNPs (rs1005165C>T, rs967591G>A, and rs735482A>C) were significantly associated with [under a recessive model for the variant allele; log-rank \( P \) for OS = 0.005, 0.003, and 0.01, respectively], but 2 SNPs (rs3212986G>T and rs11615G>A) were not associated with OS (Table 2; Supplementary Fig. S1).

### Table 2. Genotype frequencies and log-rank \( P \) value for the polymorphisms analyzed in patients with NSCLC

<table>
<thead>
<tr>
<th>ID no.</th>
<th>Base change</th>
<th>MAF</th>
<th>HWE ( P )</th>
<th>General</th>
<th>Dominant</th>
<th>Recessive</th>
<th>Asian*</th>
<th>European*</th>
<th>Af-Am*</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1005165</td>
<td>C&gt;T</td>
<td>0.502</td>
<td>0.134</td>
<td>0.02a</td>
<td>0.66</td>
<td>0.005a</td>
<td>0.513</td>
<td>0.143</td>
<td>–</td>
</tr>
<tr>
<td>rs967591</td>
<td>G&gt;A</td>
<td>0.486</td>
<td>0.313</td>
<td>0.01b</td>
<td>0.45</td>
<td>0.003f</td>
<td>–</td>
<td>0.130</td>
<td>0.150</td>
</tr>
<tr>
<td>rs735482</td>
<td>A&gt;C</td>
<td>0.477</td>
<td>0.942</td>
<td>0.02c</td>
<td>0.20</td>
<td>0.01d</td>
<td>0.478</td>
<td>0.075</td>
<td>0.278</td>
</tr>
<tr>
<td>rs3212986</td>
<td>G&gt;T</td>
<td>0.265</td>
<td>0.903</td>
<td>0.29</td>
<td>0.11</td>
<td>0.61</td>
<td>0.221</td>
<td>0.235</td>
<td>–</td>
</tr>
<tr>
<td>rs11615</td>
<td>G&gt;A</td>
<td>0.229</td>
<td>0.284</td>
<td>0.91</td>
<td>0.66</td>
<td>0.86</td>
<td>0.208</td>
<td>0.646</td>
<td>0.087</td>
</tr>
</tbody>
</table>

NOTE: Corrected \( P \) values by Bonferroni correction for 15 multiple tests, *0.30; †0.15; ‡0.075; §0.045; ¶0.15. Abbreviations: Af-Am, African-American; HWE \( P \), \( P \) for Hardy–Weinberg equilibrium test; MAF, minor allele frequency.

aThe minor allele refers to the alternate allele in National Center for Biotechnology Information (NCBI) SNP database. Information about polymorphisms and IDs and MAF in other ethnic populations (Asian, European, and African-American) were obtained from NCBI database (http://www.ncbi.nlm.nih.gov). rs1005165 C>T: -7474 from PPP1R13L or -905 from CD3EAP (translation start site denoted as +1); rs967591 G>A: -8358 from PPP1R13L or -22 from CD3EAP (translation start site denoted as +1); rs735482 A>C: ERCC1/C3931 (the nucleotide 30 of the translation termination codon was denoted by/C3931) or CD3EAP K259T; rs3212986 G>T: ERCC1/C3197; or CD3EAP Q504K; and rs11615 G>A: ERCC1 N118N.

### Table 3. OS according to haplotypes and diplotypes of polymorphisms analyzed in patients with NSCLC

<table>
<thead>
<tr>
<th>Polymorphism/genotype</th>
<th>No. of patients</th>
<th>No. of deaths (%)b</th>
<th>5Y-OSR (%)c</th>
<th>Log-rank ( P )</th>
<th>HR (95% CI)c</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype of rs1005165C&gt;T, rs967591G&gt;A, and rs735482A&gt;C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGA</td>
<td>320</td>
<td>115 (35.9)</td>
<td>60.7</td>
<td>0.03</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>TAC</td>
<td>301</td>
<td>135 (44.9)</td>
<td>49.1</td>
<td>1.32 (1.03–1.70)</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Othersd</td>
<td>35</td>
<td>10 (28.6)</td>
<td>66.9</td>
<td>0.81 (0.42–1.55)</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>Diplotype of rs1005165 C&gt;T, rs967591G&gt;A, and rs735482A&gt;C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others’/others</td>
<td>96</td>
<td>31 (32.3)</td>
<td>61.1</td>
<td>0.01</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>TAC/others</td>
<td>163</td>
<td>63 (38.7)</td>
<td>61.3</td>
<td>0.90 (0.67–1.21)</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>TAC/TAC</td>
<td>69</td>
<td>36 (52.2)</td>
<td>33.5</td>
<td>1.70 (1.18–2.44)</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>Others/others + TAC/others</td>
<td>259</td>
<td>94 (36.3)</td>
<td>61.3</td>
<td>0.003</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>TAC/TAC</td>
<td>69</td>
<td>36 (52.2)</td>
<td>33.5</td>
<td>1.91 (1.29–2.83)</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Haplotype of rs3212986G&gt;T and rs11615G&gt;A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>330</td>
<td>134 (40.6)</td>
<td>54.4</td>
<td>0.34</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>152</td>
<td>64 (42.1)</td>
<td>51.2</td>
<td>0.95 (0.71–1.29)</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>TG</td>
<td>174</td>
<td>62 (35.6)</td>
<td>61.7</td>
<td>0.81 (0.60–1.10)</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>GG + GA</td>
<td>482</td>
<td>198 (41.1)</td>
<td>53.3</td>
<td>0.14</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>TG</td>
<td>174</td>
<td>62 (35.6)</td>
<td>61.7</td>
<td>0.82 (0.62–1.09)</td>
<td>0.18</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: 5Y-OSR, 5-year OS rate.

aRow percentage.

bFive-year survival rate, proportion of survival derived from Kaplan–Meier analysis.

cHRs, 95% CIs, and their corresponding \( P \) values were calculated using multivariate Cox proportional hazard models, adjusted for age, gender, smoking status, tumor histology, pathologic stage, and adjuvant therapy.

dHaplotypes that had a frequency of less than 5%.

eAny haplotype other than the TAC haplotype.

Of the five SNPs studied, three SNPs (rs1005165C>T, rs967591G>A, and rs735482A>C) were significantly associated with [under a recessive model for the variant allele; log-rank \( P \) for OS = 0.005, 0.003, and 0.01, respectively], but 2 SNPs (rs3212986G>T and rs11615G>A) were not associated with OS (Table 2; Supplementary Fig. S1).
Figure 2. Effect of the rs967591G>A on transcription activities of CD3EAP and PPP1R13L and mRNA expression of CD3EAP, PPP1R13L, and ERCC1. A, the transcription activity was measured using the Dual Luciferase Reporter Assay System with pGL3-CD3EAP and pGL3-PPP1R13L constructs. The
the 5 SNPs, the rs1005165C>T, rs967591G>A, rs735482A>C, rs321298G>C, and rs11615G>A were in strong LD, respectively (Fig. 1). Thus, we examined the associations of the haplotypes of the rs1005165C>T, rs967591G>A, and rs735482A>C and the haplotypes of the rs321298G>C and rs11615G>A with survival outcomes. The inferred haplotypes and their associations with OS are shown in Table 3. Consistent with the results of individual genotype analyses, the rs1005165T-rs967591A-rs735482C haplotype carrying variant alleles at the three loci was associated with a significantly worse OS compared with the rs1005165C-rs967591G-rs735482A carrying wild-type alleles at the three loci (adjusted HR (aHR) for OS was 1.32; 95% CI = 1.03–1.70; \( P = 0.03 \)). In addition, patients with homozygous rs1005165T-rs967591A-rs735482C haplotype had a significantly worse OS compared with those carrying one or none of the rs1005165T-rs967591A-rs735482C haplotype (aHR for OS was 1.91; 95% CI = 1.29–2.83; \( P = 0.001 \)).

**Effect of the rs967591G>A on the activity of CD3EAP and PPP1R13L promoter**

As a consequence of the strong LD among the rs1005165C>T, rs967591G>A, and rs735482A>C, it was difficult to determine which of the three SNPs was more likely to have a functional effect on the disease association. In an attempt to resolve this problem, we first used the Alibabaprogram and PolyPhen computational programs for prediction of functional significance of the three SNPs (24–26). Analysis of the potential transcription factor–binding sites by the Alibabaprogram (24) showed that the rs967591G to A change leads to the creation of an ER-binding site. In addition, we inferred the functional relevance of the non-synonymous rs735482A>C (CD3EAP Q504K) using the PolyPhen algorithm (25, 26). The PolyPhen analysis showed that the Q to K change may possibly be benign. These results suggest that the rs967591G>A may be functional. Thus, we evaluated the effect of the rs967591G>A on the activity of the promoter of CD3EAP and PPP1R13L genes by a luciferase assay. The rs967591A allele had significantly higher activity of CD3EAP promoter compared with the rs967591G allele in H1299 and A549 cell lines (\( P = 0.002 \) and \( P = 0.003 \), respectively). However, the activity of PPP1R13L promoter was not significantly altered by the rs967591G>A in either of the two cell lines (Fig. 2A).

In addition, we tested whether E2 stimulation or 4-OHT inhibition would influence the activity of CD3EAP promoter in an allele-specific manner. The activity of CD3EAP promoter was not significantly changed by treatment with E2 or 4-OHT in either of the two alleles of rs967591G>A (Supplementary Fig. S2).

**Effect of the rs967591G>A on CD3EAP mRNA expression**

To determine whether the rs967591G>A genotypes were correlated with CD3EAP expression, we evaluated the mRNA level in 118 cases (genotype distribution: 27 GG, 51 GA, and 40 AA). As shown in Fig. 2B, the expression level of CD3EAP mRNA was significantly higher in tumor tissues than in nonmalignant lung tissues \( (P < 0.001) \). In agreement with the results of the promoter assay, the CD3EAP mRNA level was significantly higher in the rs967591 AA genotype than in the rs967591 GG or GA genotype in the nonmalignant lung tissues \( (P = 0.01, \text{Fig. 2C}) \). However, the mRNA level in tumor tissues was not significantly correlated with the genotypes although a similar trend was observed \( (P = 0.31) \). The failure to observe a significant association between the genotypes and mRNA expression level in tumor tissues may have been due to the difference in the proportion of tumor cells in macroscopically isolated samples, which contained both tumor and nonmalignant lung tissues.

**Effect of the rs967591G>A on ERCC1 and PPP1R13L mRNA expression**

Because CD3EAP is positioned in an antisense orientation to ERCC1, it is likely that CD3EAP mRNA acts as a repressor of ERCC1 expression. Therefore, we evaluated the relationship between CD3EAP and ERCC1 mRNA expressions as well as the relationship of the rs967591G>A genotypes with ERCC1 mRNA expression in the 118 cases. As shown in Fig. 2B, the level of ERCC1 mRNA was significantly higher in tumor tissues than in nonmalignant lung tissues \( (P = 0.03) \). However, there was no significant

---

rs967591A construct exhibited significantly higher transcription activity of CD3EAP than the rs967591G construct in H1299 and A549 cell lines \( (P = 0.002 \) and \( P = 0.003 \), respectively). However, transcription activity of PPP1R13L was not significantly altered by the rs967591G>A. The data represent means ± SEs from 2 independent experiments carried out in quadruplicate. B–E, CD3EAP and ERCC1 mRNA expression levels and the association with the rs967591G>A genotypes (27 GG, 51 GA, and 40 AA) were examined in 118 tumor and paired nonmalignant lung tissues and PPP1R13L mRNA expression level and the association with the rs967591G>A in 102 (23 GG, 45 GA, and 34 AA). The CD3EAP, ERCC1, and PPP1R13L mRNA expression levels were normalized with that of \( \beta \)-actin gene. The horizontal line within the box represents the median value; the upper and lower boundaries of the box represent 75th and 25th percentile, respectively; the upper and lower bars indicate the largest and smallest observed values, respectively. \( P \) values, Student t test. B, CD3EAP and ERCC1 expression levels were significantly upregulated in tumor tissues than nonmalignant lung tissues \( (P < 0.001 \) and \( P = 0.03 \), respectively). However, PPP1R13L expression level was not different between tumor and nonmalignant tissues. C, the level of CD3EAP mRNA expression was significantly higher in the rs967591 AA genotype than the rs967591 GG or GA genotype in nonmalignant lung tissues \( (P = 0.01) \). D and E, however, there was no significant difference in ERCC1 and PPP1R13L mRNA expression in either tumor or nonmalignant lung tissues according to the rs967591G>A genotypes. F, investigation of ERCC1 expression by CD3EAP overexpression in H1299 cells. Wild-type CD3EAP and vector constructs were transfected into H1299 cells. ERCC1 mRNA expression was normalized with that of \( \beta \)-actin gene in qRT-PCR analysis. The overexpression of CD3EAP was verified by Western blotting with FLAG antibody (Sigma-Aldrich), ERCC1 and \( \beta \)-actin antibodies (Santa Cruz Biotechnology) were used for ERCC1 protein level detection and internal control, respectively.
difference of ERCC1 mRNA expression levels by the rs967591G>A in either nonmalignant lung tissues or tumor tissues (Fig. 2D). To confirm these results, we overexpressed CD3EAP and measured the change of ERCC1 expression. ERCC1 expression was not changed by CD3EAP overexpression (Fig. 2F). These results suggest that the rs967591G>A SNP affect CD3EAP expression but not ERCC1 expression. In addition, because rs967591G>A is located in the PPP1R13L promoter region, we determined the relationship between rs967591G>A and PPP1R13L mRNA expression in 102 cases. As shown in Fig. 2E, there was no significant correlation between rs967591G>A and PPP1R13L mRNA expression in either nonmalignant lung tissues or tumor tissues (P = 0.24 and 0.91, respectively). These results suggest that the rs967591G>A SNP affects CD3EAP expression, but neither PPP1R13L nor ERCC1 expression.

The association of CD3EAP overexpression with gene copy number and methylation status

To investigate the mechanism of CD3EAP overexpression in tumor tissues, we first analyzed copy number of CD3EAP. No significant correlation was found between CD3EAP expression and gene copy number (Supplementary Fig S3). In addition, we determined the methylation status of CD3EAP gene to test whether demethylation of the CpG islands in CD3EAP promoter led to CD3EAP overexpression in lung tumor tissues. However, there were no methylated CpG islands in CD3EAP promoter in either tumor tissues or nonmalignant lung tissues (data not shown). Therefore, increased copy number and demethylation of CD3EAP seemed not to be associated with CD3EAP overexpression in NSCLC.

Validation study for the association between the rs967591G>A and survival outcomes

The 3 SNPs, the rs1005165C>T, rs967591G>A, and rs735482A>C, significantly associated with the OS in the stage I study, were in strong LD (Fig. 1). Of the three SNPs, the rs967591G>A was predicted to be the putative functional polymorphism via in silico analysis and thus selected for a replication study using an independent sample to confirm the observed association of the SNP with survival outcomes. In agreement with the results of the stage I study, the rs967591G>A was significantly associated with OS (Supplementary Fig. S1F). In addition, there was no evidence of heterogeneity in HRs between the two studies (under a recessive model for the variant allele; \( P_{\text{heterogeneity}} = 1.00 \); Table 4). In combined analysis of the two stages of the study, the rs967591 AA genotype exhibited a worse OS than the rs967591 GG or GA genotype (aHR for OS = 1.69; 95% CI = 1.29–2.20; \( P = 0.0001 \); Table 4 and Fig. 3).

Discussion

We evaluated the effect of SNPs in the 19q13.3 region on the prognosis of early-stage NSCLC in a large two-stage study, including 811 patients. Three SNPs (rs105165C>T;
rs967591G>A, and rs735482A>C) and their haplotypes were significantly associated with OS in patients with surgically resected NSCLC. In addition, this study provides evidence that the rs967591G>A is the causative functional SNP for the associations: the rs967591A allele had increased activity of the promoter and expression of CD3EAP gene. These findings are novel and suggest that in addition to the pathologic stage, testing for the presence of the rs967591G>A may help identify patient subgroups at high risk for poor disease outcome, thereby helping to refine therapeutic decisions in the treatment of NSCLC.

The design of two independent cohorts for the discovery and validation sets was a major strength, which would largely reduce a false-positive finding from the genetic association study (27, 28). In the present study, there was no statically difference of the association on OS between the two cohorts except age and gender. In addition to statistical power, the effect of age on survival in the stage II cohort may be caused by compounding factors such as socioeconomic factor. The rs967591G>A were associated with OS in two independent cohorts, providing credibility that this genetic variant influences survival outcomes of patients with early-stage NSCLC. In addition, the observed P value was compatible with the P value (10^-5), a more stringent level of statistical significance for candidate–gene studies that would avoid most of the false-positive associations arising from multiple comparisons (27). Taken as a whole, these results strengthen the reliability of our finding of an association between the rs967591G>A and prognosis of patients with NSCLC.

Another important finding of the present study was that the rs967591G>A was the functional causative SNP for the associations with survival outcomes. Because the rs967591G>A is located in the promoter region of CD3EAP and PPP1R13L genes, we investigated whether the rs967591G>A modulates the activity of the promoter of CD3EAP and PPP1R13L genes by an established luciferase assay. The in vitro promoter assay revealed that the rs967591G>A increased the activity of CD3EAP promoter, but not PPP1R13L. Moreover, in agreement with the result of the luciferase assay, the mRNA level in nonmalignant lung tissues was significantly higher in the AA genotype than in the GG or GA genotype. These findings suggest that the inherited rs967591G>A affects CD3EAP expression. However, the activity of CD3EAP promoter was not affected by ER stimulation or inhibition in an allele-specific manner, although the SNP was predicted to create an ER-binding site in the promoter region by computation analysis. Further investigation is needed to elucidate the mechanism of influence of this SNP on CD3EAP expression.

Several studies have shown that ERCC1 expression of lung cancer tissue is related to an objective response to the platinum-based chemotherapy and survival in patients with cancer (18). CD3EAP is positioned in an antisense orientation to ERCC1. It is possible that naturally occurring antisense RNA transcripts negatively regulate sense gene expression by modulating sense RNA transcription, pre-mRNA splicing, and mRNA stability, transport, and translation (29, 30). Therefore, we evaluated whether the rs967591G>A that affects CD3EAP expression modulates ERCC1 expression. CD3EAP expression as well as the rs967591G>A genotypes were not significantly correlated with ERCC1 expression. These findings suggest that the rs967591G>A does not modulate ERCC1 expression.

In the present study, CD3EAP mRNA expression was significantly higher in cancer tissues than in paired nonmalignant lung tissues. In addition, patients with higher production genotype for CD3EAP had poor survival outcomes. Comparable with our findings, it has been reported that the rs967591G>A was associated with an increased risk of lung cancer in Chinese population (31). Taken together, these findings suggest that CD3EAP plays an oncogenic role in lung carcinogenesis. On the basis of our results, it is not likely that somatic changes in the tumors such as increased copy number or demethylation of CD3EAP is the mechanism of CD3EAP overexpression. Therefore, upstream signaling pathway of CD3EAP may affect its overexpression in NSCLC. The biologic mechanism of CD3EAP overexpression in lung cancer remains to be elucidated.

Genetic polymorphisms often show ethnic variation. In the present study, the minor allele frequencies of the rs967591G>A and rs11615G>A were 0.49 and 0.23.
respectively among 328 patients with lung cancer in the stage I study; these were comparable with those (0.45 and 0.22, respectively) among Chinese patients with lung cancer (31). However, on the basis of the NIH Database (http://www.ncbi.nlm.nih.gov/SNP), the minor allele frequencies of the five SNPs examined in the present study were significantly different among Asians, Caucasians, and African-Americans (Table 2). Ethnic variation of the SNPs on 19q13.3 and their haplotypes warrants additional study to clarify the association of the SNPs with survival outcomes in diverse ethnic populations.

In conclusion, the present study shows that the 3 SNPs (rs105165C>T, rs967591G>A, and rs735482A>C) and their haplotypes are associated with survival outcomes of patients with surgically resected NSCLC. In addition, the rs967591G>A is the functional causative SNP for the association. This SNP may be an important prognostic marker for identifying patient subgroups at high risk for poor survival outcome, thereby helping to refine therapeutic decisions in the treatment of NSCLC. However, considering the ethnic variation of the SNPs on 19q13.3 and their haplotypes, further studies are needed to clarify the association between the SNPs on 19q13.3, particularly, the rs967591G>A and prognosis of patients with surgically resected NSCLC in diverse ethnic populations. In addition, future studies on the biologic function of CD3EAP are needed to understand the role of the CD3EAP gene in determining lung cancer prognosis.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Development of methodology: H.-S. Jeon, J.Y. Park
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): H.-S. Jeon, S.Y. Lee, Y.T. Kim, J. Lee, J.Y. Park
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.S. Yoo, E.B. Lee, S.-I. Cha, S. Jheon, J.Y. Park
Study supervision: H.-S. Jeon, Y.T. Kim, J.Y. Park

Grant Support
This study was supported by the R&D program of MKE/KEIT (10040393, development and commercialization of molecular diagnostic technologies for lung cancer through clinical validation). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 28, 2012; revised May 22, 2013; accepted June 1, 2013; published OnlineFirst June 17, 2013.

References


