MOLECULAR CARCINOGENESIS 49:913–921 (2010)

Association of a Common *AGO1* Variant With Lung Cancer Risk: A Two-Stage Case–Control Study

Jong-Sik Kim,[1] Yi Young Choi,[1] Guang Jin,[1,2] Hyo-Gyoung Kang,[1] Jin-Eun Choi,[1] Hyo-Sung Jeon,[1,3] Won-Kee Lee,[4] Dong-Sun Kim,[5] Chang Ho Kim,[6] Young Jin Kim,[7] Ji-Woong Son,[8] Tae Hoon Jung,[6] and Jae Yong Park[1,3,6]*

[1]*Department of Biochemistry and Cell Biology, School of Medicine, Kyungpook National University, Daegu, Republic of Korea*
[2]*Department of Pharmacology, Yanbian University School of Basic Medicine, Yanji, Jilin Province, China*
[3]*Cancer Research Institute, Kyungpook National University Hospital, Daegu, Korea*
[4]*Department of Preventive Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea*
[5]*Department of Anatomy, School of Medicine, Kyungpook National University, Daegu, Republic of Korea*
[6]*Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea*
[7]*Department of Chest Surgery, Konyang University Hospital, Daejeon, Republic of Korea*
[8]*Department of Internal Medicine, Konyang University Hospital, Daejeon, Republic of Korea*

Based on the important role of microRNA (miRNA) biosynthesis genes in carcinogenesis, we hypothesized that polymorphisms in the miRNA biosynthesis genes may modulate susceptibility to lung cancer. To test this hypothesis, we conducted a two-stage study to evaluate the associations between single nucleotide polymorphisms (SNPs) in the miRNA biosynthesis genes and the risk of lung cancer. In stage 1 of the study, 24 SNPs in the 11 miRNA biosynthesis genes (*DROSHA*, *DGCR8*, *RAN*, *XPO5*, *DICER*, *AGO1*, *AGO2*, *HIWI*, *GEMIN3*, *GEMIN4*, and *TRBP*) were genotyped in 100 lung cancer patients and 100 healthy controls using a sequenome mass spectrometry-based genotyping assay. One promising SNP (*AGO1* rs636832A > G) was selected for stage 2 of the study, and genotyped by a melting-curve analysis using fluorescence-labeled hybridization probes in an independent set of 552 cases and 552 controls. The *AGO1* rs636832A > G exhibited highly consistent results between the two stages of the study. In combined analysis, the 636832A > G was associated with a significantly decreased risk of lung cancer in a dose-dependent manner ($P_{trend} = 6.0 \times 10^{-4}$). Individuals with at least one rs636832G allele were at a significantly decreased risk of lung cancer compared with those with the AA genotype (adjusted odds ratio = 0.67, 95% confidence interval = 0.53–0.84, $P = 4.0 \times 10^{-4}$). This finding suggests that the *AGO1* rs636832A > G might be a useful marker for determining the susceptibility to lung cancer and that the *AGO1* gene might be involved in the development of lung cancer.
© 2010 Wiley-Liss, Inc.

Key words: genetic susceptibility; microRNA; lung cancer; polymorphisms

INTRODUCTION

MicroRNAs (miRNAs) are noncoding, single-stranded RNAs of ~22 nucleotides that repress gene expression by interacting with messenger RNA, by inhibiting mRNA translation or by inducing mRNA cleavage, depending on the degree of their complementarity to target sequences [1–3]. To date, more than 500 human miRNAs have been identified and up to 30% of protein-coding human genes are estimated to be regulated by miRNAs [2–5]. A growing body of evidence has revealed that miRNAs regulate a variety of biological processes, such as organ development, cell proliferation, cell differentiation, and apoptosis [6,7]. Moreover, several studies indicate that miRNAs play an important role in the development and progression of human cancers, including lung cancer, by regulating the expression of proto-oncogenes or tumor suppressor genes [8,9].

MiRNAs are generated by a two-step processing pathway. MiRNA genes are generally transcribed by RNA polymerase II to form large primary miRNAs, which are capped and polyadenylated [10]. After transcription, primary miRNAs are cropped in the nucleus by the microprocessor machinery, which

Abbreviations: miRNA, microRNA; AGO, Argonaute; TRBP, trans-activating response RNA binding protein; SNP, single nucleotide polymorphism; KNUH, Kyungpook National University Hospital; MAF, minor allele frequency; LD, linkage disequilibrium; OR, odds ratio; CI, confidence interval.

Jong-Sik Kim and Yi Young Choi contributed equally to this study.

*Correspondence to: Department of Internal Medicine, Kyungpook National University Hospital, Samduk 2a 50, Daegu 700-412, Republic of Korea.

Received 18 January 2010; Revised 2 June 2010; Accepted 24 June 2010

DOI 10.1002/mc.20672

Published online 18 August 2010 in Wiley Online Library (wileyonlinelibrary.com)

© 2010 WILEY-LISS, INC.

includes the RNAse III DROSHA and the double-strand RNA binding protein DGCR8/PASHA [11]. This processing releases ~70 nucleotide precursor miRNAs [12], which are exported to the cytoplasm by the RAN-GTP-dependent transporter exportin 5 (XPO5) [13]. In the second step, precursor miRNA is processed by another RNAse III enzyme DICER to generate double-stranded ~22 nucleotide miRNA molecules [14]. One strand of the miRNA duplex is incorporated into a ribonucleoprotein effector complex, termed RNA-induced silencing complex (RISC), which includes Argonaute proteins (AGO1, AGO2, and HIWI), human immunodeficiency virus transactivating response RNA binding protein (TRBP), GEMIN3, and GEMIN4 [15,16].

In addition to deregulation of miRNAs, altered regulation of the miRNA processing genes has been implicated in tumorigenesis [17–20]. It has been reported that aberrant expression of DICER modifies the development of lung cancers [17]. Knockdown of the miRNA processing enzymes (DROSHA and DICER) in cell lines has been shown to globally reduce miRNA expression, and thereby enhance cellular transformation and tumorigenesis [18]. The Argonaute proteins have also been associated with various cancers [19]. TRBP has also been shown to have oncogenic potential, and can induce tumors in nude mice [21]. Taken together, these emerging lines of evidence suggest that miRNA processing proteins may play important roles in cancer development and progression.

Single nucleotide polymorphisms (SNPs) are the most common human genetic variants and may contribute to an individual's susceptibility to cancer. Many studies have demonstrated that some SNPs affect either the expression or activities of various enzymes, and are therefore associated with cancer risk [21,22]. Based on the important role of miRNA processing genes in carcinogenesis, we hypothesized that SNPs in the miRNA processing genes may influence their expression or activity, thereby modulating susceptibility to lung cancer. To test this hypothesis, we conducted a two-stage study to evaluate the associations between SNPs in the miRNA processing genes (*DROSHA*, *DGCR8*, *RAN*, *XPO5*, *DICER*, *AGO1*, *AGO2*, *HIWI*, *GEMIN3*, *GEMIN4*, and *TRBP*) and lung cancer risk.

METHODS

Study Design and Population

A two-stage study design was used to comprehensively evaluate SNPs in the miRNA processing genes in relation to lung cancer risk and then to validate promising associations in a second independent patient population (Table 1). Cases ($n = 100$) in stage 1 of the study were randomly selected from 432 patients who were newly diagnosed with lung cancer at the Kyungpook National University Hospital (KNUH) in Daegu, Korea between January 2001 and February 2002. The control subjects ($n = 100$) were randomly selected from a pool of healthy volunteers who visited the general health check-up center at KNUH during the same period. The details of this study population, including the participation rate, are described elsewhere [22].

Table 1. Characteristics of the Study Population

	Stage 1			Stage 2		
Variables	Cases ($n = 100$)	Controls ($n = 100$)	*P*	Cases ($n = 552$)	Controls ($n = 552$)	*P*
Age (yr)	61.5 ± 10.9	61.5 ± 11.3	1.00[a]	60.9 ± 8.5	60.5 ± 9.0	0.52[a]
Sex						
Male	80 (80.0)[b]	80 (80.0)	1.00[c]	453 (82.1)	453 (82.1)	1.00[c]
Female	20 (20.0)	20 (20.0)		99 (17.9)	99 (17.9)	
Smoking status			0.001[c]			<0.0001[c]
Current	69 (69.0)	52 (52.0)		355 (64.3)	259 (46.9)	
Former	5 (5.0)	23 (23.0)		99 (17.9)	153 (27.7)	
Never	26 (26.0)	25 (25.0)		98 (17.8)	140 (25.4)	
Pack-years[d]	37.8 ± 18.8	31.2 ± 17.8	0.03[a]	39.8 ± 17.9	32.6 ± 17.3	<0.0001[a]
Histological types						
Squamous cell ca.	40 (40.0)			247 (44.8)		
Adenoca.	40 (40.0)			204 (37.0)		
Large cell ca.	0 (0.0)			9 (1.6)		
Small cell ca.	20 (20.0)			92 (16.7)		

[a] *t*-Test.
[b] Numbers in parenthesis, column percentage.
[c] χ^2 test.
[d] In current and former smokers.

The stage 2 study population consisted of 552 lung cancer cases and 552 healthy controls who were enrolled using the same method as described for our previous study. Briefly, the eligible cases included all the patients who were newly diagnosed with primary lung cancer between January 2003 and June 2004 at the KNUH and agreed to this study (participation rate, 97.2%). There were no gender, histologic, or stage restrictions; however, those patients $\geq$75 yr of age or those who had a prior history of cancer were excluded from this study. The control subjects were randomly selected from a pool of healthy volunteers who visited the general health check-up center at KNUH during the same period. All the subjects enrolled in this study (cases and controls) were ethnic Koreans who resided in Daegu City or the surrounding regions. This study was approved by the Institutional Review Board of the Kyungpook National University Hospital, and written informed consent was obtained from each participant.

SNPs and Genotyping

Twenty-four SNPs in the 11 miRNA processing genes were evaluated in stage 1 of the study. The 24 SNPs were reported as potential functional polymorphisms through extensive mining of the databases in the previous study [23], and genotyped using a sequenome mass spectrometry-based genotyping assay. Of the 24 SNPs, 6 variants for which the minor allele frequency (MAF) was $\leq$5% in the controls (*DGCR8* rs417309, *DICER* rs13078, *GEMIN3* rs197388 and rs197414, *GEMIN4* rs1062923, and *TRBP* rs784567) were excluded from further analysis. Therefore, 18 SNPs in 10 genes were retained for an association analysis. The nucleotide substitutions, amino acid changes, and reference SNP identification numbers of the 24 SNPs are shown in Table 2.

In stage 2 of the study, 1 promising SNP (*AGO1* rs636832) was genotyped by a melting-curve analysis using fluorescence-labeled hybridization probes (LightCycler 480®: Roche Diagnostic, Mannheim, Germany). Genotyping analysis was performed ''blind'' with respect to the case/control status in order to ensure quality control. Approximately 10% of samples were randomly selected to be genotyped again by a different investigator, and the results showed 100% concordance.

Statistical Analysis

The Hardy–Weinberg equilibrium was tested by comparing the observed and expected genotype frequencies using a χ^2 test with 1 degree of freedom for the cases and controls separately. Allele and genotype frequencies in the cases and controls were compared using a χ^2 test. In order to minimize type II errors, the SNPs for the stage 2 validation study were selected using a cut-off α error level of 0.10 ($P < 0.10$) for the difference in either allele or genotype frequency between the cases and controls in the stage 1 of the study. The linkage disequilibrium (LD) between SNPs was measured by using HaploView (http://broad.mit.edu/mpg/haploview). LD blocks were inferred from the definition proposed by Gabriel et al. [24]. The haplotypes and their frequencies were estimated based on a Bayesian algorithm using the Phase program [25]. The cancer risk associated with the genotypes and haplotypes was estimated as an odds ratio (OR) and 95% confidence interval (CI) using logistic regression. Crude ORs and ORs adjusted for possible confounders (gender as a nominal variable; age and pack-years smoked as continuous variables) were calculated. A heterogeneity test was performed to compare the ORs of the two study samples and different subgroups. We also performed a stratified analysis by age, gender, smoking status, and tumor histology to further explore the association between genotypes/haplotypes and the risk of lung cancer in each stratum. For the gene–smoking interaction analyses, we used the following three approaches to evaluate the consistency of results: (i) stratification analysis; (ii) gene–smoking joint effects; and (iii) a logistic regression model, including the interaction term between genotype/haplotype and smoking. For these analyses, the subjects were categorized into three groups according to the level of smoking exposure: never-smokers; light smokers (ever-smokers $\leq$33 pack-years) and heavy smokers (ever-smokers >33 pack-years). All the analyses were performed using Statistical Analysis Software for Windows, version 9.1.3 (SAS Institute, Cary, NC).

RESULTS

Of the SNPs examined in the stage 1 study, the MAF of the *AGO1* rs636832A > G was significantly different between the cases and controls (0.184 vs. 0.268, $P = 0.046$; Table 2). Individuals with the rs636832 AG or GG genotype were at a significantly decreased risk of lung cancer compared those with the AA genotype (adjusted OR = 0.53, 95% CI = 0.30–0.96, $P = 0.035$; Table 3).

To confirm the observed association of the *AGO1* rs636832A > G with the risk of lung cancer, we undertook a replication study using an independent sample. Agreement with the results of the stage 1 study, the genotype distribution, and MAF of the rs636832A > G among the cases was significantly different from the controls ($P = 0.02$ and $P = 0.004$, respectively). In addition, there was no evidence of heterogeneity in ORs between the two studies ($P = 0.41$, 0.72, and 0.41 for the AG, GG, and AG + GG genotypes vs. the AA genotype, respectively; Table 3). In combined analysis of the two stages of the study, the rs636832A > G was associated with a significantly decreased risk of lung cancer in a dose-dependent manner ($P_{\text{trend}} = 6.0 \times 10^{-4}$). Individuals with at least one rs636832G allele were at a significantly decreased risk of lung cancer

Table 2. Information for 24 Genotyped Polymorphisms, and Genotype and Allele Frequencies in Stage 1 Study Population

	Polymorphism[a]			Genotypes, n[b]			MAF			P for HWE test		MAF in healthy populations[a]	
Gene	ID no.	Region	Base change	Cases	Controls	P[c]	Cases	Controls	P[c]	Cases	Controls	Asian	European
DROSHA	rs6877842	5′UTR	G > C	81/11/1	84/8/1	0.77	0.07	0.054	0.52	0.384	0.136	0.022	0.181
	rs10719	3′UTR	T > C	59/29/9	52/38/7	0.39	0.242	0.268	0.56	0.067	0.987	0.261	0.775
DGCR8	rs3757	3′UTR	G > A	60/27/7	60/24/6	0.92	0.218	0.200	0.67	0.126	0.114	0.233	0.310
	rs417309	3′UTR	G > A	90/8/0	88/9/0	0.78	0.041	0.046	0.79	0.674	0.632	0.083	0.042
	rs1640299	3′UTR	T > G	58/33/7	52/40/5	0.52	0.240	0.258	0.68	0.449	0.444	0.318	0.467
RAN	rs14035	3′UTR	C > T	65/23/5	52/33/5	0.20	0.177	0.239	0.15	0.141	0.937	0.261	0.267
XPO5	rs11077	3′UTR	A > C	88/12/0	87/9/3	0.18	0.060	0.076	0.53	0.523	0.000	0.080	0.400
DICER	rs13078	3′UTR	T > A	8711/0	88/9/0	0.65	0.056	0.046	0.66	0.556	0.632	0.033	0.142
	rs3742330	3′UTR	A > G	31/45/24	41/40/19	0.32	0.465	0.390	0.13	0.339	0.111	0.440	0.117
AGO1	rs636832	Intron	A > G	65/30/3	50/42/5	0.11	0.184	0.268	0.046	0.837	0.308	0.375	0.908
	rs595961	Intron	G > A	65/29/4	69/28/0	0.13	0.189	0.144	0.24	0.738	0.097	0.273	0.850
AGO2	rs4965280	Promoter	C > A	83/15/1	76/21/0	0.32	0.086	0.108	0.45	0.729	0.232	—	—
HIWI	rs1106042	Exon	G > A (K527R)	70/24/3	73/23/1	0.58	0.155	0.129	0.47	0.597	0.581	0.133	0.075
GEMIN3	rs197388	Promoter	T > A	84/8/0	84/6/0	0.61	0.043	0.033	0.62	0.663	0.744	0.034	0.292
	rs197412	Exon	T > C (T636I)	28/53/11	35/45/10	0.48	0.408	0.361	0.36	0.064	0.428	0.409	0.466
	rs197414	Exon	C > A (S693R)	100/0/0	98/0/1	—	0.000	0.010	—	—	—	0.000	0.192
GEMIN4	rs910924	Promoter	C > T	72/20/1	69/20/1	0.99	0.118	0.122	0.91	0.765	0.735	0.136	0.350
	rs2740348	Exon	G > C (E450Q)	74/18/0	71/19/0	0.80	0.098	0.106	0.81	0.298	0.263	0.182	0.192
	rs910925	Exon	G > C (G579A)	41/45/11	47/39/13	0.61	0.345	0.328	0.72	0.798	0.288	0.386	0.525
	rs4968104	Exon	T > A (V593E)	74/21/2	73/22/2	0.99	0.129	0.134	0.88	0.725	0.822	0.151	0.345
	rs3744741	Exon	C > T (Q684R)	56/38/6	60/34/5	0.80	0.250	0.222	0.51	0.894	0.948	0.136	0.133
	rs1062923	Exon	T > C (T739I)	91/1/0	90/0/0	—	0.005	0.000	—	—	—	0.000	0.175
	rs7813	Exon	T > C (C1033R)	42/45/11	47/40/12	0.74	0.342	0.323	0.70	0.839	0.447	0.389	0.525
TRBP	rs784567		C > T	98/0/0	97/0/0	—	0.000	0.000	—	—	—	0.000	0.483

HWE, Hardy–Weinberg equilibrium; ID, identification number; MAF, minor allele frequency.
[a]Information about polymorphisms and IDs, and MAF in other ethnic populations (Asian and European) were obtained from NCBI database (http://www.ncbi.nlm.nih.g).
[b]Number of major allele homozygotes/number of heterozygotes/number of minor allele homozygotes.
[c]Two-sided chi-squared test either genotype distributions or allele frequencies between the cases and controls.

Table 3. Association Between *AGO1* rs636832A > G Genotypes and Lung Cancer Risk

	Stage 1			Stage 2					Stage 1 + stage 2			
Genotype	No. of subjects[a]	*P*	Adjusted OR (95% CI)[b]	*P*[b]	No. of subjects[a]	*P*	Adjusted OR (95% CI)[b]	*P*[b]	P_H^c	No. of subjects[a]	*P*	Adjusted OR (95% CI)[b]	*P*[b]
AA	65/50	0.108	1		341/295	0.02	1			406/345	0.003	1.00	
AG	30/42		0.54 (0.30–0.99)	0.046	178/212		0.71 (0.55–0.92)	0.01	0.41	208/254		0.67 (0.54–0.86)	0.001
GG	3/5		0.47 (0.11–2.07)	0.32	33/45		0.62 (0.38–1.02)	0.06	0.72	36/50		0.60 (0.38–0.96)	0.03
P_{trend}			0.042				0.005					0.0006	
AA	65/50	0.036	1		341/295	0.005	1			406/345	0.0007	1.00	
AG + GG	33/47		0.53 (0.30–0.96)	0.035	211/257		0.70 (0.54–0.89)	0.004	0.41	250/304		0.67 (0.53-0.84)[d]	0.0004

[a]Case no./control no.
[b]Odds ratios (ORs), 95% confidence intervals (CIs) and their corresponding *P*-values were calculated with logistic regression analysis, adjusted with age, gender, and pack-years of smoking.
[c]Wald test for homogeneity of adjusted ORs between the two study samples.
[d]Mantel–Haenszel estimate of the OR (95% CI) controlled for the confounding effect of smoking status = 0.69 (0.55–0.86).

compared with those with the AA genotype (adjusted OR = 0.67, 95% CI = 0.53–0.84, $P = 4.0 \times 10^{-4}$).

The association between the *AGO1* rs636832A > G and the risk of lung cancer was further examined after stratifying the subjects according to age, smoking status, and histologic types of lung cancer. The protective effect of the rs636832G allele on the risk of lung cancer was similar in younger and older individuals (*P*-value of test for homogeneity = 0.86, data not shown). When stratified according to the smoking status, the protective effect of the rs636832G allele was significant in the smokers (adjusted OR = 0.67, 95% CI = 0.53–0.91, $P = 0.002$), but not in never-smokers ($P = 0.16$). When the ever-smokers were dichotomized by the pack-years of smoking, the protective effect of the rs636832G allele was significant in heavy smokers (adjusted OR = 0.60, 95% CI = 0.41–0.86, $P = 0.006$), whereas there was no significant association in light smokers ($P = 0.11$). Lung cancers are comprised of different histologic types, and the carcinogenesis pathways are different according to the histologic type of lung cancer. Therefore, the effect of the rs636832A > G genotypes on the risk of lung cancer was estimated according to the histologic type of lung cancer. The protective effect of the rs636832G allele was more pronounced in patients with adenocarcinoma (adjusted OR = 0.56, 95% CI = 0.41–0.77, $P = 3 \times 10^{-4}$, Table 4).

In addition to the stratification analyses, the joint effect of the rs636832A > G genotypes and smoking status on the risk of lung cancer was also determined (Table 5). When the group of never-smokers with the protective rs636832G allele was used as the reference group, the group of heavy smokers with the rs636832 AA genotype was found to have the highest risk of lung cancer (adjusted OR = 5.30, 95% CI = 3.13–9.00, $P < 1 \times 10^{-4}$). Nevertheless, we did not find statistically significant evidence of a gene–smoking interaction in the multivariate logistic regression analysis ($P = 0.29$ for multiplicative interaction).

DISCUSSION

We evaluated the potential associations between genetic variations in the miRNA biosynthesis genes and the risk of lung cancer in a two-stage case–control study. Of the 24 SNPs examined, one SNP (*AGO1* rs636832A > G) was convincingly replicated across both stages of the study. In the combined analysis of two stages of the study, the rs636832 AG or GG genotype was associated with a 33% reduced risk of lung cancer compared to the AA genotype. This finding suggests that the *AGO1* rs636832A > G might be a useful marker for determining the susceptibility to lung cancer, and that the *AGO1* gene might be involved in the development of lung cancer. This is the first study to evaluate the associations of miRNA biosynthesis gene polymorphisms in relation to lung cancer risk.

Table 4. Association Between *AGO1* rs636832A > G Genotypes and Lung Cancer According to Age, Smoking Status, and Histological Types of Lung Cancer

Variables	Cases		Controls		Adjusted OR (95% CI) for AG + GG vs. AA	*P*	P_H[a]
	AA	AG + GG	AA	AG + GG			
Smoking status							
Never	73 (59.4)	50 (40.6)	86 (52.1)	79 (47.9)	0.71 (0.44–1.14)[†]	0.16	0.85
Ever	333 (63.2)	194 (36.8)	259 (53.5)	225 (46.5)	0.67 (0.53–0.91)[†]	0.002	
Pack-years of smoking							
≤36	136 (63.0)	80 (37.0)	164 (55.8)	130 (44.2)	0.75 (0.52–1.07)[†]	0.11	0.38
>36	197 (63.3)	114 (36.7)	95 (50.0)	95 (50.0)	0.60 (0.41–0.86)[†]	0.006	
Histological types							
Squamous cell ca.	174 (61.1)	111 (38.9)	345 (53.2)	304 (46.8)	0.72 (0.53–0.97)[‡]	0.03	0.53
Adenoca.	161 (66.0)	83 (34.0)			0.56 (0.41–0.77)[‡]		0.0003
Small cell ca.	68 (60.7)	44 (39.3)			0.69 (0.45–1.05)[‡]	0.08	

ORs, 95% confidence intervals (CIs) and their corresponding *P*-values were calculated by unconditional logistic analysis, [†]adjusted for age and gender; and [‡]adjusted for age, gender and pack-years of smoking.
[a]Wald test for heterogeneity of adjusted odds ratios (ORs) between groups.

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 [26,27]. In addition, the observed *P*-values ($P_{trend} = 6 \times 10^{-4}$; and *P* in a dominant model for the variant rs636832G allele $= 4 \times 10^{-4}$) were 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 [26]. Moreover, the population in Korea is genetically homogenous, reducing the risk of confounding due to the population stratification. In addition, other confounding factors such as selection bias, information bias, and observer bias that can result in a false positive association, were minimized. Taken as a whole, these strengthen the reliability of our finding of an association between the *AGO1* rs636832A > G and lung cancer risk.

The interpretation of our data was limited by the lack of published evidence showing that AGO1 contributes to the development of lung cancer. Emerging data indicate an important role for miRNA biogenesis in tumorigenesis. Impaired miRNA processing through in vitro targeting Drosha, DGCR8, and Dicer has been shown to accelerate oncogenic transformation in mouse lung cancer cell cells and in vivo tumor development [18]. In addition, a recent study has shown that the global repression of miRNAs in human cancers does not coincide with reductions in the primary miRNA transcripts, suggesting that altered regulation of the miRNA processing machinery plays a critical role in carcinogenesis [28]. AGO1 is a crucial component of the RISC complex with AGO2 and DICER, and thus plays an important role in miRNA-mediated gene regulation [16,19,29]. In addition, it has been reported that AGO1 interacts with the Wingless/Wnt pathway that involved in lung tumorigenesis [30]. Moreover, the *AGO1* gene is located at chromosome 1p34-35, which is frequently deleted in human cancers, including lung cancer [31,32]. Thus, it is possible that AGO1 contributes to the development of lung cancer by either modulating miRNA-mediated gene regulation or interacting with the Wingless/Wnt pathway. However, this hypothesis must be verified in future studies.

The rs636832 is located in intron of the *AGO1* gene, indicating that the association of the rs636832 with lung cancer may be due to LD with other functional variant(s) rather than a direct effect of the rs636832. In order to identify the variant(s) that may be in strong LD with the rs636832 that could be actually responsible for the alteration in lung cancer risk, we screened and estimated the LDs of all the common variants (with a MAF ≥0.01) within an approximately 100-kb long region around the rs636832 locus (~50 kb upstream and ~50 kb

Table 5. Interaction of *AGO1* rs636832A > G Genotype and Tobacco Smoking on Lung Cancer Risk

Smoking status	Genotype AG + GG	OR (95% CI)[a]	AA	OR (95% CI)[a]
Never smoker	50/79[b]	1.00 (reference)	73/86	1.41 (0.88–2.28)
Smoker				
≤36 pack-years	80/130	1.47 (0.87–2.50)	136/164	1.99 (1.20–3.29)[c]
>36 pack-years	114/95	3.10 (1.80–5.35)[d]	197/95	5.30 (3.13–9.00)[d]

$P = 0.29$ for the interaction term between genotype and smoking in the multivariate model.
[a]Odds ratios (ORs) and 95% confidence interval (CIs) were calculated by logistic regression, with AG + GG genotype in never-smokers as reference group and adjusted for age and gender.
[b]Case no./control no.
[c]$P = 0.008$.
[d]$P < 0.0001$.

downstream of the rs636832 locus) using the HapMap JPT data. Notably, we found that the rs636832 is located within a large haplotype block (Figure 1). Therefore, it is possible that the functional variant(s) in strong LD with the rs636832 may be the causal variant(s) for the association of the *AGO1* rs636832 with lung cancer in the present study. Thus, the rs11263830 in the *AGO1* promoter that is in strong LD with the rs636832 ($|D'| = 1.0$ and $r^2 = 0.95$) may be the causal variant. Alternatively, it is possible that rs11263839 and rs617613 in the *AGO3* promoter in strong LD ($|D'| = 1.00$ and $r^2 = 0.90$, both) may be potential sources for the observed effect. However, even though we used a comprehensive SNPs set that captured all known common SNPs in HapMap, the HapMap data do not contain a complete catalogue of all genetic variations and are derived from a small number of subjects, and thus fine mapping will be required to identify the true causal variant.

To date, there were only two published studies on miRNA biosynthesis gene polymorphisms and cancer susceptibility, which were performed by the same study group at the University of Texas M.D. Anderson Center based on US Caucasians [23,33]. Yang et al. [23] reported that the *GEMIN3* rs197414C > A, and one (CCTGCTT) among the *GEMIN4* haplotypes (rs910924C > T, rs2740348G > C, rs7813C > T, rs910925C > G, rs3744741C > T, rs1062923T > C, and rs4968104T > A) were associated with a significantly increased risk of bladder cancer. In contrast to this study [23], Horikawa et al. [33] subsequently

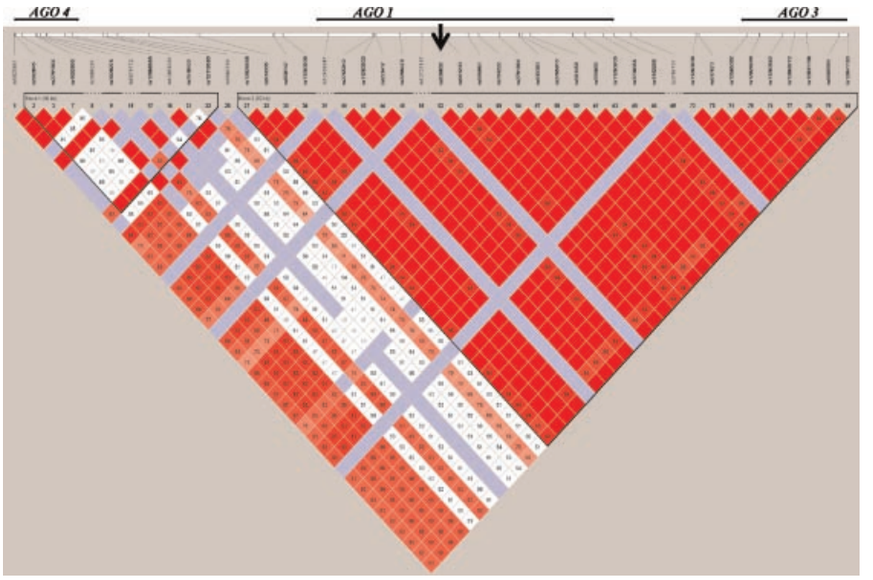

Figure 1. Reconstructed linkage disequilibrium (LD) plot using SNPs with minor allele frequency ≥1% from HapMap JPT data in a 100-kb region encompassing *AGO1*. Haplotype blocks are constructed by the confidence interval method proposed by Gabriel et al. [33]. The black boxes indicate strong LD (confidence interval for strong LD: upper 0.98, low 0.7; fraction of strong LD in informative comparisons must be at least 0.95). The white boxes indicate strong recombination (upper confidence interval maximum 0.9) and the gray boxes indicate noninformative findings. The triangles indicate haplotype blocks. The numbers in the squares are $|D'|$ (100×) values. Vertical bold arrow indicates the *AGO1* rs636832 location. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

reported that the same *GEMIN4* haplotype, the CCTNCTT (N indicates rs910925C > G [not examined]), was associated with a significantly decreased risk of renal cell carcinoma. However, in the present study, the *GEMIN3* and *GEMIN4* SNPs evaluated, as well as their haplotypes, were not significantly associated with the risk of lung cancer. Although it is difficult to decipher the basis for the discrepant finding from different studies, including an oppositely directed (flip-flop) association of the *GEMIN4* haplotypes in the previous studies [23,33], the different genetic backgrounds in the study populations or interactive effects of multiple polymorphisms might contribute to the discrepancy [34]. Another possible explanation is that the discrepancy between the different studies may be that different cell types of cancer have different etiologies and different carcinogenesis pathways. There is growing evidence that miRNA signatures are different according to human cancers [9]. Therefore, polymorphisms in miRNA biosynthesis genes might play different roles in different cell types of cancer. However, inadequacies in the study design, such as nonrandom sampling, limited sample sizes and the pitfalls arising from unknown confounders, also need to be considered.

One must consider a number of limitations of this study. This study used random sample selection for cases and controls so that the two groups were not matched for smoking status and the level of exposure to smoking, which can confound the association between the *AGO1* rs636832A > G genotype and lung cancer risk. To determine if smoking is a confounding variable, we assessed the potential interaction between genotypes and smoking by a logistic regression model, including the interaction term between genotype and smoking, and stratification analysis. The interaction term between genotype and smoking was not statistically significant ($P = 0.29$). In addition, there was no clear evidence that smoking modified the effect of the *AGO1* genotype on the risk of lung cancer in the stratified analyses (*P*-value of Wald test for heterogeneity of adjusted ORs = 0.85 and 0.38 for smoking status and level of exposure to smoking, respectively, Table 4). Moreover, the *P*-value of the Breslow–Day test for homogeneity of the ORs between the never- and ever-smoker groups was 0.70, and the Mantel–Haenszel estimate of the OR controlled for the confounding effect of smoking status was significant (0.69 [95% CI = 0.55–0.86], Table 3). Therefore, a confounding bias is unlikely. Our study was limited by the small sample size in stage 1 of the study, which did not have sufficient statistical power to detect all of the genetic factors. However, it would be possible to identify variants having a relatively large effect on the risk of lung cancer. It should be considered that the sample size of the stage 1 study might have a type II error for detection of genetic variants having small effects to disease susceptibility. Therefore, additional studies with larger sample sizes will be required to confirm our findings.

In conclusion, we found a significant association between the *AGO1* rs636832 and lung cancer risk. This association was detected in two independent sets of study populations. This finding suggests that the *AGO1* rs636832 might be a useful marker for determining the susceptibility to lung cancer. Future studies on the other sequence variants in the *AGO1* locus and their biologic function are needed to understand the role of the *AGO1* gene in determining lung cancer risk. Because the genetic polymorphisms often vary between different ethnic groups, further studies are needed to clarify the association of the *AGO1* rs636832 with lung cancer risk in diverse ethnic populations.

ACKNOWLEDGMENTS

This study was supported by National Research Foundation of Korea Grant funded by the Korean Government (grant no. 2009-0078985).

REFERENCES

1. Ambros V. The functions of animal MicroRNAs. Nature 2004;431:350–355.
2. Bartel DP. MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell 2004;116:281–297.
3. Lim LP, Lau NC, Garrett-Engele P, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 2005;433:769–773.
4. Bentwich I, Avniel A, Karov Y, et al. Identification of hundreds of conserved and nonconserved human microRNAs. Nature 2005;37:766–770.
5. Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E. Phylogenetic shadowing and computational identification of human microRNA genes. Cell 2005;120: 21–24.
6. Xu P, Guo M, Hay BA. MicroRNAs and the regulation of cell death. Trends Genet 2004;20:617–624.
7. Karp X, Ambros V. Developmental biology, encountering microRNAs in cell fate signaling. Science 2005;310:1330–1333.
8. Esquela-Kerscher A, Slack FJ. Oncomirs—microRNAs with role in cancer. Nat Rev Cancer 2006;6:259–269.
9. Calin GA, Croce CM. MicroRNA signature in human cancer. Nat Rev Cancer 2006;6:857–866.
10. Lee Y, Kim M, Han J, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J 2004;23:4051–4060.
11. Lee Y, Ahn C, Han J, et al. The nuclear RNAse III Drosha initiates microRNA processing. Nature 2003;425:4051–4060.
12. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. Processing of primary microRNAs by the microprocessor complex. Nature 2004;432:231–235.
13. Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-miRNAs and short hairpin RNAs. Genes Dev 2003;17:3011–3016.
14. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 2001;409:363–366.
15. Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. Human RISC couples microRNA biogenesis and posttranslational gene silencing. Cell 2005;123:631–640.

16. Yue J, Tigyi G. MicroRNA trafficking and human cancer. Cancer Biol Ther 2006;5:572–578.
17. Karube Y, Tanaka H, Osada H, et al. Reduced expression of Dicer associated with poor prognosis in lung cancer patients. Cancer Sci 2005;96:111–115.
18. Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet 2007;39:673–677.
19. Carmell MA, Xuan Z, Zhang MQ, Hannon GJ. The Argonaute family: Tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev 2002;16:2733–2742.
20. Benkirane M, Neuveut C, Chun RF, et al. Oncogenic potential of TAR RNA binding protein TRBP and its regulatory interaction with RNA-dependent protein kinase PKR. EMBO J 1977;16:611–624.
21. Spitz MR, Wei Q, Dong Q, Amos CI, Wu X. Genetic susceptibility to lung cancer: The role of DNA damage and repair. Cancer Epidemiol Biomark Prev 2003;12:689–698.
22. Lee SJ, Jeon HS, Jang JS, et al. DNMT3B polymorphisms and risk of primary lung cancer. Carcinogenesis 2005;26:403–409.
23. Yang H, Dinney CP, Ye Y, Zhu Y, Grossman HB, Wu X. Evaluation of genetic variants in MicroRNA-related genes and risk of bladder cancer. Cancer Res 2008;68:2530–2537.
24. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002; 296:2225–2229.
25. Stephens M, Smith MJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978–989.
26. Pharoah PDP, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer 2004;4:850–860.
27. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype–phenotype associations. Nature 2007;447:655–660.
28. Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, Hammond SM. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. Genes Dev 2006;20:2202–2207.
29. Peters L, Meister G. Argonaute proteins: Mediators of RNA silencing. Mol Cell 2007;26:611–623.
30. Kataoka Y, Takeichi M, Uemura T. Developmental roles and molecular characterization of Drosophila homologue of Arabidopsis Argonaute 1, the founder of a novel gene superfamily. Genes Cells 2001;6:313–325.
31. Sanchez-Cespedes M, Ahrendt SA, Piantoadosi S, et al. Chromosomal alterations in lung adenocarcinoma from smokers and nonsmokers. Cancer Res 2001;61:1309–1313.
32. Garnis C, Campbell J, Davis JJ, Macaulay C, Lam S, Lam WL. Involvement of multiple developmental genes on chromosome 1 in lung tumorigenesis. Hum Mol Genet 2005;14: 475–482.
33. Horikawa Y, Wood CG, Yang H, et al. Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma. Clin Cancer Res 2008;14:7956–7964.
34. Lin PI, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: The flip-flop phenomenon. Am J Hum Genet 2007;80:531–538.