Clinical Cancer Research

Imaging, Diagnosis, Prognosis

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

Hyo-Sung Jeon[1,2], Guang Jin[6], Hyo-Gyoung Kang[2], Yi Young Choi[2], Won Kee Lee[3], Jin Eun Choi[2], Eun Young Bae[2], Seung Soo Yoo[4], Shin Yup Lee[4], Eung Bae Lee[5], Young Tae Kim[7], Jaehee Lee[4], Seung-Ick Cha[4], Chang Ho Kim[4], Sanghoon Jheon[7], In San Kim[2], and Jae Yong Park[1,2,4]

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-κB, and shown to modulate NF-κB–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

Authors' Affiliations: [1]Lung Cancer Center, Kyungpook National University Medical Center; Departments of [2]Biochemistry and Cell Biology, [3]Statistics Center, [4]Internal Medicine and [5]Thoracic Surgery, School of Medicine, Kyungpook National University, Daegu, Republic of Korea; [6]Cancer Research Center, Yanbian University School of Basic Medicine, Yanji, Jilin Province, China; and [7]Department of Thoracic and Cardiovascular Surgery, Seoul National University School of Medicine, Seoul, Republic of Korea

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

H.-S. Jeon, G. Jin, and H.-G. Kang contributed equally to this work.

Corresponding Author: Jae Yong Park, Lung Cancer Center, Kyungpook National University Medical Center, 807, Hoguk-ro, Buk-gu, Daegu 702-210, Korea. Phone: 82-53-200-2631; Fax: 82-53-200-2027; E-mail: jaeyong@knu.ac.kr

doi: 10.1158/1078-0432.CCR-12-2792

©2013 American Association for Cancer Research.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024

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.

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), rs1970764T>C (*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 *1), 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>T (-7474 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 17β-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.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024

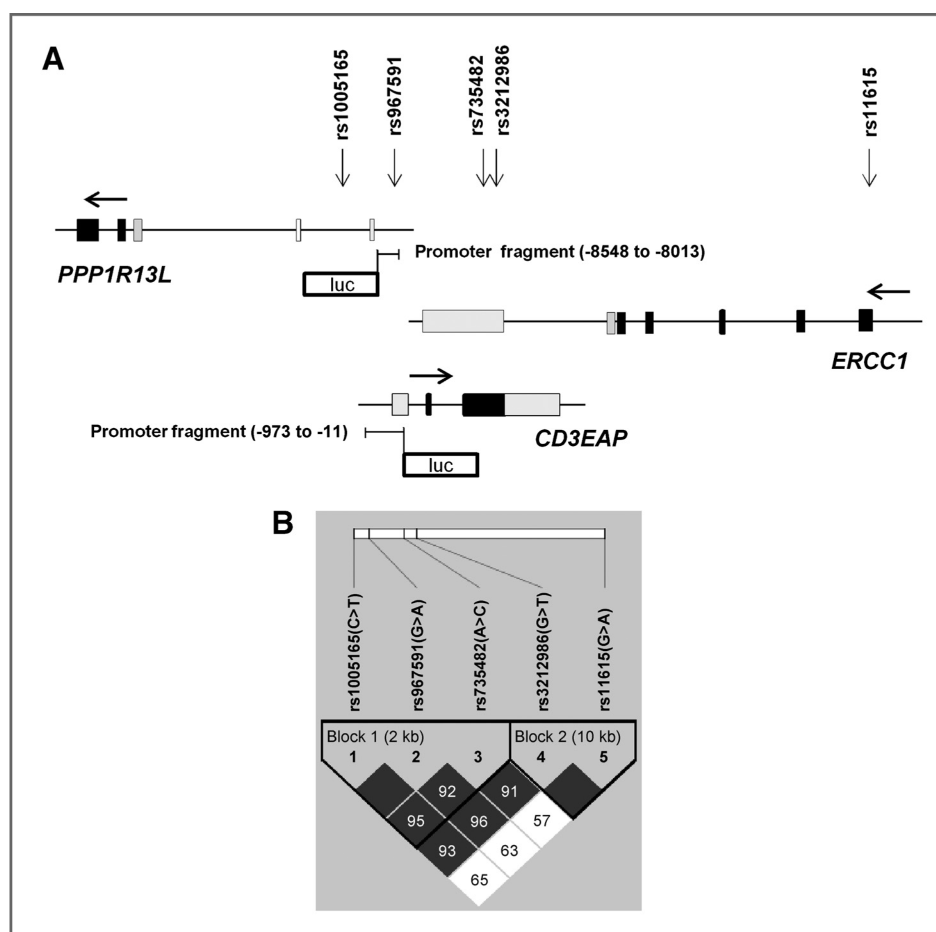

Figure 1. Gene map and LD blocks. A, *PPP1R13L*, *ERCC1*, and *CD3EAP* genes and polymorphisms in the 19q13.3 region. Coding exons are marked by black blocks and 5′- and 3′-UTRs by gray blocks, and arrows indicate the direction of each gene transcription. The first base of translation start site is denoted as +1. The boundaries of the promoter fragments cloned out for the two luciferase assays are indicated under gene map. B, the LD blocks were generated by the Haploview program using the method proposed by Gabriel and colleagues (23). Triangles indicate haplotype blocks. Numbers in squares are |D′| (× 100) values. luc, luciferase.

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 $2^{-\Delta\Delta C_t}$ 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 χ^2 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

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024

Table 1. Univariate analysis for survival by demographics, smoking status, histologic type, pathologic stage, and adjuvant therapy

	Stage I				Stage II			
Variables	No. of cases	No. of deaths (%)[a]	5Y-OSR (%)[b]	Log-rank P	No. of cases	No. of deaths (%)[a]	5Y-OSR (%)[b]	Log-rank P
Overall	328	130 (39.6)	56		483	119 (24.6)	71	
Age, y								
≤64	181	69 (38.1)	58	0.43	257	51 (19.8)	77	0.001
>64	147	61 (41.5)	53		226	68 (30.1)	63	
Gender								
Male	260	111 (42.7)	54	0.17	339	93 (27.4)	67	0.01
Female	68	19 (27.9)	64		144	26 (18.1)	79	
Smoking status								
Never	64	20 (31.3)	64	0.52	187	44 (23.5)	73	0.28
Ever	264	110 (41.7)	54		296	75 (25.3)	69	
Pack-years[c]								
<40	115	44 (38.3)	56	0.15	135	30 (22.2)	71	0.16
≥40	149	66 (44.3)	53		161	45 (28.0)	67	
Histologic type[d]								
Squamous cell carcinoma	200	78 (39.0)	57	0.77	183	43 (23.5)	72	0.63
Adenocarcinoma	122	49 (40.2)	52		277	67 (24.2)	72	
Pathologic stage								
I	190	53 (27.9)	65	3×10^{-5}	295	62 (21.0)	75	2×10^{-5}
II	52	29 (55.8)	46		124	32 (25.8)	73	
IIIA	86	48 (55.8)	43		64	25 (39.1)	47	
Adjuvant therapy[e]								
No	85	45 (53.0)	45	0.70	91	28 (30.8)	69	0.49
Yes	53	32 (60.4)	43		97	29 (29.9)	60	

[a]Row percentage.
[b]Five year-OS rate (5Y-OSR), proportion of survival derived from Kaplan–Meier analysis.
[c]In ever-smokers.
[d]Six large cell carcinomas (stage I) and 23 large cell carcinomas (stage II) were excluded from this analysis.
[e]In pathologic stage II + IIIA: stage I, 50 cases received chemotherapy, 2 cases received radiotherapy, and 1 case received chemotherapy and radiotherapy and stage II, 72 cases received chemotherapy, 9 cases received radiotherapy, and 16 cases received chemotherapy and radiotherapy.

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 15 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 ($P_{L\text{-}R}$) < 0.001] in both the stage I and stage II studies. Age and gender were also associated with OS in the stage II study ($P_{L\text{-}R}$ 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,

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024

Table 2. Genotype frequencies and log-rank *P* value for the polymorphisms analyzed in patients with NSCLC

Polymorphism[a]		Genotype		Log-rank *P* for OS			MAF in healthy populations		
ID no.	Base change	MAF	HWE *P*	General	Dominant	Recessive	Asian[a]	European[a]	Af-Am[a]
rs1005165	C>T	0.502	0.134	0.02[b]	0.66	0.005[e]	0.513	0.143	–
rs967591	G>A	0.486	0.313	0.01[c]	0.45	0.003[f]	–	0.130	0.150
rs735482	A>C	0.477	0.942	0.02[d]	0.20	0.01[g]	0.478	0.075	0.278
rs3212986	G>T	0.265	0.093	0.29	0.11	0.61	0.221	0.235	–
rs11615	G>A	0.229	0.284	0.91	0.66	0.86	0.208	0.646	0.087

NOTE: Corrected *P* values by Bonferroni correction for 15 multiple tests, [b]0.30; [c]0.15; [d]0.30; [e]0.075; [f]0.045; [g]0.15.
Abbreviations: Af-Am, African-American; HWE *P*, *P* for Hardy–Weinberg equilibrium test; MAF, minor allele frequency.
[a]The 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* *931 (the nucleotide 3′ of the translation termination codon was denoted by *1) or *CD3EAP* K259T; rs3212986 G>T: *ERCC1* *197; or *CD3EAP* Q504K; and rs11615 G>A: *ERCC1* N118N.

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 ($P_{L\text{-}R}$) 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). Among

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

Polymorphism/genotype	No. of patients	No. of deaths (%)[a]	5Y-OSR (%)[b]	Log-rank *P*	HR (95% CI)[c]	*P*[c]
Haplotype of rs1005165C>T, rs967591G>A, and rs735482A>C						
CGA	320	115 (35.9)	60.7	0.03	1.00	
TAC	301	135 (44.9)	49.1		1.32 (1.03–1.70)	0.03
Others[d]	35	10 (28.6)	66.9		0.81 (0.42–1.55)	0.52
Diplotype of rs1005165 C>T, rs967591G>A, and rs735482A>C						
Others[e]/others	96	31 (32.3)	61.1	0.01	1.00	
TAC/others	163	63 (38.7)	61.3		0.90 (0.67–1.21)	0.49
TAC/TAC	69	36 (52.2)	33.5		1.70 (1.18–2.44)	0.004
Others/others + TAC/others	259	94 (36.3)	61.3	0.003	1.00	
TAC/TAC	69	36 (52.2)	33.5		1.91 (1.29–2.83)	0.001
Haplotype of rs3212986G>T and rs11615G>A						
GG	330	134 (40.6)	54.4	0.34	1.00	
GA	152	64 (42.1)	51.2		0.95 (0.71–1.29)	0.76
TG	174	62 (35.6)	61.7		0.81 (0.60–1.10)	0.17
GG + GA	482	198 (41.1)	53.3	0.14	1.00	
TG	174	62 (35.6)	61.7		0.82 (0.62–1.09)	0.18

Abbreviation: 5Y-OSR, 5-year OS rate.
[a]Row percentage.
[b]Five-year survival rate, proportion of survival derived from Kaplan–Meier analysis.
[c]HRs, 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.
[d]Haplotypes that had a frequency of less than 5%.
[e]Any haplotype other than the TAC haplotype.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024

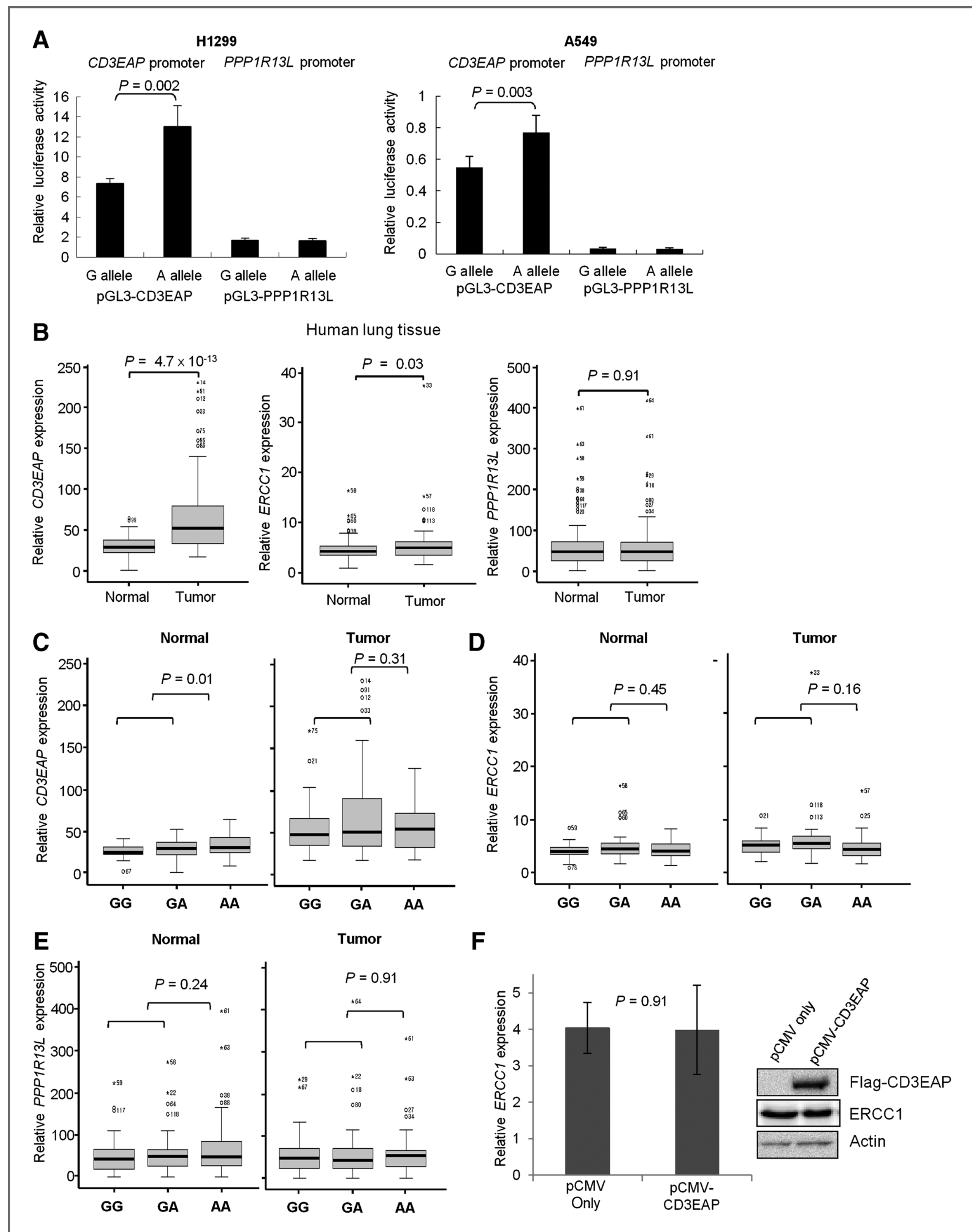

Figure 2. Effect of the rs967591G>A on transcription activities of *CD3EAP* and *PPP1R13L* and mRNA expression of *CD3EAP*, *PPP1R13L*, and *ERCC1*. A, the transcription activity was measured using the Dual Luciferase Reporter Assay System with pGL3-CD3EAP and pGL3-PPP1R13L constructs. The

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024

the 5 SNPs, the rs1005165C>T, rs967591G>A, rs735482A>C, rs3212986G>T, 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 rs3212986G>T 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 = 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 = 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 Alibaba2 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 Alibaba2 program (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 nonsynonymous 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 rs976591G>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 ($P = 0.01$, 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 *PPP1R3L* mRNA expression levels were normalized with that of β-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 β-actin gene in qRT-PCR analysis. The overexpression of CD3EAP was verified by Western blotting with FLAG antibody (Sigma-Aldrich). ERCC1 and β-actin antibodies (Santa Cruz Biotechnology) were used for ERCC1 protein level detection and internal control, respectively.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024

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 genotype 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 sillico* 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; $P_{heterogeneity} = 1.00$; 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,

Table 4. OS according to the rs967591G>A genotypes

	Stage I				Stage II					Stage I + Stage II			
Genotype	No. of deaths/ patients (%)[a]	5Y-OSR (%)[b]	HR (95% CI)[c]	P^c	No. of deaths/ patients (%)[a]	5Y-OSR(%)[b]	HR (95% CI)[c]	P^c	$P_H{}^d$	No. of deaths/ patients (%)[a]	5Y-OSR (%)[b]	HR (95% CI)[c]	P^c
GG	28/82 (34.2)	59.1	1.00		18/113 (15.9)	81.8	1.00			46/195 (23.6)	72.7	1.00	
GA	64/173 (37.0)	62.8	0.96 (0.61–1.51)	0.86	56/246 (22.8)	74.7	1.52 (0.89–2.59)	0.12	0.20	120/419 (28.6)	69.7	1.17 (0.83–1.65)	0.37
AA	38/73 (52.1)	32.4	1.75 (1.06–2.89)	0.03	45/124 (36.3)	54.2	2.43 (1.40–4.22)	0.002	0.39	83/197 (42.1)	46.9	1.88 (1.31–2.71)	0.001
Dominant		54.4	1.15 (0.75–1.77)	0.52		67.3	1.82 (1.10–3.01)	0.02	0.17		62.1	1.38 (1.00–1.91)	0.05
Recessive		61.9	1.80 (1.22–2.65)	0.003		77.0	1.80 (1.24–2.63)	0.002	1.00		70.7	1.69 (1.29–2.20)	0.0001
Codominant			1.36 (1.04–1.78)	0.03			1.57 (1.20–2.04)	0.0009	0.45			1.41 (1.17–1.69)	0.0003

Abbreviations: 5Y-OSR, 5-year OS rate.
[a]Row percentage.
[b]Five-year survival rate, proportion of survival derived from Kaplan–Meier analysis.
[c]HRs, 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.
[d]Wald test for heterogeneity of adjusted HRs between the two study samples.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024

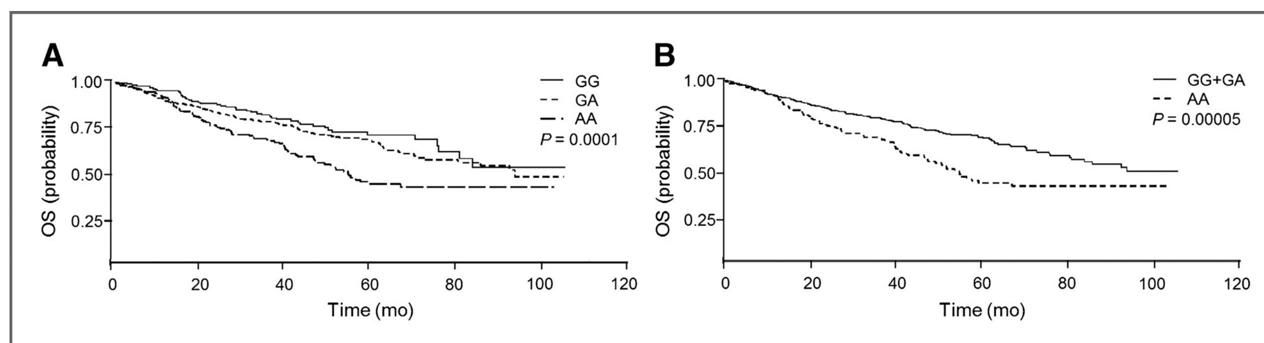

Figure 3. Kaplan–Meier plot of OS according to the rs967591G>A genotypes. Under a referent model (A), and under a recessive model for the variant allele (B). *P* values, log-rank test.

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^{-4}), 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,

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024

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

Conception and design: H.-S. Jeon, G. Jin, H.-G. Kang, W.-K. Lee, E.B. Lee, C.-H. Kim, S. Jheon, J.Y. Park
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
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): H.-S. Jeon, G. Jin, H.-G. Kang, Y.Y. Choi, W.-K. Lee, J.-E. Choi, S.S. Yoo, S.Y. Lee, S.-I. Cha, S. Jheon, I.-S. Kim, J.Y. Park
Writing, review, and/or revision of the manuscript: H.-S. Jeon, G. Jin, H.-G. Kang, Y.Y. Choi, W.-K. Lee, J.-E. Choi, E.Y. Bae, S.S. Yoo, S.Y. Lee, E.B. Lee, J. Lee, C.-H. Kim, S. Jheon, I.-S. Kim, 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

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, et al. Cancer statistics, 2008. CA Cancer J Clin 2008;58:71–96.
2. Detterbeck FC, Boffa DJ, Tanoue LT. The new lung cancer staging system. Chest 2009;136:260–71.
3. Coate LE, John T, Tsao MS, Shepherd FA. Molecular predictive and prognostic markers in non-small cell lung cancer. Lancet Oncol 2009;10:1001–110.
4. Gonzalez-Angulo AM, Hennessy BT, Mills GB. Future of personalized medicine in oncology: a systems biology approach. J Clin Oncol 2010;28:2777–83.
5. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. Science 2001;291:1284–9.
6. Wang D, Lippard SJ. Cellular processing of platinum anticancer drugs. Nat Rev Drug Disc 2005;4:307–20.
7. Yang JP, Hori M, Sande T, Okamoto T. Identification of a novel inhibitor of nuclear factor-kappaB, RelA-associated inhibitor. J Biol Chem 1999;274:15662–70.
8. Chen X, Kandasamy K, Srivastava RK. Differential roles of RelA (p65) and c-Rel subunits of nuclear factor kappa B in tumor necrosis factor-related apoptosis-inducing ligand signaling. Cancer Res 2003;63:1059–66.
9. Bergamaschi D, Samuels Y, O'Neil NJ, Trigiante G, Crook T, Hsieh JK, et al. iASPP oncoprotein is a key inhibitor of p53 conserved from worm to human. Nat Genet 2003;33:162–7.
10. Laska MJ, Lowe SW, Zender L, Hearn S, Vogel U, Jensen UB, et al. Enforced expression of PPP1R13L increases tumorigenesis and invasion through p53-dependent and p53-independent mechanisms. Mol Carcinog 2009;48:832–42.
11. Whitehead CM, Winkfein RJ, Fritzler MJ, Rattner JB. ASE-1: a novel protein of the fibrillar centres of the nucleolus and nuclus organizer region of mitotic chromosomes. Chromosoma 1997;106:493–502.
12. Yin J, Rockenbauer E, Hedayati M, Jacobsen NR, Vogel U, Grossman L, et al. Multiple single nucleotide polymorphisms on human chromosome 19q13.2-3 associate with risk of basal cell carcinoma. Cancer Epidemiol Biomarkers Prev 2002;11:1449–53.
13. Nexo BA, Vogel U, Olsen A, Ketelsen T, Bukowy Z, Thomsen BL, et al. A specific haplotype of single nucleotide polymorphisms on chromosome 19q13.2-3 encompassing the gene RAI is indicative of post-menopausal breast cancer before age 55. Carcinogenesis 2003;24:899–904.
14. Vogel U, Laros I, Jacobsen NR, Thomsen BL, Bak H, Olsen A, et al. Two regions in chromosome 19q13.2-3 are associated with risk of lung cancer. Mutat Res 2004;546:65–74.
15. Isla D, Sarries C, Rosell R, Alonso G, Domine M, Taron M, et al. Single nucleotide polymorphisms and outcome in doxetaxel-cisplatin-treated advanced non-small cell lung cancer. Ann Oncol 2004;15:1194–203.
16. Takenaka T, Yano T, Kiyohara C, Miura N, Kouso H, Ohba T, et al. Effects of excision repair cross-complementation group 1 (ERCC1) single nucleotide polymorphisms on the prognosis of non-small cell lung cancer patients. Lung Cancer 2010;67:101–7.
17. Yin M, Yan J, Voutsina A, Tibaldi C, Christiani DC, Heist RS, et al. No evidence of an association of ERCC1 and ERCC2 polymorphisms with clinical outcomes of platinum-based chemotherapies in non-small cell lung cancer: a meta-analysis. Lung Cancer 2011;72:370–7.
18. Chen S, Zhang J, Wang R, Luo X, Chen H. The platinum-based treatments for advanced non-small cell lung cancer, is low/negative ERCC1 expression better than high/positive ERCC1 expression? A meta-analysis. Lung Cancer 2010;70:63–70.
19. Osawa K. SNPs in ERCC1 and drug response to cisplatin in non-small-cell lung cancer patients. Pharmacogenomics 2011;12:445–7.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024

20. Park JY, Lee WK, Jung DK, Choi JE, Park TI, Lee EB, et al. Polymorphisms in the FAS and FASL genes and survival of early stage non-small cell lung cancer. Clin Cancer Res 2009;15:1794–800.

21. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 2001;25:402–8.

22. Stephens M, Smith MJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978–89.

23. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225–9.

24. Grabe N. Alibaba2: context specific identification of transcription factor binding sites. In Silico Biol 2002;2:S1–15.

25. Rebbeck TR, Spitz M, Wu X. Assessing the function of genetic variants in candidate gene association studies. Nat Rev Genet 2004;5:589–97.

26. Sunyaev S, Ramensky V, Koch I, Lathe W III, Kondrashov AS, Bork P. Prediction of deleterious human alleles. Hum Mol Genet 2001; 10:591–7.

27. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer 2004;4:850–60.

28. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al.: Replicating genotype-phenotype associations. Nature 2007;447:655–60.

29. Knee R, Murphy PR. Regulation of gene expression by natural antisense RNA transcripts. Neurochem Int 1997;31:379–92.

30. Timmons JA, Good L. Does everything now make (anti)sense? Biochem Soc Trans 2006;34:1148–50.

31. Yin J, Vogel U, Ma Y, Wang H, Wang C, Liang D, et al. A specific diplotype defined by PPP1R13L rs1970764, CD3EAP rs967591 and ERCC1 rs11615 and lung cancer risk in a Chinese population. Lung Cancer 2012;76:286–91.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/19/15/4185/2013491/4185.pdf by guest on 06 December 2024