

Polymorphisms in MicroRNAs Are Associated with Survival in Non–Small Cell Lung Cancer

Yang Zhao¹, Qingyi Wei², Lingming Hu¹, Feng Chen¹, Zhibin Hu¹, Rebecca S. Heist³, Li Su⁴, Christopher I. Amos⁵, Hongbing Shen¹, and David C. Christiani^{1,3,4}

Abstract

Background: MicroRNAs (miRNA) play important roles in the regulation of eukaryotic gene expression and are involved in human carcinogenesis. Single-nucleotide polymorphisms (SNP) in miRNA sequence may alter miRNA functions in gene regulation, which, in turn, may affect cancer risk and disease progression.

Methods: We conducted an analysis of associations of 142 miRNA SNPs with non–small cell lung cancer (NSCLC) survival using data from a genome-wide association study (GWAS) in a Caucasian population from the Massachusetts General Hospital (Boston, MA) including 452 early-stage and 526 late-stage NSCLC cases. Replication analyses were further performed in two external populations, one Caucasian cohort from The University of Texas MD Anderson Cancer Center (Houston, TX) and one Han Chinese cohort from Nanjing, China.

Results: We identified seven significant SNPs in the discovery set. Results from the independent Caucasian cohort demonstrated that the C allele of rs2042253 (hsa-miRNA-5197) was significantly associated with decreased risk for death among the patients with late-stage NSCLC (discovery set: HR, 0.80; $P = 0.007$; validation set: HR, 0.86; $P = 0.035$; combined analysis: HR, 0.87; $P = 0.007$).

Conclusions: These findings provide evidence that some miRNA SNPs are associated with NSCLC survival and can be used as predictive biomarkers.

Impact: This study provided an estimate of outcome probability for survival experience of patients with NSCLC, which demonstrates that genetic factors, as well as classic nongenetic factors, may be used to predict individual outcome. *Cancer Epidemiol Biomarkers Prev*; 23(11); 2503–11. ©2014 AACR.

Introduction

Lung cancer is the leading cause of cancer-related deaths in the United States (1). More than 80% of lung cancers are non–small cell lung cancer (NSCLC; ref. 2). It is reported that the 5-year survival rate of patients with NSCLC ranges from 11% to 17% (3). Both environmental and genetic factors contribute to the mortality of patients with NSCLC. Although it is well known that the major risk factor for lung cancer–related deaths is smoking, several genes, including loci at 3p22.1, 5p14.1, 6q16.1, 7q31.31, 9p21.3, 11p15.1, and 14q24.3, have also been identified to have an influence on overall survival or prognosis

through genetic association studies, as well as genome-wide association studies (GWAS; refs. 4, 5).

MicroRNAs (miRNA) are small, single-stranded, non-protein-coding RNAs that regulate gene expression. Although their biologic functions remain largely unknown, aberrant expression of miRNA seems to affect expression of various protein-coding oncogenes and tumor suppressors that are related to the etiology, diagnosis, and prognosis of many cancers (6–8). Recently, a growing body of evidence suggests that altered expression of some miRNAs can affect tumorigenesis and progression of NSCLC (9–20).

By altering the expression or maturation of a miRNA, single-nucleotide polymorphisms (SNP) within these 22-nucleotide sequences may lead to a dysregulation of gene expression that thereby contributes to cancer risk and survival (21–27). For example, Pu and colleagues reported that three miRNA SNPs (i.e., rs713065, rs6886834, and rs2234978) were associated with clinical outcomes in patients with early-stage NSCLC (28). The present study tested the hypothesis that SNPs in miRNAs contribute to survival of NSCLC. An association analysis was performed using genotype data from a GWAS in a Caucasian population with 452 patients with early-stage and 526 patients with late-stage NSCLC. Replication analyses were conducted in a Caucasian NSCLC cohort from The

¹Nanjing Medical University, Nanjing, China. ²Duke Cancer Institute, Duke University, Durham, North Carolina. ³Massachusetts General Hospital, Boston, Massachusetts. ⁴Harvard School of Public Health, Boston, Massachusetts. ⁵Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: David C. Christiani, Harvard School of Public Health, Room 1407, Building 1, 665 Huntington Avenue, Boston, MA 02115. Phone: 617-432-3323; Fax: 617-432-3441; E-mail: dchris@hsph.harvard.edu

doi: 10.1158/1055-9965.EPI-14-0389

©2014 American Association for Cancer Research.

University of Texas MD Anderson Cancer Center and a Han Chinese NSCLC cohort from Nanjing, China.

Materials and Methods

Study populations

The dataset from the discovery phase (The Harvard Lung Cancer Susceptibility Study GWAS) included 452 early-stage (I and II) and 526 late-stage (III and IV) patients with NSCLC recruited from Massachusetts General Hospital (Boston, MA). Details of participant recruitment for the study have been described previously (29). DNA was extracted from the whole blood and genotyped using the Illumina 610K Quad chip (Illumina).

Positive hits from the discovery set were validated in patients with NSCLC from two replication patient cohorts. One of the two cohorts comprised Caucasians in a lung cancer-control study at The University of Texas MD Anderson Cancer Center (Houston, TX). Sampling, genotyping, and quality control procedures have been described previously (5, 30, 31). The dataset included 788 newly diagnosed and histopathologically confirmed lung cancer cases. The second replication cohort was from a Han Chinese population in Nanjing, China (4, 32). This study included 609 patients with histopathologically or cytologically confirmed NSCLC.

A written informed consent was obtained from each subject at the time of recruitment, and the study was approved by the Institutional Review Boards of each participating institution.

Quality control in GWAS

We conducted systematic quality control (QC) on the raw genotyping data to filter both unqualified samples and SNPs (33). SNPs were excluded if they met any one of the following conditions: (i) SNPs that did not map on autosomal chromosomes; (ii) SNPs that had a call rate <95%; or (iii) SNPs that had minor allele frequency (MAF) <0.05. Samples with low call rates (<95%), ambiguous sex, familial relationships ($PI_HAT > 0.25$), outliers in the principal component analysis, and extreme heterozygote rate (>6 SD from nearest neighbor) were removed. Finally, a total of 543,697 SNPs passed the general QC procedure.

Extracting of miRNA SNPs

A list of miRNAs was downloaded from an online miRNA database (miRBase: <http://www.mirbase.org>, release 18; refs. 34, 35). We used *liftover* (<http://genome.ucsc.edu/util.html>), an online tool to lift the version of assembly from Hg18 to Hg19. However, because of the short length of miRNA regions, only 16 SNPs in the Illumina 610K Quad chip could be matched to miRNAs.

To increase the opportunity to capture miRNA SNPs, we first performed a genotype imputation procedure. The reference CEU (Utah residents with ancestry from northern and western Europe) panel was downloaded from the 1000 Genomes Project (phase I, release 2010-6; <http://www.1000genomes.org>). MACH (<http://www.sph.umich.edu/csg/abecasis/MACH/>) was used to impute the ungenotyped SNPs (36). Among the 4,649,540

SNPs that passed QC, there were 142 miRNA SNPs. Similarly, the imputation procedure for the MD Anderson cohort was also performed by using MACH and the CEU reference panel from the 1000 Genomes Project. For the Nanjing cohort, the reference panel was based on HapMap phase II database (CHB+JPT, released July 17, 2006).

Statistical analysis

We performed a two-stage association analysis. In the first stage, survival analyses were performed on the basis of the discovery GWAS dataset. To reach satisfactory power, we used a significant level of 0.01. We also used the false discovery rate (FDR) to evaluate the proportion of false positives among our findings (37). The survival time was defined as the length of period (unit, month) from the time of diagnosis until death or the latest follow-up. Cox proportional hazards model analysis was performed on both early- and late-stage patients. For the 452 early-stage patients, covariates adjusted included age, sex, smoking status, cell type (adenocarcinoma, squamous, and the others), stage (I vs. II), and the top four principal components (PC). For the 526 late-stage patients, age, sex, smoking status, cell type (adenocarcinoma, squamous, and others), stage (III vs. IV), surgery (yes vs. no), and the top four PCs were adjusted in the multivariate Cox model. To remove the possible adverse influence of some long-term survivors and allow for easy comparisons with other similar studies, those late-stage patients with more than 5 years of overall survival were right-censored. The PCs included in the model were generated by the EIGENSTRAT analysis, which were used to control for the confounding effect of population stratification (38, 39). The significant SNPs observed in stage I were then evaluated in two independent cohorts in stage II with a significance level of 0.05. Meta-analysis was performed to synthesize the results from different study cohorts. To evaluate the association of the validated miRNA SNPs on NSCLC survival, we performed a time-dependent receiver operating characteristic (ROC) analysis to calculate the cumulative area-under-the-curve (AUC) of the miRNA SNPs, as proposed by Chambless and Diao (40).

We used PLINK 1.07 for GWAS data management and general statistical analysis (41). The "survival" package in R (PLINK plug-in; <http://www.r-project.org/>) was used to conduct the survival analysis. Meta-analysis was performed by using "metan" package in Stata (version 12). The time-dependent ROC analysis was performed using the "survAUC" package in R. The target mRNAs of the miRNAs, including miR-#-5p and miR-#-3p, were predicted by using TargetScan (42), and the predicted mRNAs were further queried for GO functional enrichments using CapitalBio Molecule Annotation System V4.0 (MAS, <http://bioinfo.capitalbio.com/mas3/>).

Results

Characteristics of the NSCLC cases in the discovery set and the two replication patient cohorts are described in Table 1. In the discovery set, mean ages of patients

Table 1. Basic characteristics of the study populations from the three cohorts

	Harvard cohort (discovery)		MD Anderson cohort (replication)		Nanjing cohort (replication)
	Early stage	Late stage	Early stage	Late stage	Late stage
Total	452	526	241	547	609
Age					
Mean \pm SD	67.86 \pm 9.66	63.47 \pm 10.87	65.43 \pm 9.85	60.23 \pm 10.51	60.19 \pm 10.38
Sex					
Male	214 (47.35%)	259 (49.24%)	122 (50.62)	331 (60.51)	439 (72.09)
Female	238 (52.65%)	267 (50.76%)	119 (49.38)	216 (39.49)	170 (27.91)
Smoking status					
Never	40 (8.85%)	52 (9.89%)	0	0	232 (38.10)
Former	238 (52.65%)	260 (49.43%)	141 (58.51)	271 (49.54)	73 (11.99)
Current	174 (38.50%)	214 (40.68%)	100 (41.49)	276 (50.46)	304 (49.91)
Histology					
Adeno	199 (44.03%)	285 (54.18%)	119 (49.38)	275 (50.27)	395 (64.86)
SQC	135 (29.87%)	79 (15.02%)	74 (30.71)	147 (26.87)	214 (35.14)
Others	118 (26.11%)	162 (30.80%)	48 (19.92)	125 (22.85)	0
Stage					
I	381 (84.29%)	—	176 (73.03%)	—	—
II	71 (15.71%)	—	65 (27.97%)	—	—
III	—	238 (45.25%)	—	302 (55.21%)	376 (61.74)
IV	—	288 (54.75%)	—	245 (44.79%)	233 (38.26)

Abbreviations: adeno, adenocarcinoma; SQC, squamous carcinoma.

with early- and late-stage NSCLC were 67.86 and 63.47 years, respectively. The proportions of males with early and late-stage NSCLC were 47.35% and 49.24%, respectively. There were 174 (52.65%) current and 238 (38.50%) former smokers in the early-stage group. For the late-stage patients, the proportions of current and former smokers were 49.43% and 40.68%, respectively. For the histology type, more than half of the patients in both early and later stages were squamous (early stage, 29.87%; late stage, 15.02%) or adenocarcinoma (early stage, 44.03%; late stage, 54.48%).

We identified seven miRNA SNPs in the Harvard cohort; all were imputed. In the early-stage survival analysis, we identified three miRNA SNPs with P values < 0.01 , assuming an additive genetic model. The G>A variation of rs11048315 (hsa-mir-4302) was associated with increased survival time [HR, 0.64; 95% confidence interval (CI), 0.46–0.89; $P = 0.008$; FDR $q = 0.298$], whereas the variations of chr:129197463 (hsa-mir-182) and rs7522956 (hsa-mir-4742) were associated with increased risk of death (HR, 1.99; 95% CI, 1.35–2.93; $P = 0.0005$; FDR $q = 0.054$ and HR, 1.39; 95% CI, 1.12–1.74; $P = 0.003$; FDR $q = 0.195$, respectively). In the late-stage survival analysis, we found that the variations of rs17111728 (hsa-mir-4422), rs2042253 (hsa-mir-5197), and rs550894 (hsa-mir-612) were associated with a better survival outcome (HR, 0.60; 95% CI, 0.45–0.80; $P = 0.0005$; FDR $q = 0.053$; HR, 0.79; 95% CI, 0.67–0.94; $P = 0.007$; FDR $q = 0.193$; and HR, 0.70; 95% CI, 0.55–0.89; $P = 0.003$; FDR $q = 0.115$, respectively). In addition, rs7227168 (hsa-mir-4741) was associ-

ated with increased risk of death (HR, 1.35; 95% CI, 1.11–1.65; $P = 0.003$; FDR $q = 0.115$). Detailed results of the survival analysis are shown in Tables 2 and 3, as well as Supplementary Table S1 and Supplementary Figs. S1 and S2.

To validate these findings, we analyzed miRNA SNPs with $P < 0.01$ from the discovery set in the two replication populations. Of the seven SNPs, rs7522956 and rs2042253 were found in the MD Anderson cohort's imputed dataset. Only rs2042253, was significant at the level of 0.05 with the effect in the same direction as the discovery set (HR, 0.86; 95% CI, 0.74–0.99; $P = 0.035$). And only rs2042253 was identified in the imputed dataset from the Nanjing cohort but was not significantly associated with survival (Supplementary Table S2). We then performed a meta-analysis on rs7522956 and rs2042253, which existed in the datasets of at least two of the three cohorts. The results are presented in Table 4. None of the effects of the two SNPs were significantly heterogeneous among the cohorts (rs7522956, $P = 0.106$; rs2042253, $P = 0.219$). Thus, we used fixed effects model for data synthesis. Rs2042253 was significantly associated with the survival of patients with late-stage NSCLC (HR, 0.87; 95% CI, 0.80–0.96; $P = 0.003$). For rs7522956, there was a significant association between the genotype of rs7522956 and the survival of patients with early-stage NSCLC (HR, 1.22; 95% CI, 1.05–1.42; $P = 0.011$).

Several additional analyses were also performed. For the patients with early-stage NSCLC, rs7522956 was significantly associated with the progression-free survival

Table 2. Significant miRNA SNPs ($P < 0.01$) in the early-stage survival analysis (discovery set)

SNP	Chr	Position (bp, Hg19)	Genotype	Death N (%)	Censored N (%)	Median survival time (mo)	Comparison	Unadjusted model			Adjusted model			MIRNA
								HR (95% CI)	P	FDR q	HR (95% CI)	P	FDR q	
chr7:129197463	7	129410227	CC	183 (48.67)	193 (51.33)	91.4		1.00 (reference)		1.00 (reference)				hsa-mir-182
			TC	30 (69.77)	13 (30.23)	64.9	1 vs. 0	1.63 (1.1–2.4)	1.41E–02	1.87 (1.25–2.78)	2.19E–03			
			TT	1 (100)	0 (0)	9	2 vs. 0	31.75 (4.17–241.9)	8.46E–04	29.29 (3.53–242.65)	1.75E–03			
rs11048315	12	26026988	GG	189 (55.26)	153 (44.74)	78		1.00 (reference)		1.00 (reference)				hsa-mir-4302
			AG	41 (40.59)	60 (59.41)	111	1 vs. 0	0.69 (0.49–0.96)	2.94E–02	0.59 (0.41–0.84)	3.45E–03			
			AA	3 (37.5)	5 (62.5)	112	2 vs. 0	0.97 (0.31–3.05)	9.62E–01	0.88 (0.28–2.80)	8.33E–01			
rs7522956	1	224585958	AA	127 (49.42)	130 (50.58)	91.3		1.00 (reference)		1.00 (reference)				hsa-mir-4742
			AC	88 (53.66)	76 (46.34)	78	1 vs. 0	1.21 (0.92–1.59)	1.71E–01	1.30 (0.98–1.73)	6.73E–02			
			CC	18 (62.07)	11 (37.93)	78.8	2 vs. 0	1.82 (1.11–3.00)	1.81E–02	2.20 (1.29–3.77)	3.90E–03			
							Additive	1.28 (1.04–1.58)	1.89E–02	1.39 (1.12–1.74)	3.48E–03	1.95E–1		

(PFS; HR, 1.36; 95% CI, 1.10–1.68; $P = 0.005$), which demonstrates the potential prognostic value of rs7522956. To account for the potential confounding effects of other treatments, we also included the platinum chemotherapy (yes vs. no) and radiation treatment (yes vs. no) in the analysis. The results were similar to the original analysis (rs2042253 for late-stage patients: HR, 0.80; 95% CI, 0.67–0.94; $P = 0.0074$; rs7522956 for early-stage patients: HR, 1.35; 95% CI, 1.08–1.69; $P = 0.0079$).

We performed a time-dependent ROC analysis to evaluate the predictive utility of rs2042253 and rs7522956 for NSCLC survival outcome. For the late-stage patients, when the model included both rs2042253 and clinical risk score (derived from stage, cell type, and surgical operation), the cumulative AUC estimates at different time points were greater than those when the model included clinical risk score only (Fig. 1). Consistently, the summary measure of AUC for the combined model was greater than the one with clinical risk score only (0.63 vs. 0.57). For the early-stage patients, although rs7522956 was not replicated in the MD Anderson cohort, the model with rs7522956 and clinical score (derived from stage and cell type) provided greater summary measure of AUC than the model with clinical score only (0.57 vs. 0.46; Supplementary Fig. S3).

Discussion

There are an increasingly number of reports on the association between miRNA SNPs and survival of patients with lung cancer. Hu and colleagues reported that the C allele of rs11614913 (hsa-mir-196a2) was significantly associated with a decreased survival of patients with NSCLC (23). A meta-analysis published by Chen and colleagues demonstrated that hsa-mir-196a2 could also be a potential biomarker of lung cancer risk (43). Pu and colleagues reported that some miRNA-related SNPs (*FZD4*:rs713065, *DROSHA*:rs6886834, *FAS*:rs2234978) may be associated with NSCLC patients' clinical outcomes through altered miRNA regulation of the target genes (28). In a Han Chinese population, Cheng and colleagues suggested that the functional SNP rs2240688A>C in *CD133* could be a functional biomarker to predict risk and prognosis of lung cancer (44).

In the present study, we used genotype data generated from three GWAS datasets to examine the association of miRNA SNPs with the survival of patients with NSCLC. In the first stage, by using the Harvard Cohort, we found seven miRNA SNPs to be associated with survival of patients with early- or late-stage NSCLC. In the second stage, the positive hits were validated in one Caucasian and one Han Chinese cohort, resulting in one SNP, rs2042253, associated with improved survival for patients with late-stage NSCLC. Furthermore, in the time-dependent ROC analysis, we observed an improvement of 11% of the AUC when compared the combined risk model with the clinical score model only. This demonstrates the

Table 3. Significant miRNA SNPs ($P < 0.01$) in the late-stage survival analysis (discovery set)

SNP	Chr	Position (bp, Hg19)	Genotype	Death N (%)	Censored N (%)	Median survival time (mo)	Comparison	Unadjusted model			Adjusted model			MIRNA
								HR (95% CI)	P	FDR q	HR (95% CI)	P	FDR q	
rs17111728	1	55691384	TT	372 (83.22%)	75 (16.78%)	13.9	1 vs. 0	1.00 (reference)			1.00 (reference)			hsa-mir-4422
			TC	48 (76.19%)	15 (23.81%)	22.3		0.74 (0.55-1.01)	5.71E-02	0.63 (0.46-0.86)	3.87E-03			
			CC	2 (40%)	3 (60%)	—		0.36 (0.09-1.43)	1.45E-01	0.25 (0.06-1.01)	5.20E-02			
rs2042253	5	143059433	TT	223 (83.52%)	44 (16.48%)	13.5	1 vs. 0	1.00 (reference)			1.00 (reference)			hsa-mir-5197
			TC	159 (84.13%)	30 (15.87%)	16.1		0.88 (0.72-1.09)	2.47E-01	0.79 (0.64-0.97)	2.43E-02			
			CC	22 (73.33%)	8 (26.67%)	23.4		0.72 (0.46-1.11)	1.40E-01	0.65 (0.42-1.01)	5.44E-02			
rs550894	11	6521940	CC	339 (83.91%)	65 (16.09%)	14.8	1 vs. 0	1.00 (reference)			1.00 (reference)			hsa-mir-612
			AC	69 (74.19%)	24 (25.81%)	20.1		0.75 (0.58-0.98)	3.31E-02	0.74 (0.57-0.97)	2.88E-02			
			AA	3 (50%)	3 (50%)	25.2		0.40 (0.13-1.25)	1.16E-01	0.30 (0.09-0.95)	4.03E-02			
rs7227168	18	20513374	CC	320 (80.6%)	77 (19.4%)	16.4	1 vs. 0	1.00 (reference)			1.00 (reference)			hsa-mir-4741
			TC	89 (85.58%)	15 (14.42%)	10.1		1.38 (1.09-1.76)	8.49E-03	1.49 (1.17-1.91)	1.44E-03			
			TT	9 (100%)	0 (0%)	11.9		1.69 (0.87-3.28)	1.23E-01	1.31 (0.66-2.6)	4.43E-01			
							Additive	1.35 (1.11-1.65)	3.16E-03	1.35 (1.11-1.65)	3.11E-03	1.15E-1		

Table 4. Meta-analysis on rs7522956 and rs2042253 from the three cohorts

SNP	miRNA	Harvard cohort (discovery)			MD Anderson cohort (replication)			Nanjing cohort (replication)			Test of heterogeneity among studies			Fixed effect model		
		HR (95% CI)	P	P	HR (95% CI)	P	P	HR (95% CI)	P	P	I ²	P	HR (95% CI)	P		
rs7522956	hsa-mir-4742	1.39 (1.12–1.74)	3.48E–03	4.80E–1	1.08 (0.87–1.33)	4.80E–1	—	—	0.106	1.22 (1.05–1.42)	0.011	1.22 (1.05–1.42)	0.011			
rs2042253	hsa-mir-5197	0.79 (0.67–0.94)	6.95E–03	3.50E–2	0.86 (0.74–0.99)	3.50E–2	0.97 (0.82, 1.13)	6.66E–1	0.219	0.87 (0.80–0.96)	0.003	0.87 (0.80–0.96)	0.003			

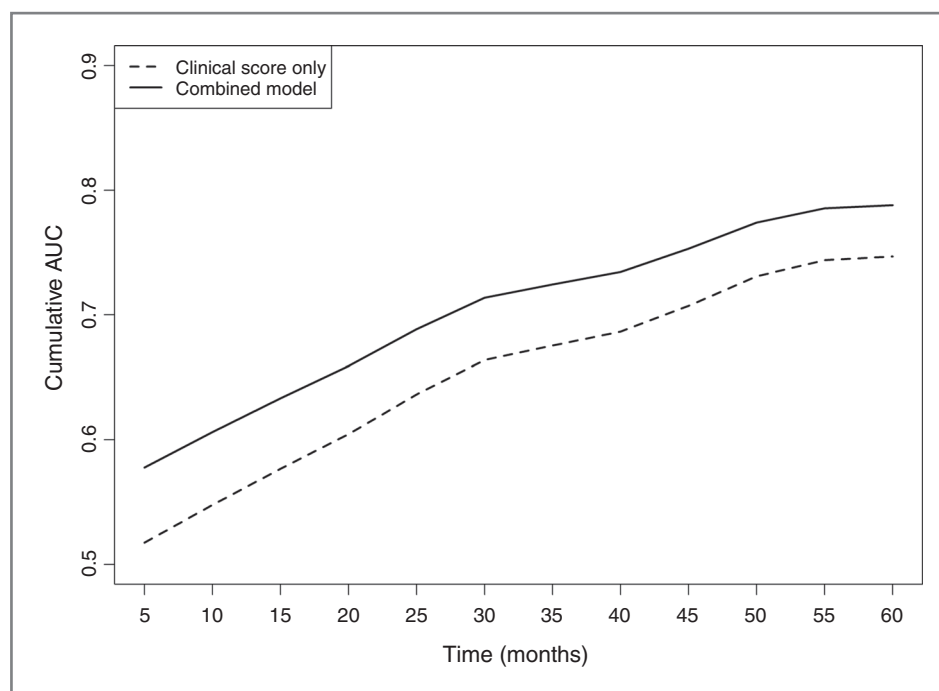
potential predictive value of hsa-mir-5197 (rs2042253) on the survival of patients with late-stage NSCLC. For the patients with early-stage NSCLC, although rs7522956 was not significant in the replication cohort, it can also improve the predictive ability of the model with clinical score only.

Hsa-mir-5197 (rs2042253), located on the long arm of chromosome 5, was significant in Caucasian populations from both Harvard and MD Anderson; the T>C variation of this SNP provided a protective effect on lung cancer survival. This novel miRNA is reported to have a high read frequency for pediatric acute lymphoblastic leukemia (ALL) through high-throughput sequencing (45). Although not significant in the Nanjing replication cohort, our meta-analysis showed that hsa-mir-4742 (rs7522956) was significantly associated with the survival of patients with early-stage NSCLCs, which was also associated with the PFS in early-stage patients in the Harvard cohort. Rs2042253 is located in the adjacent region of miR-5197-5p, also called a microRNA-offset RNA (moRNA; refs. 46–48), and rs7522956 is located in loop sequences of the mir-4742 gene (Supplementary Fig. S4_A). The nucleotide variations in pre-miRNAs may have an effect on the stability of the stem-loop structure, and even contribute to pre-miRNA processing via affecting recognition and cleavage of Drosha and Dicer. Although mature miRNA sequences can be generated by these pre-miRNAs with varied nucleotides, multiple isomiRs in miRNA processing and maturation processes may be regulated (49–53). These mature miRNAs, including miR-#5p and miR-#3p, have important biologic roles through contributing to basic multiple biologic processes (such as TGF-beta signaling pathway, cytokine–cytokine receptor interaction, and the insulin signaling pathway) and development of some human cancers (such as colorectal cancer, prostate cancer, and endometrial cancer; Supplementary Fig. S4_B). Further studies are needed to understand the roles of these two miRNA SNPs in the survival of patients with NSCLC.

The present study has several strengths. We used three relatively large datasets from three independent GWASs for the discovery and confirmation of the association between miRNA SNPs and overall survival. Thus, SNPs that were identified by the analysis should have a high probability of being true-positive findings. We also performed a time-dependent AUC to demonstrate that the identified miRNA SNP could be a biomarker for the survival of patients with NSCLC.

However, we acknowledge that there are several limitations of the present study. First, most microRNA-related variants may not be well covered by current GWAS platforms. Although we used imputed datasets for SNP extracting, it is possible that some miRNA SNPs identified in one cohort may not exist in the dataset of another cohort, because the SNPs may not in the reference panel or the imputed SNPs are of low quality. Second, although one miRNA SNP was validated in both Caucasian cohorts,

Figure 1. Time-dependent ROC analysis of rs2052253. The cumulative AUC estimates of the combined model (rs2042253 + clinical score) at different time points were greater than those when the model included clinical risk score only.



no SNPs were positive in the Nanjing Cohort. Possible reasons for this discrepancy include different ethnic background, demographic, and clinic characteristics of the cohorts, as well as potential gene–gene interactions between functional SNPs and gene–environmental interactions, which may result in the failure to replicate for some statistically significant SNPs in an independent dataset (54). Third, *in vitro* functional assays are needed for evidence of biologic plausibility of the identified miRNA SNPs.

In conclusion, the present study provides evidence that SNPs in some miRNAs are associated with NSCLC survival. Our result provides evidence for the application of miRNA SNPs as predictive biomarkers in future personalized medicine for patients with NSCLC. Further investigation is needed to illustrate the precise mechanism by which the miRNA SNPs affect NSCLC survival.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: Y. Zhao, Q. Wei, F. Chen, H. Shen, D.C. Christiani
Development of methodology: L. Su, C.I. Amos, D.C. Christiani
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): Q. Wei, L. Hu, R.S. Heist, L. Su, C.I. Amos, D.C. Christiani

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y. Zhao, R.S. Heist, C.I. Amos, D.C. Christiani

Writing, review, and/or revision of the manuscript: Y. Zhao, Q. Wei, R.S. Heist, C.I. Amos, H. Shen, D.C. Christiani

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Z. Hu, L. Su, D.C. Christiani

Study supervision: Y. Zhao, Q. Wei, F. Chen, Z. Hu, H. Shen, D.C. Christiani

Acknowledgments

The authors thank Dr. Margaret Spitz for her suggestions on preparing this article. The authors also thank Dr. Li Guo and Dr. Guangfu Jin for their support of this work. The authors thank the participants and the physicians and staff of the Massachusetts General Hospital, the M.D. Anderson Cancer Center, and the Nanjing Medical University affiliated hospitals. The authors also greatly appreciate the constructive comments from the two anonymous reviewers.

Grant Support

This study was supported by NIH grants (R01CA092824, P50CA090578, and P30ES000002 to D.C. Christiani), the Natural Scientific Funding of China [NSFC; 30901232, 81373102, and NIH grants 5R01CA092824 (PI, D. C. Christiani) to Y. Zhao; NSFC 81072389 to F. Chen], Scientific Research Grants for High Education of Jiangsu Province (12KJB310003 to Y. Zhao), and The National Science Foundation for Distinguished Young Scholars of China (81225020 to Z. Hu).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 15, 2014; revised July 17, 2014; accepted July 22, 2014; published OnlineFirst August 7, 2014.

References

- Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012;62:10–29.
- Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* 2008;83:584–94.
- Minna J. Neoplasms of the lung. In: Fauci A, Braunwald E, Kasper D, et al. (editors). *Harrison's principles of internal medicine*. New York: McGraw-Hill; 2005. p. 506–15.
- Hu L, Wu C, Zhao X, Heist R, Su L, Zhao Y, et al. Genome-wide association study of prognosis in advanced non-small cell lung cancer

- patients receiving platinum-based chemotherapy. *Clin Cancer Res* 2012;18:5507–14.
5. Wu X, Ye Y, Rosell R, Amos CI, Stewart DJ, Hildebrandt MA, et al. Genome-wide association study of survival in non-small cell lung cancer patients receiving platinum-based chemotherapy. *J Natl Cancer Inst* 2011;103:817–25.
 6. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 2006;6:259–69.
 7. Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet* 2007;39:673–7.
 8. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature* 2005;435:834–8.
 9. Chen Y, Min L, Zhang X, Hu S, Wang B, Liu W, et al. Decreased miRNA-148a is associated with lymph node metastasis and poor clinical outcomes and functions as a suppressor of tumor metastasis in non-small cell lung cancer. *Oncol Rep* 2013;30:1832–40.
 10. Wang XC, Wang ZH, Zhang ZB, Zhao J, Tan XG, Luo JC. Overexpression of miRNA-21 promotes radiation-resistance of non-small cell lung cancer. *Radiat Oncol* 2013;8:146.
 11. Kasinski AL, Slack FJ. miRNA-34 prevents cancer initiation and progression in a therapeutically resistant K-ras and p53-induced mouse model of lung adenocarcinoma. *Cancer Res* 2012;72:5576–87.
 12. Salim H, Akbar NS, Zong D, Vaculova AH, Lewensohn R, Moshfeh A, et al. miRNA-214 modulates radiotherapy response of non-small cell lung cancer cells through regulation of p38MAPK, apoptosis and senescence. *Br J Cancer* 2012;107:1361–73.
 13. Gao W, Lu X, Liu L, Xu J, Feng D, Shu Y. MiRNA-21: a biomarker predictive for platinum-based adjuvant chemotherapy response in patients with non-small cell lung cancer. *Cancer Biol Ther* 2012;13:330–40.
 14. Gao W, Xu J, Shu YQ. miRNA expression and its clinical implications for the prevention and diagnosis of non-small-cell lung cancer. *Expert Rev Respir Med* 2011;5:699–709.
 15. Wang XC, Tian LL, Jiang XY, Wang YY, Li DG, She Y, et al. The expression and function of miRNA-451 in non-small cell lung cancer. *Cancer Lett* 2011;311:203–9.
 16. Wang ZX, Bian HB, Wang JR, Cheng ZX, Wang KM, De W. Prognostic significance of serum miRNA-21 expression in human non-small cell lung cancer. *J Surg Oncol* 2011;104:847–51.
 17. Keller A, Leidinger P, Gislefoss R, Haugen A, Langseth H, Staehler P, et al. Stable serum miRNA profiles as potential tool for non-invasive lung cancer diagnosis. *RNA Biol* 2011;8:506–16.
 18. Jeong HC, Kim EK, Lee JH, Lee JM, Yoo HN, Kim JK. Aberrant expression of let-7a miRNA in the blood of non-small cell lung cancer patients. *Mol Med Rep* 2011;4:383–7.
 19. Wang XC, Du LQ, Tian LL, Wu HL, Jiang XY, Zhang H, et al. Expression and function of miRNA in postoperative radiotherapy sensitive and resistant patients of non-small cell lung cancer. *Lung Cancer* 2011;72:92–9.
 20. Wang XC, Tian LL, Wu HL, Jiang XY, Du LQ, Zhang H, et al. Expression of miRNA-130a in nonsmall cell lung cancer. *Am J Med Sci* 2010;340:385–8.
 21. Duan R, Pak C, Jin P. Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. *Hum Mol Genet* 2007;16:1124–31.
 22. Horikawa Y, Wood CG, Yang H, Zhao H, Ye Y, Gu J, et al. Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma. *Clin Cancer Res* 2008;14:7956–62.
 23. Hu Z, Chen J, Tian T, Zhou X, Gu H, Xu L, et al. Genetic variants of miRNA sequences and non-small cell lung cancer survival. *J Clin Invest* 2008;118:2600–8.
 24. Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A. Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc Natl Acad Sci U S A* 2008;105:7269–74.
 25. Tian T, Shu Y, Chen J, Hu Z, Xu L, Jin G, et al. A functional genetic variant in microRNA-196a2 is associated with increased susceptibility of lung cancer in Chinese. *Cancer Epidemiol Biomarkers Prev* 2009;18:1183–7.
 26. 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;68:2530–7.
 27. Yu Z, Li Z, Jolicoeur N, Zhang L, Fortin Y, Wang E, et al. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic Acids Res* 2007;35:4535–41.
 28. Pu X, Roth JA, Hildebrandt MA, Ye Y, Wei H, Minna JD, et al. MicroRNA-related genetic variants associated with clinical outcomes in early-stage non-small cell lung cancer patients. *Cancer Res* 2013;73:1867–75.
 29. Asomaning K, Miller DP, Liu G, Wain JC, Lynch TJ, Su L, et al. Second hand smoke, age of exposure and lung cancer risk. *Lung Cancer* 2008;61:13–20.
 30. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008;40:616–22.
 31. Yu H, Zhao H, Wang LE, Han Y, Chen WV, Amos CI, et al. An analysis of single nucleotide polymorphisms of 125 DNA repair genes in the Texas genome-wide association study of lung cancer with a replication for the XRCC4 SNPs. *DNA Repair* 2011;10:398–407.
 32. Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat Genet* 2011;43:792–6.
 33. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc* 2010;5:1564–73.
 34. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res* 2008;36:D154–8.
 35. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011;39:D152–7.
 36. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009;10:387–406.
 37. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001;125:279–84.
 38. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–9.
 39. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* 2010;11:459–63.
 40. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med* 2006;25:3474–86.
 41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
 42. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell* 2003;115:787–98.
 43. Chen Z, Xu L, Ye X, Shen S, Li Z, Niu X, et al. Polymorphisms of microRNA sequences or binding sites and lung cancer: a meta-analysis and systematic review. *PLoS ONE* 2013;8:e61008.
 44. Cheng M, Yang L, Yang R, Yang X, Deng J, Yu B, et al. A microRNA-135a/b binding polymorphism in CD133 confers decreased risk and favorable prognosis of lung cancer in Chinese by reducing CD133 expression. *Carcinogenesis* 2013;34:2292–9.
 45. Schotte D, Akbari Moqadam F, Lange-Turenhout EA, Chen C, van Ijcken WF, Pieters R, et al. Discovery of new microRNAs by small RNAome deep sequencing in childhood acute lymphoblastic leukemia. *Leukemia* 2011;25:1389–99.
 46. Langenberger D, Bermudez-Santana C, Hertel J, Hoffmann S, Khaitovich P, Stadler PF. Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics* 2009;25:2298–301.

47. Shi WY, Hendrix D, Levine M, Haley B. A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. *Nat Struct Mol Biol* 2009;16:183–9.
48. Umbach JL, Cullen BR. In-depth analysis of kaposi's sarcoma-associated herpesvirus MicroRNA expression provides insights into the mammalian MicroRNA-processing machinery. *J Virol* 2010;84:695–703.
49. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 2007;129:1401–14.
50. Morin RD, Aksay G, Dolgosheina E, Ebhardt HA, Magrini V, Mardis ER, et al. Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. *Genome Res* 2008;18:571–84.
51. Guo L, Yang Q, Lu J, Li H, Ge Q, Gu W, et al. A comprehensive survey of miRNA repertoire and 3' addition events in the placentas of patients with pre-eclampsia from high-throughput sequencing. *PLoS ONE* 2011;6:e21072.
52. Neilsen CT, Goodall GJ, Bracken CP. IsomiRs - the overlooked repertoire in the dynamic microRNAome. *Trends Genet* 2012;28:544–9.
53. Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, et al. Complexity of the microRNA repertoire revealed by next generation sequencing. *RNA* 2010;16:2170–80.
54. Greene CS, Penrod NM, Williams SM, Moore JH. Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE* 2009;4:e5639.