A genetic variation in microRNA target site of KRT81 gene is associated with survival in early-stage non-small-cell lung cancer


1Lung Cancer Center, Kyungpook National University Medical Center, Daegu; 2Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu; 3Department of Biochemistry and Cell Biology, School of Medicine; 4The Molecular Diagnostics & Imaging Research Institute; Departments of 5Thoracic Surgery; 6Pathology; 7Biostatistics Center, School of Medicine, Kyungpook National University, Daegu; 8Department of Thoracic and Cardiovascular Surgery, Seoul National University School of Medicine, Seoul; 9Department of Internal Medicine, Konkuk University Hospital, Daejeon; 10BK21 Plus KNU Biomedical Convergence Program, Department of Biomedical Science, Kyungpook National University, Daegu, Republic of Korea

Received 2 July 2014; revised 22 January 2015; accepted 12 February 2015

Background: MicroRNAs (miRNAs) have a key role in carcinogenesis through negative regulation of their target genes. Therefore, genetic variations in miRNAs or their target sites may affect miRNA–mRNA interactions, thereby resulting in altered expression of target genes. This study was conducted to investigate the associations between single-nucleotide polymorphisms (SNP) located in the miRNA target sites (poly-miRTSs) and survival of patients with early-stage non-small-cell lung cancer (NSCLC).

Methods: Using public SNP database and miRNA target sites prediction program, 354 poly-miRTSs were selected for genotyping. Among these, 154 SNPs applicable to Sequenom’s MassARRAY platform were investigated in 357 patients. A replication study was carried out on an independent patient population (n = 479). Renilla luciferase assay and reverse transcription-polymerase chain reaction were conducted to examine functional relevance of potentially functional poly-miRTSs.

Results: Of the 154 SNPs analyzed in a discovery set, 14 SNPs were significantly associated with survival outcomes. Among these, KRT81 rs3660G>C was found to be associated with survival outcomes in the validation cohort. In the combined analysis, patients with the rs3660 GC + CC genotype had a significantly better overall survival compared with those with GG genotype [adjusted hazard ratio (aHR) for OS, 0.65; 95% confidence interval (CI) 0.50–0.85; P = 0.001]. An increased expression of the reporter gene for the C allele of rs3660 compared with the G allele was observed by luciferase assay. Consistently, the C allele was associated with higher relative expression level of KRT81 in tumor tissues.

†These authors contributed equally to this paper.

*Correspondence to: Dr Jin Eun Choi, Department of Biochemistry and Cell Biology, School of Medicine, Kyungpook National University, 2-101 Dongn-dong, Jung-gu, Daegu 702-422, Republic of Korea. Tel. +82-53-200-3590; Fax: +82-53-252-5830; E-mail: jechoi.9711@gmail.com
Conclusions: The rs3660G>C affects KRT81 expression and thus influences survival in early-stage NSCLC. The analysis of the rs3660G>C polymorphism may be useful to identify patients at high risk of a poor disease outcome.

Key words: non-small cell lung cancer, miRNA target sites, polymorphisms, KRT81

Introduction
MicroRNAs (miRNAs) are endogenous small (~22 nucleotides) non-coding RNAs that regulate target gene expression by complementary binding to 3'-untranslated region (3'-UTR) of messenger RNAs (mRNAs), leading to translational repression or mRNA cleavage [1]. More than 1000 miRNAs have been identified in human genome and are thought to regulate the expression of at least 30% of all protein-coding genes, suggesting that miRNAs have important roles in a variety of biological function such as cell proliferation and survival, DNA repair, and immune response [2–4]. Evidence indicates that miRNAs are critically involved in the development and progression of diverse human cancers as tumor suppressors and/or oncogenes depending on the tissue type and the presence of specific targets [4–7].

The interaction of miRNAs and their target genes to which miRNAs bind to exert their effect is complicated: one miRNA can regulate multiple target genes whereas multiple miRNAs may coordinate to regulate one target gene [8, 9]. It has been demonstrated that mostly perfect complementarity to the ‘seed region’ in target sequence—nucleotides 2–7 from the 5′ end of the mature miRNA—is essential for miRNA–mRNA interactions [9]. Therefore, genetic variations in miRNAs or their target sites may affect miRNA–mRNA interactions, thereby result in altered expression of target genes. Recently, a growing body of evidence suggests that single-nucleotide polymorphisms (SNPs) in miRNA target sites (poly-miRTSs) are associated with the risk [10, 11] and the prognosis of diverse types of cancer [12, 13], including lung cancer.

Based on the important role of miRNA network in carcinogenesis, we hypothesized that poly-miRTSs may influence miRNA–mRNA interaction and consequently the expression of target genes, thereby influencing the prognosis of lung cancer. To test this hypothesis, we conducted a two-stage study to evaluate the associations between poly-miRTSs and the prognosis of lung cancer.

Materials and Methods

Study population
A two-stage approach is applied to evaluate the effect of poly-miRTSs on survival outcome in NSCLC patients. The discovery set included 357 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, Kyungpook National University Hospital, which is supported by the Ministry of Health, Welfare and Family Affairs. Written informed consent was obtained from all patients before surgery. All materials derived from the National Biobank of Korea, KNUH were obtained under institutional review board (IRB)-approved protocols (Approval No., KNUMCBIO_12-1005). The validation set included 479 patients with pathologic stages I, II, or IIIA NSCLC who underwent curative surgical resection at KNUH (n = 27) and Seoul National University Hospital (n = 277), Seoul National University Bundang Hospital (n = 175). Written informed consent was obtained from all patients before surgery and research protocol was approved by the IRB at each hospital. Patients who underwent chemotherapy or radiotherapy before 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 [14].

SNP selection and genotyping
To select all the potentially functional polymorphisms in miRNA target sites, we processed the following procedures: We searched all the polymorphisms in 3′ UTR of genes in public SNP database (http://www.ncbi.nlm.nih.gov/SNP) and Ensemble data, and select 12 877 SNPs in predicted miRNA target sites using the miRNA target prediction program miRanda algorithm version 1.2 released in September 2008 (http://www.microrna.org/microrna/). Among those, 1678 SNPs with the minor allele frequency ≥0.05 in the HapMap IPT data were collected after excluding those in linkage disequilibrium. Using RNAhybrid (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/submission.html), we calculated minimum free energy (MFE) for binding between all the predicted miRNAs and target mRNAs both for wild and variant alleles. And then we ranked the values of MFE difference between wild and variant alleles and chose (arbitrarily) the upper quintile of distribution (MFE difference ≥6.5) as the significant cutoff. Because a single target miRNA may bind to multiple miRNAs, we used the largest value of MFE difference for each target miRNA to rank the value. Among the resulting 354 SNPs, 154 SNPs applicable to Sequenom MassARRAY® iPLEX assay (Sequenom Inc., San Diego, CA) were genotyped. Genomic DNA was extracted from tissues with QIAamp® genomic DNA kit (Quagen, Hilden, Germany) according to the manufacturer’s protocol. For validation of genotyping, ~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
Because only KRT81 rs3660 was consistently associated with the survival outcome of patients in discovery and validation sets among 154 SNPs investigated, we decided to examine whether rs3660 is a functional SNP. We investigated whether miR-17-5p and miR-20b, which are known to be overexpressed in lung cancer [6] and predicted to bind to KRT81 mRNA (http://www.microrna.org/microrna/), change KRT81 expression on its 3′UTR including rs3660G>C by luciferase report assay (supplementary Methods, available at Annals of Oncology online).

RNA preparation and quantitative reverse transcription-PCR
KRT81 mRNA expression was examined by quantitative reverse transcription-PCR (qRT-PCR) in 69 pairs of tumor and paired non-malignant lung tissues (supplementary Methods, available at Annals of Oncology online).

Statistical analysis
Hardy–Weinberg equilibrium was tested using a goodness-of-fit χ² test with 1 degree of freedom. Overall survival (OS) was measured from the day of...
surgery to the date of the last follow-up or until the date of death. The survival estimates were calculated using the Kaplan–Meier method. The difference in OS 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 (≥65 years versus <65), gender (male versus female), smoking status (ever versus never), tumor histology (squamous versus non-squamous), and pathologic stage (II–IIIA versus I). 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).

**Results**

**Patient characteristics and clinical predictors**

The clinical and pathologic characteristics of patients in the discovery and validation sets and the association with OS are shown in Table 1. Upon univariate analysis, pathologic stage was significantly associated with OS in both sets (log-rank P [P_{LR}] = 6.3 × 10^{-5} and 0.01, respectively). Age and gender were also associated with OS in the validation set (P_{LR} for OS = 0.01 and 0.03, respectively).

**Associations between SNPs and survival outcomes**

One hundred fifty-four miRNA binding site polymorphisms were analyzed in this study. The SNP ID, gene information, miRNA, and minor allele frequencies are shown in supplementary Table S1, available at Annals of Oncology online. Of the 154 SNPs analyzed in the discovery set, 14 SNPs listed in supplementary Table S2, available at Annals of Oncology online were significantly associated with survival outcomes when adjusted for age, gender, smoking status, tumor histology, and pathologic stage. Among the fourteen poly-miRTSs, KRT81 rs3660G>C was found to be associated with survival outcomes in the same direction as the discovery set in an independent validation set. In combined analysis, patients with rs3660 GC + CC genotype had a significantly better overall survival (OS) compared with those with GG genotype [adjusted hazard ratio (aHR) for OS, 0.65; 95% confidence interval (CI), 0.50–0.85; P = 0.001; Table 2 and Figure 1].

**Effect of rs3660G>C on the binding of miR-17-5p and miR-20b to KRT81 3′UTR**

The rs3660 G-to-C change produces the loss of a nucleotide binding in the seed sequence region of these miRNAs as showed in Figure 2A. MFE was predicted using RNAhybrid (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/submission.html; Figure 2A): −23.40 and −19.10 kCal/mol for binding of miR-20b to rs3660 G and C alleles, respectively, and −22.70 and −18.40 kCal/mol for binding of miR-17-5p to rs3660 G and C alleles, respectively. To investigate whether rs3660G>C in the 3′UTR of KRT81 can modulate the binding of miR-17-5p and miR-20b and thereby alter the expression of KRT81, we generated psiCHECK2-KRT81 constructs containing 3′UTR of KRT81 including rs3660G>C and co-transfected the constructs into 293T cells and H1299 cells with miR-20b and miR-17-5p. As shown in Figure 2B and C, the Renilla luciferase activity was significantly decreased by both hsa-miR-17-5p and hsa-miR-20b compared with negative control miRNA in both rs3660G and rs3660C (P < 0.001, for all comparisons; Figure 2B and C). In addition, the Renilla luciferase activity was significantly higher in rs3660C construct than in rs3660G construct when the cells were co-transfected with miR-17-5p and miR-20b in 293T cells (P = 0.05 and 0.01, respectively; Figure 2B).

**Table 1. Univariate analysis for overall survival by clinicopathologic features of the discovery and validation cohorts**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Discovery cohort</th>
<th>Validation cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of cases</td>
<td>No. of death (%)</td>
</tr>
<tr>
<td>Overall</td>
<td>357</td>
<td>54</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤64</td>
<td>195</td>
<td>64 (32.8)</td>
</tr>
<tr>
<td>&gt;64</td>
<td>162</td>
<td>66 (40.7)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>273</td>
<td>109 (39.9)</td>
</tr>
<tr>
<td>Female</td>
<td>84</td>
<td>21 (25.0)</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>84</td>
<td>23 (27.4)</td>
</tr>
<tr>
<td>Ever</td>
<td>273</td>
<td>107 (39.2)</td>
</tr>
<tr>
<td>Histological type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous cell ca.</td>
<td>189</td>
<td>68 (36.0)</td>
</tr>
<tr>
<td>Adenoca.</td>
<td>163</td>
<td>60 (36.8)</td>
</tr>
<tr>
<td>Large cell ca.</td>
<td>5</td>
<td>2 (40.0)</td>
</tr>
<tr>
<td>Pathologic stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>184</td>
<td>48 (26.1)</td>
</tr>
<tr>
<td>II–IIIA</td>
<td>173</td>
<td>82 (47.4)</td>
</tr>
</tbody>
</table>

*Row percentage.

bFive-year OS rate, proportion of survival derived from Kaplan–Meier analysis.
**Table 2. OS according to the rs3660 G>C genotypes in the discovery and validation cohorts, and the combined cohort**

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Discovery cohort</th>
<th>Validation cohort</th>
<th>Combined cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5Y-OSRb HR (95% CI) p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5Y-OSRb HR (95% CI) p&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5Y-OSRb HR (95% CI) p&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>No. of deaths/patients (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>No. of deaths/patients (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>No. of deaths/patients (%)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>GG</td>
<td>90/211 (42.7) 48 1 81/275 (29.5) 65 1 171/486 (35.2) 58 1</td>
<td>171,046 (35.2) 70 0.67 (0.46–1.00) 0.05 0.08 40/176 (22.7) 74 0.68 (0.46–0.99) 0.04 0.07 75/302 (24.8) 70 0.65 (0.50–0.86) 0.002 0.01</td>
<td></td>
</tr>
<tr>
<td>GC</td>
<td>35/126 (27.8) 63 0.67 (0.46–0.99) 0.04 0.07 40/176 (22.7) 74 0.68 (0.46–0.99) 0.04 0.07 75/302 (24.8) 70 0.65 (0.50–0.86) 0.002 0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>5/20 (25.0) 69 0.69 (0.28–1.71) 0.42 0.42 6/27 (22.2) 77 0.69 (0.30–1.59) 0.38 0.42 11/47 (23.4) 74 0.67 (0.36–1.23) 0.20 0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dominant</td>
<td>40/146 (27.4) 64 0.67 (0.46–0.99) 0.04 0.07 46/203 (22.7) 74 0.68 (0.47–0.97) 0.04 0.07 86/349 (24.6) 70 0.65 (0.50–0.85) 0.001 0.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Discussion**

We conducted a two-stage study to investigate the associations between SNPs located in the miRNA target sites and survival of patients with early-stage non-small-cell lung cancer (NSCLC) to identify genetic variations associated with prognosis of patients with surgically resected early-stage NSCLC. The KRT81 rs3660 G>C was replicated across both stages of the study in terms of survival. In addition, this study provides evidence that the KRT81 rs3660 G>C is a functional SNP. These findings suggest that KRT81 rs3660 G>C could be used as prognostic markers for early-stage NSCLC.

In the present study, KRT81 rs3660 G>C was associated with the prognosis of patients with surgically resected early-stage NSCLC. In vitro luciferase assay suggested that the KRT81 rs3660 G-to-C change reduced binding efficiency of miR-20b and miR-17-5p to KRT81 mRNA, leading to decreased translational repression, thereby increased KRT81 expression in rs3660 C allele compared with rs3660 G allele. In addition, the same association was observed between the relative expression level of KRT81 and the rs3660 G>C genotypes in tumor tissues. However, the biologic mechanism of the observed associations between the SNPs and survival outcomes remains unclear because the functional role of KRT81 in development and progression of lung cancer is unknown. Keratins are the primary intermediate filament proteins of epithelial cells, and more than 50 keratins have been identified in human. They are expressed in all types of epithelial cells with highly specific patterns related to the epithelial type and the stage of cellular differentiation [15]. Since keratins also exhibit characteristic expression patterns in human tumors, they are extensively used as immunohistochemical diagnostic markers for carcinomas. As part of the epithelial cytoskeleton, keratins are important for the mechanical stability and integrity of epithelial cells and tissues [15, 16]. In addition, keratins have also been recognized as regulators of cellular functions and intracellular signaling such as tissue...
Notably, emerging evidences have suggested that keratins are involved in malignant transformation, cancer cell invasion, and metastasis [16, 17]. The KRT81 encodes for KRT81 protein, also known as Hb-1, a type of hair keratin. In the present study, KRT81 expression was significantly higher in tumor tissues than in non-malignant lung tissues, suggesting a possibility of the oncogenic role of KRT81 in carcinogenesis. Although a study reported that breast carcinomas ectopically express KRT81 [18], the functional role of KRT81 in human malignancy is largely unknown. Further studies on the biologic function of KRT81, such as analyzing the effect of suppression of KRT81 gene expression using siRNA in cell lines with KRT81 overexpression, are needed to understand their roles in determining lung cancer prognosis. In addition, further studies are needed to clarify the association between the SNPs and prognosis of patients with surgically resected NSCLC.

### Table 1: SNP Analysis in NSCLC Patients

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. of event/no. at risk</th>
<th>Overall survival (probability)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td>0/211</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>GC</td>
<td>0/126</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>0/20</td>
<td>0.50</td>
<td></td>
</tr>
</tbody>
</table>

**A. Discovery cohort**

**B. Replication cohort**

**C. Combined analysis**

*Figure 1.* Kaplan–Meier plot of overall survival curves according to KRT81 rs3660G>C genotype in discovery cohort (A), replication cohort (B) and combined cohort (C). *P* values, log-rank test.
molecular pathogenesis in different cancers, might be signi-
cificant across studies, several genetic and environmental factors relevant to each cell and tissue type of cancers. This might play different roles in different cancer cell and tissue types, the effect of genetic polymorphism in miRNA target sites on tumorigenesis might be different in various cancers depending on each cell and tissue type of cancers.

In this study, the association between KRT81 rs3660G>C and survival outcomes was replicated across both set of the study, which would largely reduce false-positive findings in the genetic association study [22, 23]. In addition, functional study results support the association of KRT81 rs3660G>C and survival outcomes. However, several limitations in the present study should be considered. First, SNPs which were not applicable to Sequenom’s MassARRAY system might have been significantly associated with the survival of patients. Second, the modest sample size of both cohorts does not have optimal statistical power for discovering and validating the association, so the observed P-value did not reach a more stringent level of statistical significance that would avoid most of the false-positive associations arising from multiple comparisons. However, the association had similar effect size with the same direction in both study sets. In addition, the association showed higher level of significance (P = 0.001) in the combined analysis including larger population, supporting the credibility of statistical significance of the association. Future studies with larger number of patients are required to validate our results.

In conclusion, this study shows that KRT81 rs3660G>C influences survival outcomes of patients with surgically resected early-stage NSCLC. Further studies are required to confirm the effect of this SNP in a larger population with diverse ethnicity

Figure 2. Functional analysis of the rs3660G>C in the KRT81 3’UTR. (A) The predicted effect of allelic variant at rs3660 on hsa-miR-20b and hsa-miR-17-5p recognition and the construct of psiCHECK2-KRT81 3’UTR. MFE was predicted using RNAhybrid (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/submission.html). Renilla luciferase assay for miRNA suppression on rs3660G>C polymorphism using 293T (B) and H1299 cells (C). Cells were co-transfected with hsa-miR-17-5p, hsa-miR-20b, or negative control miRNA, and psiCHECK2 plasmid containing KRT81 3’UTR with G or C allele. Each bar represents mean ± SEM of Renilla luciferase activity normalized to firefly luciferase activity. Experiments were carried out in triplicate. P-value, a Student’s t-test.

*P < 0.001, compared with negative control miRNA.
and to understand the biologic function of KRT81 in the development and progression of lung cancer.

**funding**

This work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (NRF-2011-0013670), in part by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2013R1A1A1009443).

**disclosure**

The authors have declared no conflicts of interest.

**references**