

Functional intronic *ERCC1* polymorphism from regulomeDB can predict survival in lung cancer after surgery

Shin Yup Lee^{1,2,*}, Mi Jeong Hong^{3,10,*}, Hyo-Sung Jeon³, Yi Young Choi³, Jin Eun Choi^{3,4}, Hyo-Gyoung Kang^{3,4}, Deuk Kju Jung^{3,10}, Chengcheng Jin^{3,10}, Sook Kyung Do^{3,10}, Seung Soo Yoo^{1,2}, Yangki Seok⁵, Eung Bae Lee⁵, Kyung Min Shin⁶, Ji Yun Jeong⁷, Won Kee Lee⁸, Jaehee Lee¹, Seung Ick Cha¹, Chang Ho Kim¹, Young Tae Kim⁹, Sanghoon Jheon⁹, Jae Yong Park^{1,2,3,4,10}

¹Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

²Lung Cancer Center, Kyungpook National University Medical Center, Daegu, Republic of Korea

³Department of Biochemistry and Department of Cell Biology, and Kyungpook National University, Daegu, Republic of Korea

⁴Cell and Matrix Research Institute, Kyungpook National University, Daegu, Republic of Korea

⁵Thoracic Surgery, Kyungpook National University, Daegu, Republic of Korea

⁶Radiology, Kyungpook National University, Daegu, Republic of Korea

⁷Pathology, Kyungpook National University, Daegu, Republic of Korea

⁸Biostatistics Center, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

⁹Department of Thoracic and Cardiovascular Surgery, Seoul National University School of Medicine, Seoul, Republic of Korea

¹⁰BK21 Plus KNU Biomedical Convergence Program, Department of Biomedical Science, Kyungpook National University, Daegu, Republic of Korea

*These authors have contributed equally to this work

Correspondence to:

Jae Yong Park, e-mail: jaeyong@knu.ac.kr

Keywords: *ERCC1*, polymorphisms, regulomeDB, lung cancer,

Received: March 11, 2015

Accepted: May 13, 2015

Published: May 27, 2015

ABSTRACT

We searched for potential regulatory single nucleotide polymorphisms (SNPs) in *excision repair cross-complementing group 1 (ERCC1)* using RegulomeDB, a database integrating information from the Encyclopedia of DNA Elements (ENCODE) project, and investigated their association with survival after surgery in non-small cell lung cancer (NSCLC). Among 364 SNPs found within *ERCC1* region using RegulomeDB, four top priority SNPs (rs2298881C>A, rs1049739A>G, rs10415949A>G and rs6509214G>T) were selected for this study. The four SNPs were investigated in 316 patients. A replication study was performed ($n = 579$). Of the four SNPs analyzed in the discovery set, rs2298881C>A and rs6509214G>T were significantly associated with survival outcomes. The association was consistently observed only for rs2298881C>A in the validation cohort. In combined analysis, rs2298881C>A was significantly associated with worse overall survival and disease-free survival ($P = 0.0002$ and 0.02 , respectively). A decreased reporter gene expression for rs2298881 A allele was observed compared with C allele by luciferase assay ($P = 0.02$). *ERCC1* rs2298881C>A, an intronic SNP, is the first genetic polymorphism with functional evidence of regulating its expression, and the SNP is associated with prognosis of NSCLC. Our result supports the role of RegulomeDB as a comprehensive source of prioritized candidate SNPs for genetic association studies.

INTRODUCTION

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 [1, 2]. Therefore, ERCC1 has been linked to protection against development and progression of cancer, and resistance to platinum-based anticancer drugs at the same time: the double-edged sword. Based on the biological significance of ERCC1, its use as a predictive or prognostic biomarker has been pursued by a large number of cancer researchers. The expression of *ERCC1* by quantitative real-time polymerase chain reaction or immunohistochemistry has been correlated with the clinical outcomes of non-small cell lung cancer (NSCLC) [3–5]. Genetic polymorphism of *ERCC1* has also been investigated for the association with the risk and clinical outcome of many types of cancer including NSCLC [6–14]. The most widely studied single nucleotide polymorphisms (SNPs) include rs11615T>C (N118N) which is the only SNP tested in the exon region of *ERCC1*, and rs3212986C>A in 3'-UTR of *ERCC1* (Q504K for *CD3EAP*, antisense to *ERCC1*). Not surprisingly, the association of these SNPs with NSCLC has not been consistent across studies.

The human genome project has revealed that only 2% of human genome contains protein-coding genes, with the vast majority of human genome remained as 'junk DNA' [15]. However, despite intensive studies focused on protein-coding genes, our understanding of the genome has been far from complete [16]. In addition, nearly 90% of the variants identified as phenotype-associated SNPs in genome-wide association studies (GWAS) have been located within intergenic or intronic regions, posing an obstacle to its interpretation [16, 17]. Therefore, it has been suggested that the genome region outside the protein-coding genes may have the key to open the treasure chest of the vast genetic information of human genome.

The Encyclopedia of DNA Elements (ENCODE) project has the aim of describing all functional elements encoded in the human genome [16]. It revealed that 80% of the genome, especially outside of protein-coding regions, contains elements linked to biochemical functions such as DNA-transcription factor binding, providing new insights into the mechanisms of gene regulation [16]. RegulomeDB is a database which integrates a large collection of regulatory information from ENCODE and other data sources [17], being a rich source of information that may provide putative mechanistic explanations for genetic association studies including GWAS [18]. Until recently, a few studies utilized RegulomeDB to predict regulatory function of SNPs in non-coding regions that were identified by GWAS or candidate gene study [19–22].

RegulomeDB provides a scoring system prioritizing SNPs based on the degree of experimental or computational evidence that a variant lies in a functional location and likely results in a functional consequence, e.g., alteration of transcription factor binding and gene expression [17]. Therefore, selection of potential regulatory SNPs using RegulomeDB may help to improve power to detect true causal variants in genetic association studies. In the present study, we selected SNPs with high confidence of functional consequence in *ERCC1* gene region using RegulomeDB and investigated the association between those SNPs and the survival of NSCLC patients after curative surgery.

RESULTS

Patient characteristics and clinical predictors

The clinical and pathologic characteristics of patients in the discovery and validation sets and the association with OS and DFS are shown in Table 1. Upon univariate analysis, pathologic stage was significantly associated with OS and DFS in both sets (log-rank P [P_{L-R}] for OS = 2.0×10^{-6} and 0.0006; and P_{L-R} for DFS = 4.0×10^{-10} , 5.0×10^{-7} , respectively). Gender was associated with OS in the discovery set (P_{L-R} for OS = 0.04), and age was associated with OS and DFS in the validation set (P_{L-R} for DFS = 0.0001 and 0.01, respectively).

Associations between SNPs and survival outcomes

Among the four SNPs analyzed in the discovery set, the rs2298881C>A and rs6509214G>T were significantly associated with survival outcomes when adjusted for age, gender, smoking status, tumor histology, pathologic stage, and adjuvant therapy (Table 2). However, the association was consistently observed only for the rs2298881C > A in an independent validation set, which was in the same direction as the discovery set. In combined analysis, the rs2298881C > A was significantly associated with worse OS and DFS (adjusted HR [aHR] for OS, 1.37; 95% CI, 1.16–1.63; P = 0.0002; aHR for DFS, 1.17; 95% CI, 1.03–1.34; P = 0.02; under additive genetic model; Table 2 and Figure 1).

Effect of rs2298881C > A on the promoter activity of *ERCC1*

To investigate whether rs2298881C > A affects promoter activity of *ERCC1*, we generated three pGL3-*ERCC1* constructs: pGL3-*ERCC1*pro with cloned *ERCC1* promoter region alone, and pGL3-*ERCC1*pro_C

Table 1: Univariate analysis for overall survival and disease-free survival by clinicopathological characteristics in the discovery cohorts and validation cohorts

Variables	Discovery set						Validation set					
	Overall survival			Disease-free survival			Overall survival			Disease-free survival		
	No. of cases	No. of death (%) [†]	5Y-OSR (%) [†]	No. of event (%) [*]	5Y-DFSR (%) [†]	No. of cases	No. of death (%) [*]	5Y-OSR (%) [†]	No. of event (%) [*]	5Y-DFSR (%) [†]	No. of event (%) [*]	5Y-DFSR (%) [†]
Overall	316	126 (39.9)	53	160 (50.6)	43	579	164 (28.3)	67	296 (51.1)	45		
Age (years)												
≤ 63	161	60 (37.3)	57	81 (50.3)	44	306	70 (22.9)	74	146 (47.7)	50		0.01
> 63	155	66 (42.6)	48	79 (51.0)	41	273	94 (34.4)	58	150 (55.0)	38		
Gender												
Female	74	19 (25.7)	62	33 (44.6)	45	164	39 (23.8)	73	81 (49.4)	46		0.54
Male	242	107 (44.2)	50	127 (52.5)	42	415	125 (30.1)	65	215 (51.8)	45		
Smoking status												
Never	72	21 (29.2)	63	34 (47.2)	43	202	55 (27.2)	69	109 (54.0)	40		0.61
Ever	244	105 (43.0)	50	126 (51.6)	43	377	109 (28.9)	66	187 (49.6)	48		
Pack-years*												
< 40	106	43 (40.6)	51	75 (51.4)	42	175	47 (26.9)	68	85 (48.6)	42		0.37
≥ 40	138	62 (44.9)	49	51 (52.0)	43	202	62 (30.7)	64	102 (50.5)	47		

(continued)

Variables	Discovery set						Validation set						
	Overall survival			Disease-free survival			Overall survival			Disease-free survival			
	No. of cases	No. of death (%) [*]	5Y-OSR (%) [†]	Log-Rank P	No. of event (%) [*]	5Y-DFS (%) [†]	No. of cases	No. of death (%) [*]	5Y-OSR (%) [†]	Log-Rank P	No. of event (%) [*]	5Y-DFS (%) [†]	Log-Rank P
Histological type													
SCC	172	67 (39.0)	54	0.83	81 (47.1)	46	241	67 (27.8)	68	0.18	111 (46.1)	51	0.16
AC	139	56 (40.3)	51		76 (54.7)	37	312	86 (27.6)	68		169 (54.2)	41	
LCC	5	3 (60.0)	60		3 (60.0)	60	26	11 (42.3)	52		16 (61.5)	42	
Pathologic stage													
I	194	55 (28.4)	61	2.0 × 10 ⁻⁶	72 (37.1)	53	348	82 (23.6)	71	0.0006	152 (43.7)	51	5.0 × 10 ⁻⁷
II+III	122	71 (58.2)	40		88 (72.1)	27	231	82 (35.5)	61		144 (62.3)	36	
Adjuvant therapy													
No	67	37 (55.2)	39	0.79	47 (70.1)	29	110	40 (36.4)	65	0.87	72 (65.5)	37	0.75
Yes	55	34 (61.8)	41		41 (74.6)	26	121	42 (34.7)	56		72 (59.5)	34	

Abbreviations: SCC, Squamous cell carcinoma; AC, adenocarcinoma; LCC, Large cell carcinoma.

^{*}Row percentage.

[†]Five year-overall survival rate (5Y-OSR) and 5 year-disease free survival rate (5Y-DFS), proportion of survival derived from Kaplan-Meier analysis.

[‡]In ever-smokers.

Table 2: Association of rs2298881C>A and rs6519214G>T and survival outcomes in the discovery and validation sets

ID No.*	Genotypes	Discovery cohort			Validation cohort			Combined cohort					
		Overall survival		Disease-free survival	Overall survival		Disease-free survival	Overall survival		Disease-free survival			
		HR (95%CI) [†]	P [‡]	HR (95%CI) [†]	P [‡]	HR (95%CI) [†]	P [‡]	HR (95%CI) [†]	P [‡]	HR (95%CI) [†]	P [‡]		
rs2298881 [‡]	CC	1.00		1.00		1.00		1.00		1.00		1.00	
	CA	1.18 (0.75–1.86)	0.47	1.07 (0.73–1.57)	0.73	1.47 (0.97–2.23)	0.07	1.26 (0.95–1.68)	0.11	1.30 (0.96–1.75)	0.09	1.19 (0.95–1.49)	0.13
	AA	1.92 (1.15–3.20)	0.01	1.59 (1.01–2.48)	0.04	2.00 (1.26–3.16)	0.003	1.31 (0.93–1.84)	0.13	1.88 (1.34–2.63)	0.000	1.38 (1.05–1.80)	0.02
	Dominant	1.37 (0.90–2.09)	0.14	1.20 (0.84–1.72)	0.31	1.63 (1.10–2.42)	0.02	1.28 (0.98–1.67)	0.08	1.46 (1.10–1.94)	0.01	1.24 (1.00–1.54)	0.05
	Recessive	1.73 (1.14–2.63)	0.01	1.52 (1.04–2.22)	0.03	1.55 (1.09–2.19)	0.01	1.12 (0.85–1.49)	0.42	1.58 (1.21–2.06)	0.001	1.23 (0.98–1.54)	0.07
	Additive	1.39 (1.06–1.81)	0.02	1.25 (0.99–1.58)	0.06	1.41 (1.13–1.77)	0.003	1.15 (0.97–1.36)	0.11	1.37 (1.16–1.63)	0.000	1.17 (1.03–1.34)	0.02
rs6509214 [§]	GG	1.00		1.00		1.00		1.00		1.00		1.00	
	GT	1.47 (0.97–2.24)	0.07	1.63 (1.13–2.37)	0.01	0.91 (0.60–1.38)	0.65	0.97 (0.71–1.31)	0.82	1.17 (0.87–1.57)	0.30	1.21 (0.95–1.52)	0.12
	TT	1.67 (0.94–2.97)	0.08	1.61 (0.97–2.67)	0.06	0.87 (0.51–1.49)	0.61	0.75 (0.50–1.12)	0.16	1.10 (0.75–1.61)	0.64	0.99 (0.73–1.35)	0.95
	Dominant	1.51 (1.01–2.26)	0.05	1.63 (1.14–2.33)	0.01	0.90 (0.60–1.33)	0.59	0.90 (0.67–1.20)	0.45	1.15 (0.87–1.52)	0.33	1.14 (0.91–1.43)	0.24
	Recessive	1.31 (0.80–2.16)	0.29	1.18 (0.76–1.83)	0.45	0.93 (0.58–1.48)	0.75	0.77 (0.54–1.09)	0.14	1.00 (0.71–1.40)	0.99	0.89 (0.68–1.16)	0.38
	Additive	1.32 (1.01–1.73)	0.05	1.32 (1.04–1.66)	0.02	0.93 (0.72–1.21)	0.59	0.88 (0.73–1.06)	0.19	1.06 (0.89–1.28)	0.52	1.02 (0.88–1.18)	0.78

Abbreviations: HR, hazard ratio; CI, confidence interval.

[†]Information about polymorphisms and IDs were obtained from NCBI database (<http://www.ncbi.nlm.nih.gov/SNP>).

[‡]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, adjuvant therapy.

[§]The numbers of patients in discovery, validation, and combined cohort for rs2298881 were 316, 579, and 895, respectively.

[¶]The numbers of patients in discovery, validation, and combined cohort for rs6509214 were 316, 412, and 728, respectively, due to lack of available samples for validation.

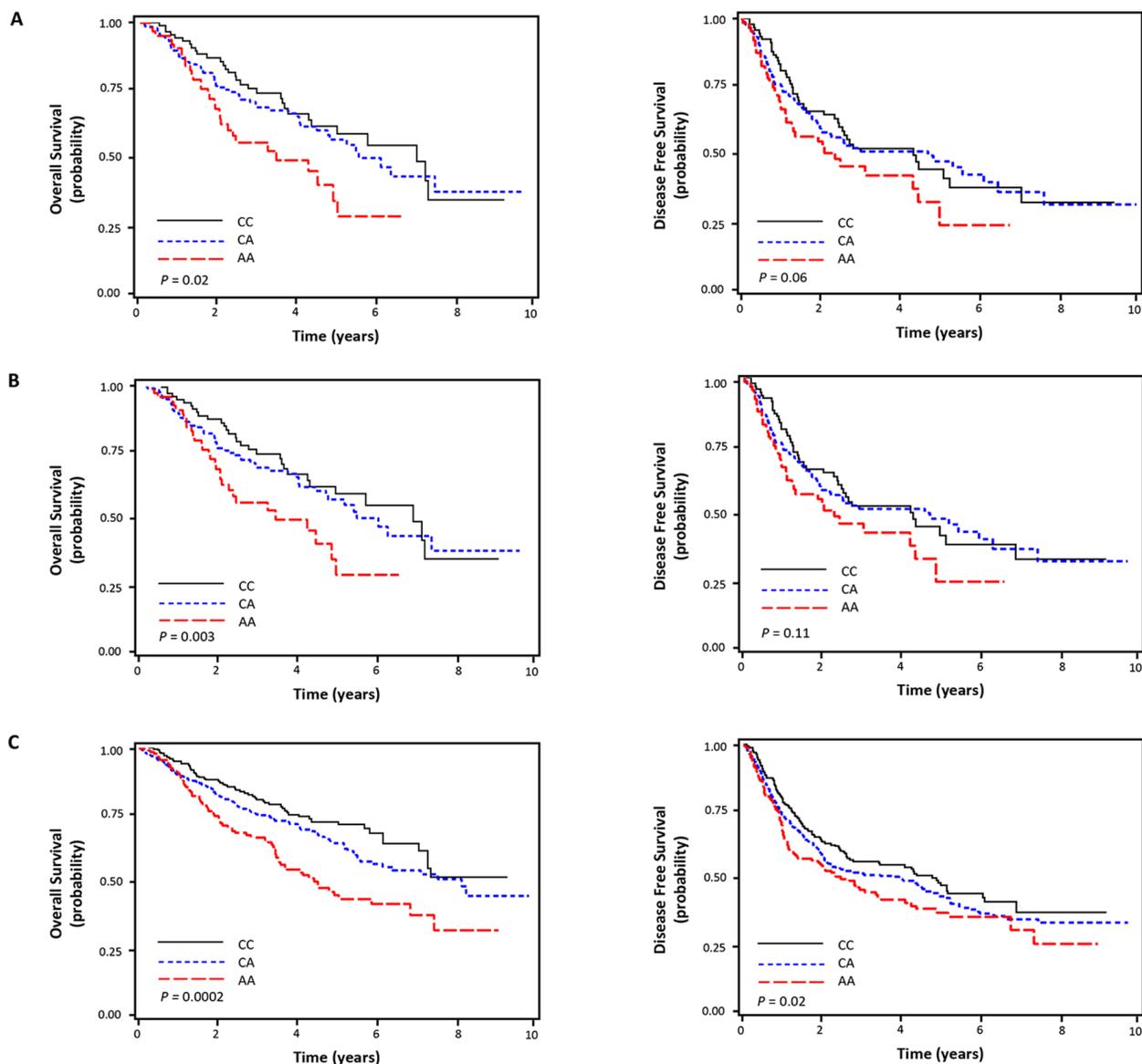


Figure 1: Kaplan-Meier plot of overall and disease free survival curves according to *ERCC1* rs2298881C > A genotype in discovery cohort. A. replication cohort B. and combined cohort C. *P* values in the multivariate Cox proportional hazard model.

and pGL3-*ERCC1*pro_A with both promoter region and the fragment containing rs2298881C>A (Figure 2A). As shown in Figure 2B, luciferase activity was significantly higher in H1299 cells transfected with pGL3-*ERCC1*pro_C or pGL3-*ERCC1*pro_A compared with pGL3-*ERCC1*pro, suggesting that the fragment containing rs2298881C>A enhanced the activity of *ERCC1* promoter. A decreased expression of the reporter gene for the A allele of rs2298881C>A was observed compared with the C allele by luciferase assay ($P = 0.02$; Figure 2B). These results suggest that an intronic SNP rs2298881C>A may alter *ERCC1* expression by affecting *ERCC1* promoter activity.

DISCUSSION

We investigated the association between potential regulatory SNPs in *ERCC1* gene region selected from RegulomeDB and survival of patients with surgically resected early stage NSCLC in a relatively large two-stage study including 895 patients. Our study showed significant association between *ERCC1* rs2298881C>A and the prognosis of patients with early stage NSCLC, which was reproducible in an independent set of patients. We also report that rs2298881C>A, an intronic SNP of *ERCC1*, is the first genetic polymorphism with functional evidence of regulating *ERCC1* expression. These findings suggest

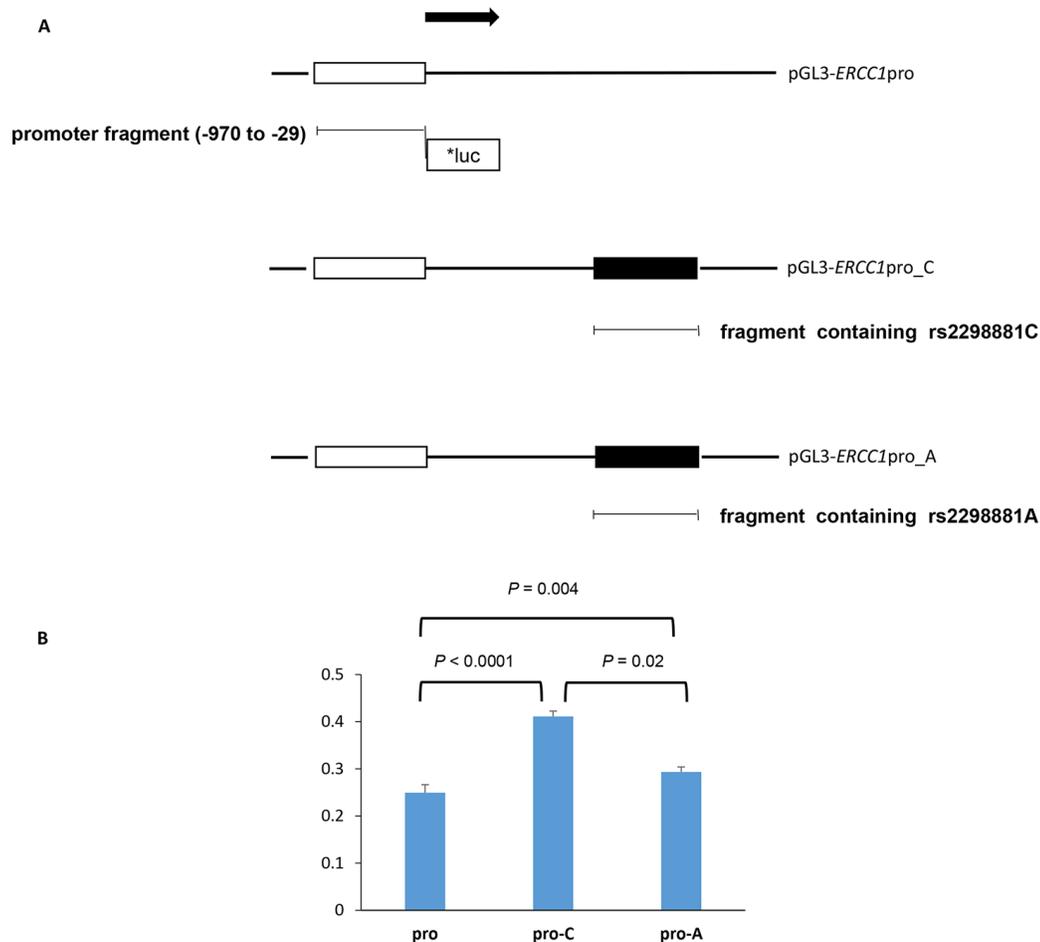


Figure 2: Functional analysis of the *ERCC1* rs2298881C>A. **A.** Schematic representation of the constructs that were used for the reporter gene assays. Promoters are marked by white blocks and the fragments including rs2298881C> A site by black blocks, and arrow indicates the direction of transcription. The first base of translation start site is denoted as +1. *ERCC1* promoter was amplified from human genomic DNA and cloned into the pGL3 basic vector (pGL3-*ERCC1*pro). DNA fragments containing the SNP site were cloned into other multi cloning site (pGL3-*ERCC1*pro_C and pGL3-*ERCC1*pro_A). **B.** Luciferase activity according to *ERCC1* rs2298881C>A. H1299 cells were transfected with pGL3-*ERCC1*pro, pGL3-*ERCC1*pro_C and pGL3-*ERCC1*pro_A constructs, respectively. Each bars represent mean \pm S.E.M. of firefly luciferase activity normalized to *Renilla* luciferase activity. Experiments were performed in triplicate. *P* value, a Student's *t*-test. luc, luciferase.

that *ERCC1* rs2298881C>A could be used as a prognostic marker for early stage NSCLC, and that RegulomeDB may be useful in selecting potentially functional SNPs in the regulatory region for genetic association studies.

In the present study, we searched for regulatory SNPs in *ERCC1* gene region using RegulomeDB and showed that rs2298881C>A was associated with worse prognosis of NSCLC patients after curative resection. *In vitro* luciferase assay showed that the *ERCC1* rs2298881C-to-A change was associated with reduced promoter activity of *ERCC1*. According to RegulomeDB, the rs2298881C>A has the highest level of evidence for regulatory role among SNPs in *ERCC1* gene region. In addition, based on RegulomeDB, rs2298881C>A is the only SNP throughout the whole genome reported to be in the eQTL that is predicted to regulate the expression of *ERCC1*. Our result is in line with the realization of

regulatory function of non-coding DNA and suggests the need for investigating variants in regulatory region.

Recent results of ENCODE project provided evidence revealing that genetic variation in non-coding DNA play an important role in the regulation of gene expression. The ENCODE data show that the results of GWAS are typically enriched for variants within non-coding functional units, suggesting that many of these regions could be causally linked to disease [17, 18, 24]. Therefore, RegulomeDB containing ENCODE data is a powerful tool for predicting the likelihood of a SNP being in a functional location, thereby facilitates prioritizing SNPs for genetic association studies. The result of our study supports the role of RegulomeDB in selecting putative regulatory SNPs for future genetic association studies.

Genetic polymorphisms of *ERCC1* have been investigated in terms of the risk and the clinical outcomes

in many types of cancer including NSCLC [6–14]. However, most of the studies have focused on only a few SNPs, such as *ERCC1* rs11615T>C (N118N) and rs3212986C>A in 3'-UTR, and the results have not been consistent among studies. We previously investigated these two SNPs in terms of the clinical outcomes of early-stage NSCLC after surgery and advanced NSCLC after platinum-based chemotherapy in Koreans [13, 25, 26]. However, neither rs11615T>C nor rs3212986C>A showed significant association with the outcome of NSCLC [13, 25, 26]. In the present study, we searched RegulomeDB for potential regulatory SNPs in *ERCC1*, and investigated their association with survival after surgery in NSCLC. In fact, rs2298881C>A and other intronic SNPs such as rs3212961A>C and rs3212948G>C were selected as haplotype tagging SNPs in previous studies that first evaluated those SNPs [27, 28]. However, a small number of studies on rs2298881C>A with variable number of patients have shown discrepant association with various types of cancer [27, 29–31].

In this study, we included a total of 895 patients which is relatively large for studies on surgically resected NSCLC. The association between *ERCC1* rs2298881C>A and survival outcomes was replicated across both discovery and validation sets of the study, which would largely reduce false positivity [32, 33]. In addition, the association of rs2298881C>A with survival outcome was biologically plausible. It is possible that the *ERCC1* rs2298881 C-to-A change in the putative regulatory region may lead to reduced promoter activity and decreased *ERCC1* expression, resulting in decreased DNA repair capacity and therefore worse disease outcome. However, in our preliminary analysis, significant difference in the relative expression level of *ERCC1* mRNA among genotypes of rs2298881C>A was not observed in either tumor or paired non-malignant lung tissues (Supplementary Figure 1). Future studies are required to understand the biologic mechanism of the observed association between the SNPs and survival outcomes.

In conclusion, this study showed that *ERCC1* rs2298881C>A could predict the survival outcomes of patients with surgically resected early stage NSCLC. RegulomeDB may be useful as a practical tool for selecting potentially functional SNPs in the regulatory region for future genetic association studies.

MATERIALS AND METHODS

Study population

The discovery set included 316 patients with pathologic stages I, II, or IIIA (micro-invasive N2) NSCLC who underwent curative surgical resection at Kyungpook National University Hospital (KNUH) between September 1998 and August 2007. Genomic DNA samples from tumor and corresponding non-malignant

lung tissue specimens were provided by the National Biobank of Korea - KNUH, which is supported by the Ministry of Health, Welfare and Family Affairs. Written informed consent was obtained from all patients prior to surgery. All materials derived from the National Biobank of Korea - KNUH were obtained under institutional review board-approved protocols. The validation set included 579 patients with pathologic stages I, II, or IIIA NSCLC who underwent curative surgical resection at KNUH ($n = 99$) and Seoul National University Hospital ($n = 307$), Seoul National University Bundang Hospital ($n = 173$). Written informed consent was obtained from all patients before surgery and research protocol was approved by the institutional review board at each hospital. All of the patients included in this study were ethnic Koreans. Patients who underwent chemotherapy or radiotherapy prior to surgery were excluded to avoid the effects on DNA. The pathologic staging of the tumors was determined according to the International System for Staging Lung Cancer [23].

SNP selection and genotyping

Three hundred and sixty four SNPs were found within *ERCC1* gene region, NC_000019.9 (45910591..45982241, complement) by Genome Reference Consortium Human Build 37 patch release 13 (GRCh37.p13) assembly, using RegulomeDB (<http://regulome.stanford.edu>). The RegulomeDB provides a scoring system with categories ranging from 1 to 6 based on the degree of experimental or computational evidence of functional consequence of a given variant. Category 1 includes variants that are known expression quantitative trait loci (eQTLs), which have been shown to be associated with expression of target genes, and is further divided into subcategories 1a to 1f. Because the lower score indicates the stronger evidence for a variant to be located in a functional region, a variant scored as 1a most likely affects transcription factor binding and expression of a target gene. We prioritized the 364 SNPs using the RegulomeDB, and selected five SNPs that were classified into category 1: two SNPs (rs2298881C>A and rs1049739A>G) had a score of 1b, and three SNPs had a score of 1f (rs10415949A>G, rs6509214G>T, rs7245548C>T). Among those, rs7245548 was not genotyped because it was in linkage disequilibrium with rs6509214 by HapMap JPT database. Finally, four SNPs (rs2298881C>A, rs1049739A>G, rs10415949A>G, and rs6509214G>T) were selected for genotyping to investigate the relationship with the survival of NSCLC. Genomic DNA was extracted from tissues with QIAamp® genomic DNA kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The rs2298881C>A was genotyped using the Taq-Man® assay (Applied Biosystems, Foster City, CA) following the manufacturer's instructions. The Taq-Man probes were pre-designed and synthesized by Applied Biosystems. The rs10415949A>G and

rs6509214G>T were genotyped using SEQUENOM's MassARRAY® iPLEX assay according to instructions of the manufacturer. Duplicate samples and negative controls were included to ensure accuracy of genotyping. For validation of genotyping, approximately 5% of samples of the cohort were randomly selected to be genotyped again with a restriction fragment length polymorphism assay by a different investigator and the results were 100% concordant.

Cloning of the luciferase reporter gene and luciferase assay

The rs2298881C>A is an eQTL for *ERCC1* gene located in intron 1. RegulomeDB suggests that the SNP lies in a location which overlaps transcription binding site and regulates gene expression. We investigated whether rs2298881C>A affects *ERCC1* promoter activity by luciferase reporter assay. The pGL3-Basic Vector (Promega, Madison, WI, USA) was used to construct luciferase reporter plasmids using manufacturer's protocols. Briefly, promoter region of *ERCC1* (-970 to -29 bp, the transcriptional start site is designated as +1) was synthesized by polymerase chain reaction from human genomic DNA and cloned into the pGL3-Basic vector to generate pGL3-*ERCC1*pro. Two fragments including rs2298881C or rs2298881A allele of *ERCC1* rs2298881C>A were amplified from genomic DNA sample and the 160 bp products were cloned into pGL3-*ERCC1*pro, respectively. All constructs were verified by direct sequencing before use. Human non-small cell lung cancer cells (H1299) were maintained at 37°C in 5% CO₂ atmosphere in RPMI-1640 medium containing 10% heat-inactivated fetal bovine serum (FBS). The cells were transfected with 300 ng of each plasmid DNA (pGL3-*ERCC1*pro, pGL3-*ERCC1*pro_C, or pGL3-*ERCC1*pro_A) and 30 ng of pRL-SV40 Vector (Promega, Madison, WI, USA) using Effectene® (Qiagen, Hilden, Germany) according to manufacturer's protocol. Luciferase activity was measured on an Orion L Microplate Luminometer (Berthold Detection Systems GmbH, Pforzheim, Germany) using the Dual-Luciferase® Reporter Assay System (Promega, Madison, WI, USA). Firefly luciferase activity measurements were normalized with respect to pRL-SV40 *Renilla* luciferase activity to correct for variations in transfection efficiency. Each experiment was conducted in triplicate.

Statistical analysis

Differences in the distribution of genotypes according to the clinicopathologic factors of patients were compared using χ^2 tests. Hardy-Weinberg equilibrium was tested using a goodness-of-fit χ^2 test with 1 *degree of freedom*. The genotypes for each SNP were analyzed as a three-group categorical variable, and those were also grouped according to the dominant, recessive and additive model. Overall survival (OS) was measured from the day of surgery to the date of the last follow-up or until the date

of death. Disease-free survival (DFS) was calculated from the day of surgery until recurrence or death. The survival estimates were calculated using the Kaplan-Meier method. The difference in OS and DFS according to the SNPs was compared using log-rank tests. Cox's proportional hazard regression model was used for the multivariate survival analyses, and the analyses were always adjusted for age (> 63 years versus \leq 63), gender (male versus female), smoking status (ever versus never), tumor histology (squamous vs. non-squamous), pathologic stage (II-IIIa versus I), and adjuvant therapy (yes vs. no). The hazard ratio (HR) and 95% confidence interval (CI) were also estimated. A cut-off *p*-value of 0.05 was adopted for all the statistical analyses. The statistical data were obtained using SAS Genetic software (SAS Institute, Cary, NC).

ACKNOWLEDGMENTS AND FUNDING

This study was supported in part by the R&D program of MKE/KEIT (10040393, Development and commercialization of molecular diagnostic technologies for lung cancer through clinical validation), in part by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI14C0402).

REFERENCES

1. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. *Science*. 2001; 291:1284–1289.
2. Wang D, Lippard SJ. Cellular processing of platinum anticancer drugs. *Nat Rev Drug Discov*. 2005; 4:307–320.
3. Roth JA, Carlson JJ. Prognostic role of ERCC1 in advanced non-small-cell lung cancer: a systematic review and meta-analysis. *Clin Lung Cancer*. 2011; 12:393–401.
4. Simon GR, Sharma S, Cantor A, Smith P, Beppler G. ERCC1 expression is a predictor of survival in resected patients with non-small cell lung cancer. *Chest*. 2005; 127:978–983.
5. Olausson KA, Dunant A, Fouret P, Brambilla E, André F, Haddad V, Taranchon E, Filipits M, Pirker R, Popper HH, Stahel R, Sabatier L, Pignon JP, et al. DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *N Engl J Med*. 2006; 355:983–991.
6. Xu TP, Shen H, Liu LX, Shu YQ. Association of ERCC1-C118T and -C8092A polymorphisms with lung cancer risk and survival of advanced-stage non-small cell lung cancer patients receiving platinum-based chemotherapy: a pooled analysis based on 39 reports. *Gene*. 2013; 526:265–274.
7. Zhu J, Hua RX, Jiang J, Zhao LQ, Sun X, Luan J, Lang Y, Sun Y, Shang K, Peng S, Ma J. Association studies of ERCC1 polymorphisms with lung cancer susceptibility: a systematic review and meta-analysis. *PLoS One*. 2014; 9:e97616.

8. Tseden-Ish M, Choi YD, Cho HJ, Ban HJ, Oh IJ, Kim KS, Song SY, Na KJ, Ahn SJ, Choi S, Kim YC. Disease-free survival of patients after surgical resection of non-small cell lung carcinoma and correlation with excision repair cross-complementation group 1 expression and genotype. *Respirology*. 2012; 17:127–33.
9. Sullivan I, Salazar J, Majem M, Pallarés C, Del Río E, Páez D, Baiget M, Barnadas A. Pharmacogenetics of the DNA repair pathways in advanced non-small cell lung cancer patients treated with platinum-based chemotherapy. *Cancer Lett*. 2014; 353:160–166.
10. Rodriguez-Hernandez I, Perdomo S, Santos-Briz A, Garcia JL, Gomez-Moreta JA, Cruz JJ, Gonzalez-Sarmiento R. Analysis of DNA repair gene polymorphisms in glioblastoma. *Gene*. 2014; 536:79–83.
11. Zhang L, Wang J, Xu L, Zhou J, Guan X, Jiang F, Wu Y, Fan W. Nucleotide excision repair gene ERCC1 polymorphisms contribute to cancer susceptibility: a meta-analysis. *Mutagenesis*. 2012; 27:67–76.
12. Kong JH, Mun YC, Kim S, Choi HS, Kim YK, Kim HJ, Moon JH, Sohn SK, Kim SH, Jung CW, Dennis Kim DH. Polymorphisms of ERCC1 genotype associated with response to imatinib therapy in chronic phase chronic myeloid leukemia. *Int J Hematol*. 2012; 96:327–333.
13. Jeon HS, Jin G, Kang HG, Choi YY, Lee WK, Choi JE, Bae EY, Yoo SS, Lee SY, Lee EB, Kim YT, Lee J, Cha SI, et al. A functional variant at q13.3, rs967591G>A, is associated with shorter survival of early-stage lung cancer. *Clin Cancer Res*. 2013; 19:4185–4195.
14. Findlay JM, Middleton MR, Tomlinson I. A systematic review and meta-analysis of somatic and germline DNA sequence biomarkers of esophageal cancer survival, therapy response and stage. *Ann Oncol*. 2015; 26:624–644.
15. Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB. Annotating non-coding regions of the genome. *Nat Rev Genet*. 2010; 11:559–571.
16. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74.
17. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012; 22:1790–1797.
18. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res*. 2012; 22:1748–1759.
19. Rosenthal SL, Barmada MM, Wang X, Demirci FY, Kamboh MI. Connecting the dots: potential of data integration to identify regulatory SNPs in late-onset Alzheimer's disease GWAS findings. *PLoS One*. 2014; 9:e95152.
20. Sillé FC, Conde L, Zhang J, Akers NK, Sanchez S, Maltbaek J, Riby JE, Smith MT, Skibola CF. Follicular lymphoma-protective HLA class II variants correlate with increased HLA-DQB1 protein expression. *Genes Immun*. 2014; 15:133–136.
21. Pirim D, Wang X, Radwan ZH, Niemsiri V, Hokanson JE, Hamman RF, Barmada MM, Demirci FY, Kamboh MI. Lipoprotein lipase gene sequencing and plasma lipid profile. *J Lipid Res*. 2014; 55:85–93.
22. Sun Y, Zuo X, Zheng X, Zhou F, Liang B, Liu H, Chang R, Gao J, Sheng Y, Cui H, Wang W, Andiappan AK, Rotzschke O, et al. A comprehensive association analysis confirms ZMIZ1 to be a susceptibility gene for vitiligo in Chinese population. *J Med Genet*. 2014; 51:345–53.
23. Dettnerbeck FC, Boffa DJ, Tanoue LT. The new lung cancer staging system. *Chest*. 2009; 136:260–271.
24. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009; 106:9362–9367.
25. Kim M, Kang HG, Lee SY, Lee HC, Lee EB, Choi YY, Lee WK, Cho S, Jin G, Jheon HS, Son JW, Lee MH, Jung DK, et al. Comprehensive analysis of DNA repair gene polymorphisms and survival in patients with early stage non-small-cell lung cancer. *Cancer Sci*. 2010; 101:2436–2442.
26. Lee SY, Kang HG, Yoo SS, Kang YR, Choi YY, Lee WK, Choi JE, Jeon HS, Shin KM, Oh IJ, Kim KS, Lee J, Cha SI, et al. Polymorphisms in DNA repair and apoptosis-related genes and clinical outcomes of patients with non-small cell lung cancer treated with first-line paclitaxel-cisplatin chemotherapy. *Lung Cancer*. 2013; 82:330–339.
27. Yin J, Vogel U, Ma Y, Qi R, Wang H, Yue L, Liang D, Wang C, Li X, Song T. HapMap-based study of a region encompassing ERCC1 and ERCC2 related to lung cancer susceptibility in a Chinese population. *Mutat Res*. 2011; 713:1–7.
28. Ma H, Xu L, Yuan J, Shao M, Hu Z, Wang F, Wang Y, Yuan W, Qian J, Wang Y, Xun P, Liu H, Chen W, et al. Tagging single nucleotide polymorphisms in excision repair cross-complementing group 1 (ERCC1) and risk of primary lung cancer in a Chinese population. *Pharmacogenet Genomics*. 2007; 17:417–423.
29. Hui L, Yue S, Gao G, Chang H, Li X. Association of single-nucleotide polymorphisms in ERCC1 and ERCC2 with glioma risk. *Tumour Biol*. 2014; 35:7451–7457.
30. Lu ZM, Luo TH, Nie MM, Fang GE, Ma LY, Xue XC, Wei G, Ke CW, Bi JW. Influence of ERCC1 and ERCC4 polymorphisms on response to prognosis in gastric cancer treated with FOLFOX-based chemotherapy. *Tumour Biol*. 2014; 35:2941–2948.

31. He J, Xu Y, Qiu LX, Li J, Zhou XY, Sun MH, Wang JC, Yang YJ, Jin L, Wei QY, Wang Y. Polymorphisms in ERCC1 and XPF genes and risk of gastric cancer in an eastern Chinese population. *PLoS One*. 2012; 7:e49308.
32. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. *Nat Rev Cancer*. 2004; 4:850–860.
33. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, et al. Replicating genotype-phenotype associations. *Nature*. 2007; 447:655–660.