The Lung Cancer Associated MicroRNAs and Single Nucleotides Polymorphisms: a Mendelian Randomization Analysis

Ruixuan HUANG, William C. CHO, Yanni SUN, and Kei Hang Katie CHAN

Abstract— **Lung cancer is a major public health burden and among the highest incidence and mortality rates of the cancers. MicroRNAs (miRNAs) play an important role in the development of lung cancer. The aim of this study was to investigate whether there was a potential causal relation between miRNAs and non-small-cell lung cancer (NSCLC). 1,026 patients with NSCLC from The Cancer Genome Atlas (TCGA) were analyzed. NSCLC associated SNPs' allele scores were established, and candidate miRNAs were filtered from differential expression analysis. Mendelian randomization (MR) analysis was conducted for 5 candidate miRNA (hsa-miR-135b, hsa-miR-142, hsa-miR-182, hsa-miR-183 and hsa-miR-3607) and 76 candidate SNPs in lung adenocarcinoma (LUAD) group. According to the core assumptions of MR, there was no clear evidence of a causal relation between the 5 candidate miRNAs and LUAD. The reads per million miRNAs mapped (RPM) level of candidate miRNAs changed less than 3% per allele score. To our knowledge, this is the first study using the TCGA data set to investigate the causal relation between miRNAs and lung cancer using the MR approach, and also one of the first MR studies to use miRNA expression as an exposure factor, with the SNPs as instrumental variables.**

I. Introduction

Lung cancer is the leading cause of cancer deaths in the world. Its etiology is multiple, including genetic and epigenetic damage, as well as tobacco smoking [1]. Non-small-cell lung cancer (NSCLC) accounts for more than 85% of lung cancer [2]. Lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most common types of NSCLC. LUAD is the most common type of lung cancer among the non-smokers, which accounts for about 40% of the total number of patients with lung cancer [3]. LUSC is not as common as LUAD, but its relation to smoking history is stronger than LUAD [4]. Moreover, studies have shown that LUSC is associated with gender and it is more common in men than in women [5]. Experimental and previous studies have shown that some microRNAs [6] and single nucleotide polymorphisms (SNPs) [7] are associated with lung cancer. However, not much previous study investigated whether there is a causal relation between microRNAs and lung cancer, particularly using Mendelian Randomization methodology with SNPs as an instrument.

SNPs has been classified as commonly occurring (>1%) genetic variation in the general population, whereas the rare variants with obvious functional consequences on the protein have been classified as mutations [8]. SNPs are widely found in the human genome, with an average of 1 out of every 500 or 1,000 base pairs, with an estimated total of three million or more in the genome [8]. According to some epidemiological studies, lung cancer had been shown to be a typical environment-related disease. Cigarette smoking, air pollution, and sulfuric acid mist are important risk factors for lung cancer [6, 37-38]. Studies also reported that only a small number of non-smokers (10% - 15%), who suffered from lung cancer, were exposed to the same carcinogenic factors including tar, soot, arsenic, chromium, silica dust and asbestos [9]. This indicated that the existence of genetic susceptibility based on individual differences played an important role in the occurrence of lung cancer [9]. Studies have shown that the differential survivability and mortality of NSCLC are related to genetic variation such as SNPs, which can be used as a potential prognostic indicator or predictor. For example, POLA2+1747 GG/GA (rs487989) has the potential to be used as a prognostic biomarker of patient outcome in NSCLC pathogenesis [10]; *BAG6* rs3117582 SNP is associated with lung cancer in Europeans and can be used as an independent predictor of lung cancer risk [11].

MicroRNA (miRNA) is a class of non-coding single-stranded RNA molecules with a length of about 22 nucleotides [12]. MiRNA acts as a negative regulator of gene expression by binding to 3'- UTRs of target mRNAs. MiRNAs involve in many crucial biological processes, including cell cycle, growth, death, stem cell differentiation, and stress response. In recent years, increasing evidence indicated that miRNAs mutations and dysregulations are associated with the occurrence and development of many complex human diseases such as Alzheimer's disease [13], colorectal carcinoma [14], leukemia, etc. [15]. Studies have shown that the over or under expression of miRNAs in lung cancer tissues plays important roles in the development and progression of lung cancer [16]. Eight microRNAs were found to have prognostic effect on NSCLC (hsa-miR-375, hsa-miR-148a, hsa-miR-29b-1 and hsa-miR-584 had a prognostic effect on the prognosis of LUAD, and hsa-miR-4746, hsa-miR-326, hsa-miR-93 and hsa-miR-671 had prognostic effect on LUSC) [16]. Other studies have shown that compared with the healthy control group, the serum level of miR-182, miR-183 is significantly higher in patients with non-small-cell lung cancer, which may act as a sensitive and specific biomarker for the early diagnosis of non-small-cell lung cancer [17, 39-40]. Although there have been many studies on lung cancer and microRNAs in recent years, the research is not thorough

* Research supported by City University of Hong Kong New Research Initiatives/Infrastructure Support from Central (APRC).

R. Huang. Author, is with City University of Hong Kong, Hong Kong SAR. (e-mail: rxhuang4-c@my.cityu.edu.hk).

W.C. Cho. Author, is with Department of Clinical Oncology, Queen Elizabeth Hospital, Kowloon, Hong Kong SAR.

Yanni SUN. Author, is with City University of Hong Kong, Hong Kong SAR. (e-mail: yannisun@cityu.edu.hk).

K. H. K. Chan. Author is with City University of Hong Kong, Hong Kong SAR; and Brown University, Providence RI 02912 USA (phone: 852-34426661; e-mail: katie.kh.chan@cityu.edu.hk).

enough, and the role of some microRNAs in lung cancer is still unclear. Some studies suggested that hsa-miR-9-1 and hsa-miR-9-2 are overexpressed in lung cancer [18-19]. However, some studies suggested that hsa-9 group miRNAs had similar effects as hsa-let-7g, which is associated with NFκB1, and significantly downregulated in NSCLC [20]. The transformation of the SNP that is associated with the risk of lung cancer has been demonstrated to be related to miRNAs. For example, the transformation (from rs2240688A to rs2240688C) is associated with hsa-miR-135a/b and is demonstrated to be associated with the risk of lung cancer [21], which may be a functional biomarker for predicting the risk and prognosis of lung cancer. At present, databases such as the Human microRNA Disease Database (HMDD) [22] and algorithms for dissecting human miRNA-disease associations have been developed, but their limitations are that most of them only show the association between miRNAs and disease (up- / down- regulation of miRNAs), without resolving whether the association is causal or not [23]. However, causal inference is essential in cancer research. It is necessary to understand whether the change in miRNAs expression is before cancer occurs (as a predictive indicator) or after cancer syndrome (as a biomarker). Therefore, causality analysis is very important to enhance the understanding of the role of miRNA in specific cancer mechanisms.

Mendelian randomization (MR) is an analytical method based on genetic variables (instrumental variables) to determine whether the correlation between observed risk factors and outcomes is causal or not [24]. Compared to the traditional experimental verification method, MR method is more concise, more intuitive and saves time. The selection of instrumental variables is critical to a successful MR analysis. Valid instrumental variables must be in accordance with three core assumptions: a) instrumental variables must be reproducible and strongly related to exposure; b) instrumental variables are not related to confounding factors; c) instrumental variables are related to the results only through exposure factors [25]. Mendelian randomization analysis consists of two main steps: i) to check three basic core assumptions; ii) to assess the causality between exposure and outcomes. In the HMDD database, there are 201 miRNAs related to LUAD, of which more than 150 miRNAs are not confirmed to be causal; 57 miRNAs that related to LUSC and only 50 miRNAs are not proved causal. Therefore, MR is a good analytical method to fill the gap of causal inference of miRNAs in the development of complex diseases such as lung cancer at this stage.

In this study, we used MR methodology to examine the causal relation between miRNAs and two sub-types of NSCLC (LUAD and LUSC), using SNPs as an instrument.

II. Methods

The overall analytical approach is illustrated in Fig. 1, representing data collection, data preprocessing, MR hypothesis validation, and MR analysis. The clinical data, miRNA expression data and part of Nucleotide Variation data of two sub-types of lung cancer, including LUAD and LUSC were collected from The Cancer Genome Atlas (TCGA, https://www.cancer.gov/tcga). The remaining part of Single Nucleotide Variation (SNV) data were collected from the NHGRI-EBI GWAS Catalog, which were used as reference data in this study (https://www.ebi.ac.uk/gwas/). All data were downloaded using the R package "TCGAbiolinks".

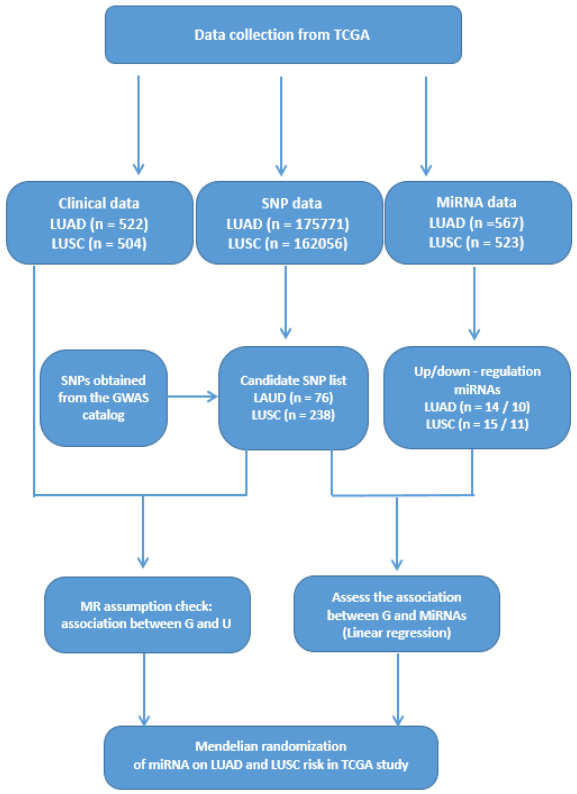

Figure 1. The overall analytical approach. TCGA: The Cancer Genome Atlas; SNP: Single Nucleotide Polymorphism; miRNA: microRNA; GWAS: genome-wide association study; G: Candidate SNP list; U: Age, Gender, Ethnicity, Cigarettes per day, Years of smoking.

A. Clinical Data Collection

Clinical files of LUAD (n = 522) and LUSC (n = 504) were collected from TCGA. The diagnoses, treatments, demographic, and exposures information (including height, weight, smoking history, alcohol history, alcohol intensity, and BMI.) were included. Basic information of patients (for example, submitter id, gender, age, race, and ethnicity) and lifestyle information related to lung cancer (for example, cigarettes per day, and years of smoking) were extracted and recorded. There were 37% and 49% missing values in the "cigarettes per day", and "years of smoking" variables, respectively. Multiple imputations were adopted to deal with the missing values. Briefly, this imputation method utilized the Monte Carlo algorithm to predict and fill the missing data according to different data types with different fitting and regression modes [26]. Morphology and tumor stage data were included in this missing value filling process. The network based on age, gender, race, ethnicity, morphology, tumor stage, cigarettes per day, and years of smoking is constructed, and the missing values in the network were simulated by generalized linear simulation method, and then a complete data set was generated. The distribution of clinical data is illustrated in Fig. 2, which shows the distribution of imputed data being similar to the original data, the applicability of Multiple Imputation method was demonstrated.

B. SNP Data Collection

Single Nucleotide Polymorphism (SNP) data was collected from SNV data, including Hugo Symbol, reference allele, tumor risk allele, dbSNP ID, reported gene, risk allele, risk allele frequency, P-value). The candidate SNP lists of LUAD and LUSC were generated by sharing the same gene, risk allele frequency ≥ 0.5, P value ≤ 0.05. The SNPs that were not recorded in the dbSNP were excluded. 76 out of 763 SNPs from LUAD and 238 out of 1816 SNPs from LUSC were selected in the candidate SNP lists. The tumor risk allele was coded as 1 while the other allele was coded as 0. An allele score, which was calculated as the total number of risk alleles from the SNPs in the candidate SNP list was used. The allele score was used as the instrumental variable in the subsequent analysis.

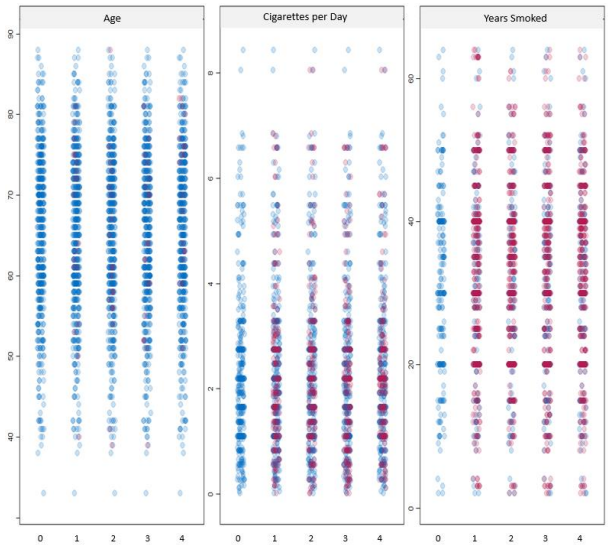

Figure 2. Distribution of clinical data. Group 0 represents the data before imputation, group 1-4 represent the different iterations of imputation. Blue dots represent the original data, red dots denote imputed data. The distribution of imputed data is similar to the original data.

C. MiRNA Data Collection

MiRNA expression data was obtained using the R package "TCGAbiolinks". There were 567 LUAD-associated miRNAs and 523 LUSC-related miRNAs. MiRNAs with the expression of 0 in more than 30% samples were excluded. The clinical data, SNP data and miRNA data were merged and 510 LUAD patients and 473 LUSC patients retained for subsequent statistical analyses.

D. Differential Expression Analysis

To estimate the relation between miRNAs and the two lung cancer sub-types, differential expression analysis (linear models for microarray data) was applied. Fold change and P-value of each miRNA were calculated based on the reads per million miRNAs mapped (RPM) value. MiRNAs with more than 30% of samples' RPM level are 0 were excluded, in order to avoid false high / low expression. 27 up-regulated miRNAs and 10 down-regulated miRNAs from LUAD samples, along with 50 up-regulated miRNAs and 15 down-regulated miRNAs from LUSC samples were screened out. The threshold was set as absolute log2 fold change ≥ 2 and P-value ≤ 0.05.

E. Mendelian Randomization

The assumption that the instrumental variables had no relation with potential confounding variables for the association between SNPs and two sub-types of lung cancer, was tested. Linear regression method was used to estimate the associations between the allele score and the potential confounding variables (i.e. age, cigarettes per day, and years of smoking). Logistic regression was employed in corresponding analyses of binary covariates (i.e. gender, ethnicity). Linear regression was applied to calculate the P-value and R^2 value between SNPs or the allele score and miRNA expression level (RPM).

In order to calculate the MR estimation of LUAD and LUSC risk by candidate miRNAs, the beta coefficient and standard error of the linear regression of allele score and RPM were extracted. Cox semiparametric hazards model was used to verify allele score correlation with LUAD and LUSC. Patients' alive/death status in clinical data were used as an outcome variable in the model, from the date of diagnosis, the date of last follow-up in the surviving patients, and the time of death in dead patients, patient's illness time were calculated, and were used as the time scale in the model. Inverse-variance weighted (IVW) method was used for the summarized data to calculate MR estimates of miRNAs for lung cancer. The MR analysis was carried out by using the package Mendelian Randomization (version 0.3.0) in R (version 3.5.2).

III. Results

In this study, SNP is the genetic variant, G; miRNA is exposure, P; LUAD and LUSC are the disease states, D; age, year of smoking and ethnicity are the confounding factors, U (Fig. 3).

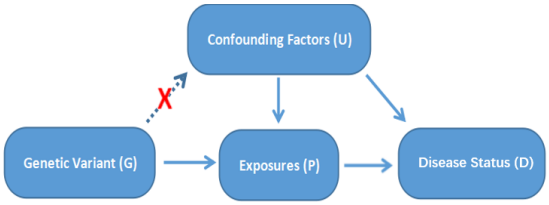

Figure 3. One stage Mendelian Randomization diagram. U: Age, Gender, Ethnicity, Cigarettes per day, Years of smoking; G: Candidate SNP list; P: Up/down- regulated miRNAs; D: LUAD and LUSC; Solid arrows: associations that conform to the MR hypothesis; Red cross: association that should not exist.

Table I shows the distribution of baseline characteristics of patients with the two sub-types of lung cancer (LUAD and LUSC). The number of patients, average number of age, and cigarettes per day were similar in the LUAD and LUSC groups. In terms of gender, the ratio of male to female in the LUAD group was similar, and the number of males (n = 350) in the LUSC group was more than twice than that of female (n = 123). For Ethnicity, the Hispanic or Latino population in both LUAD and LUSC groups accounted for less than 4% of the total, which is a feature with skewed distribution. For years of smoking, the average years of smoking in the LUSC group (40.71 years) was around 10 years longer than that in the LUAD group (30.79 years).

TABLE I. DISTRIBUTION OF BASELINE CHARACTERISTICS IN NSCLC

Subjects	Non-Small-Cell Lung Cancer	
	LUAD	LUSC
Number of patients	510	473
Age (Average Years)	65.25	67.39
Gender (Male / Female)	237 / 273	350 / 123
Ethnicity (Hispanic or Latino / not Hispanic or Latino)	20 / 490	18 / 455
Cigarettes Per Day (Average)	2.299	2.874
Years Smoked (Average)	31.49	40.71

The confounders check was performed by using 76 candidate SNPs from LUAD group, 238 candidate SNPs from LUSC group and clinical data. Linear regression was used for continuous features and logistic regression for categorical features. Table II and Table III illustrate the relation between SNPs (allele score) and confounders in LUAD and LUSC receptively. Although the distribution of each subject varies in LUAD group and LUSC group, the coefficient value of each confounder is close to 0, and P-value of each confounder is larger than 0.05, which shows that G, the SNPs are not associated with U, the confounding factors.

TABLE II. ASSOCIATIONS BETWEEN ALLELE SCORE AND POTENTIAL CONFOUNDERS IN LUAD (N = 510)

Subjects	Coefficient[a]	95%CI	P-value
Age	-0.002	(-0.004 to 0.001)	0.298
Gender (Male vs Female)	0.012	(-0.052 to 0.075)	0.722
Ethnicity (Hispanic / Latino vs not)	-0.117	(-0.281 to 0.046)	0.161
Cigarettes Per Day	0.001	(-0.021 to 0.022)	0.996
Years Smoked	0.001	(-0.001 to 0.004)	0.267

a. The coefficient was derived from linear regression for continuous variables and from logistic regression for categorical variables.

TABLE III. ASSOCIATIONS BETWEEN ALLELE SCORE AND POTENTIAL CONFOUNDERS IN LUSC (N = 473)

Subjects	Coefficient[a]	95%CI	P-value
Age	-0.004	(-0.014 to 0.006)	0.467
Gender (Male vs Female)	-0.042	(-0.237 to 0.153)	0.673
Ethnicity (Hispanic / Latino vs not)	-0.218	(-0.665 to 0.229)	0.339
Cigarettes Per Day	0.029	(-0.021 to 0.078)	0.254
Years Smoked	-0.001	(-0.008 to 0.006)	0.774

a. The coefficient was derived from linear regression for continuous variables and from logistic regression for categorical variables.

Fig. 4 presents the volcano plots of miRNAs, denoting the up- and down-regulated miRNAs in LUAD (left) and LUSC (right) groups. Black dots represent the non-regulated miRNAs, red dots represent the up-regulated miRNAs, while the green dots represent the down-regulated miRNAs. 27 up-regulated miRNAs and 10 down-regulated miRNAs in LUAD group, along with 50 up-regulated miRNAs and 15 down-regulated miRNAs in LUSC group were marked out.

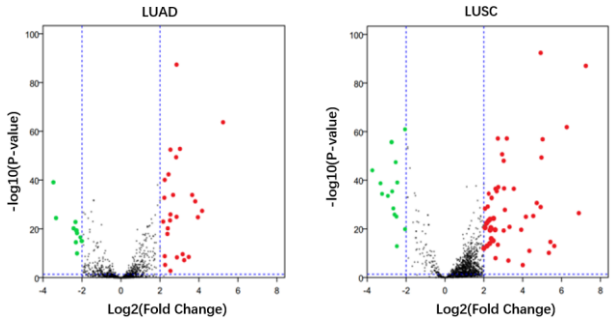

Figure 4. Volcano plots of miRNAs in LUAD (left) and LUSC (right). Black dots are the non-regulated miRNAs, red dots represent the up-regulated miRNAs, and green dots represent the down-regulated miRNAs. Threshold: -log10 (P-value) ≥ 1.3 and log2 (Fold Change) ≥ 2.

The relation between allele score and candidate miRNAs in LUAD and LUSC are presented in Table IV and Table V respectively, including the results of differential analyses (corresponding log2-fold change and P-value), and the results of linear regression between allele score and miRNA(P-value and adjusted R^2 value). Five miRNAs in LUAD were up-regulated, while in LUSC, hsa-miR-1293 was up-regulated, hsa-miR-1-1 and hsa-miR-1-2 were down-regulated. All eight miRNAs had P-values less than 0.05, indicating a strong correlation between the SNPs (allele score) and the candidate miRNA list, which met the assumption in the MR analysis that the instrumental variables must be reproducible and strongly related to exposure.

TABLE IV. CANDIDATE MIRNAS IN LUAD

MiRNAs	Log2FC	P-value[a]	P-value[b]	Adjusted R^2
hsa-miR-135b	2.779	9.59E-26	0.020	0.009
hsa-miR-142	2.680	6.04E-35	0.042	0.006
hsa-miR-182	2.550	3.90E-53	0.010	0.011
hsa-miR-183	2.356	2.09E-42	0.011	0.011
hsa-miR-3607	3.759	3.02E-32	0.026	0.008

a. P-value of differential analyses
b. P-value of linear regression

TABLE V. CANDIDATE MIRNAS IN LUSC

MiRNAs	Log2FC	P-value[a]	P-value[b]	Adjusted R^2
hsa-miR-1-2	-2.567	1.87E-26	0.013	0.011
hsa-miR-1293	3.275	8.12E-08	0.018	0.001
hsa-miR-1-1	-2.511	8.90E-26	0.032	0.008

a. P-value of differential analyses
b. P-value of linear regression

The result of Cox semiparametric hazards model was showed in Table VI, which illustrates the association between SNPs (allele score) and the survival of two sub-types of lung cancer patients. The HR in LUAD group shows that the SNP list was not associated with the survival of LUAD, with the HR being 1.012, 95% CI from 0.986 to 1.047, per allele score, and P-value being 0.397. However, in the LUSC group, the SNP list had significant association with the survival of LUSC, with the HR being 0.866 per allele score, 95% CI from 0.767 to 0.977, and P-value being 0.019. Based on the Cox semiparametric hazards model, SNP list in LUAD did not

show statistically significant association with LUAD, which meet the third assumption of MR, that instrument should not be directly associated with outcome. The SNPs among the SNP list in LUSC group showed a significant relation with the survival rate, with an average increase per allele score, the survival rate of the patients decreased by 0.866 times. Because the LUSC group did not meet the assumption of MR, it could not continue to be applied in the subsequent MR analysis.

TABLE VI. THE ASSOCIATION BETWEEN ALLELE SCORE AND SURVIVAL OF LUNG CANCER

NSCLC	Coefficient	HR[a]	95% CI[b]	P-value
LUAD	0.011	1.012	(0.986, 1.047)	0.397
LUSC	-0.144	0.866	(0.767, 0.977)	0.019

a. hazard ratio

b. 95%CI of HR

The MR analyses result was summarized in Table VII. The relation between SNPs (allele score) effect on LUAD and SNPs (allele score) effect on miRNA expression level (RPM) were not statistically significant (with p-value = 0.412). The RPM level for hsa-miR-135b and hsa-miR-3607 only changed around 2 per allele score while the original expression level were around 100 in cancer samples. RPM for hsa-miR-142, hsa-miR-182 and hsa-miR-183 changing were higher per allele score (10.750, 46.150, and 30.020 respectively), while the original RPM were from 5000 to more than 30000 in the cancer samples. Consequently, there was no clear evidence to show the causal association between the five candidate miRNAs (hsa-miR-135b, hsa-miR-142, hsa-miR-182 hsa-miR-183, and hsa-miR-3607) with LUAD.

TABLE VII. SUMMARY OF RESULTS FROM MENDELIAN RANDOMIZATION ANALYSES

MiRNA	Estimate of RPM Change	95% CI[a]	P-value
hsa-miR-135b	2.604	(-0.882, 3.613)	0.412
hsa-miR-142	10.750	(-3.643, 14.92)	
hsa-miR-182	46.150	(-64.040, 156.340)	
hsa-miR-183	30.020	(-4.167, 101.700)	
hsa-miR-3607	-2.801	(-0.949, 3.888)	

a. 95% CI of RPM Change

IV. DISCUSSION

In the MR analysis of 510 LUAD patients and 473 LUSC patients based on TCGA database, we found that there was no obvious evidence of a causal relation between the expression of hsa-miR-135b, hsa-miR-142, hsa-miR-182, hsa-miR-183 and hsa-miR-3607 and the occurrence of LUAD.

In the LUSC group, the dysregulation results of the three miRNAs (hsa-miR-1-1, hsa-miR-1-2, and hsa-miR-1293) were found. From the differential expression analysis, hsa-miR-1-1, hsa-miR-1-2 were down-regulated (with Log2FC = -2.511and -2.567, respectively), and hsa-miR-1293 was up-regulation (with Log2FC = 3.275). These 3 miRNAs were not involved in the MR analysis because of not fitting the assumptions of MR. According to the HMDD, none of these miRNAs were reported to have causal association with LUSC.

The standard method for determining causal relations is randomized controlled perturbation experiments. This way of obtaining a causal relation is widely used, however, the weakness is that the experiment is complex, and the experimental time cost is high and not necessarily definitive [27]. Therefore, we adopted a MR approach to examine the causal relation between the mentioned miRNAs and the two subtypes of lung cancer.

Our observation of no causal relation between hsa-miR-135b and hsa-miR-3607 with LUAD occurrence are consistent with other studies. However, there are studies from HMDD database showing that hsa-miR-142, hsa-miR-182 and hsa-miR-183 had causal correlation between these three miRNAs and non-small-cell lung cancer or lung tumors. Although there are many other miRNA-related databases, such as database of Differentially Expressed MiRNAs in human Cancers (dbDEMC) [28], miRCancer [29], and PhenomiR [30], these databases do not provide causal association information, so these databases can be used as a reference for correlation information (up or down regulation) and biological mechanism.

The criteria of the HMDD database for studies reflecting causal associations are as follows: a) the target miRNAs' functional acquisition/loss experiments were included in the study; b) the functional experiments should be performed on cell lines or diseased animals; c) miRNAs that only enhanced efficacy but did not contribute to the diseases were excluded [31]. Based on these criteria, there are 53 records indicate that hsa-miR-135b, hsa-miR-142, hsa-miR-182, hsa-miR-183, and hsa-miR-3607 are associated with lung cancer, including two subtypes of NSCLC: LUAD and LUSC. None of the archived previous studies reported about hsa-miR-135b and hsa-miR-3607 reported causal association results.

On the other hand, hsa-miR-142 was reported to have a causal association with NSCLC. MiR-142-3p was positively and negatively correlated with the expression of transforming growth factor beta receptor 1 (TGFβR1), a tumor suppressor, and influence the TGF-1 signal transduction pathway. Up-regulated hsa-miR-142-3p in NSCLC A549 cells suppressed expression of TGFβR1 mRNA and protein, while in the knockdown experiment, the zero expression of hsa-miR142-3p led to a completely opposite result. The expression level of TGFβR1 was increased significantly (P-value < 0.01). The expression level of the TGFβR1 protein increased or decreased in sync with the expression level of TGFβR1 mRNA in stable cells. In a downstream analysis, TGF-1-induced phosphorylation of SMAD3 (pSMAD3), an indispensable downstream effector in canonical TGF-/Smad signaling, was attenuated because of the overexpression of hsa-mir-142-3p in NSCLC cells A549, but was augmented in the down-regulated hsa-miR-142-3p environment. Thus, hsa-miR-142-3p may affect the proliferation of NSCLC cells by inhibiting the expression of TGFβR1 [32]. More studies should be done to further investigate the role of hsa-mir-142 in NSCLC because the findings from different studies have been inconsistent. Hsa-miR-142 was listed as down-regulated miRNA in HMDD in some studies. Another possible function of hsa-miR-142 was reported to down-regulate the expression of high-mobility group box 1 (HMGB1) in NSCLC patients [33]. HMGB1 was found to play roles in multiple biological processes, such as

DNA and tissue repair, cell mobility and inflammation, which was predicted to be the directly target of miR-142 in NSCLC cells.

Out of the 22 records related to hsa-miR-182 in the HMDD, one reported potential causal relation with LUAD. Programmed cell death 4 (PDCD4), a 64-kDa protein, also known as a tumor suppressor inhibiting TPA-induced neoplastic transformation and tumor promotion and progression, was reported to be the target of hsa-miR-182. Methyl thiazolyl tetrazolium and colony formation assays was performed, which provided the evidence that down-regulation of hsa-miR-182 inhibits cell proliferation in NSCLC cell lines A549 and SPC-A-1. Transwell and wound healing assays also demonstrated that down-regulation of hsa-miR-182 restricts cell invasion and migration ability of A549 and SPC-A-1 cells. Up-regulation of hsa-miR-182 may be the cause of the down-regulation of PDCD4 in lung cancer cells [34]. The mechanism of hsa-miR-182 remains to be further studied and confirmed, because there are still studies in HMDD and dbDEMC that reported that this miRNA was down-regulated in lung cancer [35].

One out of the 20 records related to hsa-miR-183 reported potential causal outcomes with LUAD. Hsa-miR-183 was reported as an up-regulated miRNA in this subpopulation in both NSCLC cell line and primary tumors. Cells marked by CD133+/CD326+ that could represent tumor-initiating cells (TICs) or cancer stem-like cells (CSLCs) of the A549 NSCLC cell line. Stable hsa-miR-183 overexpressed and knockdown CD133+/CD326+ CSLCs were established, which proved that the invasion of CD133+/CD326+ CSLCs was promoted with overexpressed hsa-miR-183, but reduced in the environment with inhibition of hsa-miR-183. Protein tyrosine phosphatase non-receptor type 4 (PTPN4), which participates in signal transduction, mediates cell growth, differentiation and regulates the function of pro-apoptotic cells. PTPN4 was reported to be the potential target gene of hsa-miR-183. There was a negative correlation between PTPN4mRNA levels and hsa-miR-183. Hsa-miR-183 played an invasive role in down-regulating PTPN4, which can be used as a therapeutic target to inhibit the migration ability of cancer stem-like cells in NSCLC [36].

To our knowledge, this study is the first known study using the TCGA data set to investigate the causal relation between miRNAs and lung cancer using the MR method, and also one of the first MR studies to use miRNA expression as an exposure factor, with the SNPs as instrumental variables. In recent years, hsa-miR-135b, hsa-miR-142, hsa-miR-182, hsa-miR-183 have been studied extensively in lung cancer research, but the number of hsa-miR-3607 related research is still limited. Same in LUSC group, hsa-miR-1-1and hsa-miR-1-2 were studied a lot, however, role of hsa-miR-1293 in LUSC is not clear. Hsa-miR-3607 and hsa-miR-1293 were significantly dysregulated in LUAD and LUSC. Therefore, this study provides a new direction for the study of miRNAs in NSCLC. In recent years, machine learning methods have been put forward, which can be used as an alternative to statistical methods, including differential expression analysis. A decision tree-based classifiers for lung cancer diagnosis and subtyping was reported [41], in which both of hsa-miR-135b and hsa-miR-183 were used as key nodes to separate lung cancer samples from normal samples. The miRNAs in the results of this study have potential of key nodes or prior knowledge in classifiers model building or other models in machine learning.

This study has a few limitations. In the sample of patients with LUAD, 86.86% were white people, 11.18% were African American, and 1.765% were from Asian, only 0.1961% were American Indian or Alaska native. In LUSC group, 87.53% were white, 9.937% were African American, and 2.537% were from Asian. There was an issue of race disparity in the current study, so the result may be more applicable to whites than to global universality. The sample capacity also limits the universality of the study to a certain extent. This problem can be solved by using larger databases or consolidating multiple databases, which also provides the ability to extract random samples for validation analysis. In the association test between SNPs (allele score) and miRNAs (RPM), the correlation conclusion could be obtained from the p-value, but it could be seen from the F-parameter which represent the strength of the instrumental variables were not large. The F-parameter for hsa-miR-135b, hsa-miR-142, hsa-miR-182, hsa-miR-183 and hsa-miR-3607 were 5.436, 4.152, 6.636, 6.566, and 5.007, respectively. Adjusting the way of defying the allele score may have the possibility of improving the SNPs' ability (as instrument variable). In this study, all the SNPs emerging were treated equally. But in fact, the influence ability of different SNPs on disease is different, and the allele frequency varies among SNPs. Therefore, meta-analysis that focusing on the influence ability and allele frequency of SNPs may be needed. The construction of a new allele score computing network, which can be weighted according to the influence of SNP and allele frequency on the basis of the original score, may enhance the utility of SNPs as an instrument variable.

V. Conclusion

Mendelian randomization analysis indicated that hsa-miR-135b, hsa-miR-142, hsa-miR-182, hsa-miR-183 and hsa-miR-3607 were not causally associated with the risk of LUAD.

References

[1] D. Wang, J. Wang, M. Lu, F. Song and Q. Cui, "Inferring the human miRNA functional similarity and functional network based on miRNA-associated diseases", Bioinformatics, vol. 26, no. 13, pp. 1644-1650, 2010. Available: 10.1093/bioinformatics/btq241.

[2] "What Is Lung Cancer? | Types of Lung Cancer", Cancer.org, 2019. [Online]. Available: https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/what-is-non-small-cell-lung-cancer.html.

[3] J. Subramanian and R. Govindan, "Lung Cancer in Never Smokers: A Review", Journal of Clinical Oncology, vol. 25, no. 5, pp. 561-570, 2007. Available: 10.1200/jco.2006.06.8015.

[4] Y. Sun et al., "Serum 25-hydroxyvitamin D levels and risk of lung cancer and histologic types: a Mendelian randomisation analysis of the HUNT study", European Respiratory Journal, vol. 51, no. 6, p. 1800329, 2018. Available: 10.1183/13993003.00329-2018.

[5] S. Kenfield, E. Wei, M. Stampfer, B. Rosner and G. Colditz, "Comparison of aspects of smoking among the four histological types of lung cancer", Tobacco Control, vol. 17, no. 3, pp. 198-204, 2008. Available: 10.1136/tc.2007.022582.

[6] W. Cho, C. Kwan, S. Yau, P. So, P. Poon and J. Au, "The role of inflammation in the pathogenesis of lung cancer", Expert Opinion on Therapeutic Targets, vol. 15, no. 9, pp. 1127-1137, 2011. Available: 10.1517/14728222.2011.599801.

[7] T.-Y. Li, F. Zhang. Screening of lung cancer related SNPs and CNVs with SNP microarrays. Eur Rev Med Pharmacol Sci. 2015. Vol. 19 - N. 2. 225-234.

[8] V. Onay et al., "SNP-SNP interactions in breast cancer susceptibility", BMC Cancer, vol. 6, no. 1, 2006. Available: 10.1186/1471-2407-6-114.

[9] A. Alberg, M. Brock, J. Ford, J. Samet and S. Spivack, "Epidemiology of Lung Cancer", Chest, vol. 143, no. 5, pp. e1S-e29S, 2013. Available: 10.1378/chest.12-2345.

[10] T. Mah et al., "Novel SNP improves differential survivability and mortality in non-small cell lung cancer patients", BMC Genomics, vol. 15, no. 9, 2014. Available: 10.1186/1471-2164-15-s9-s20.

[11] G. Etokebe et al., "Association of the FAM46A Gene VNTRs and BAG6 rs3117582 SNP with Non-Small Cell Lung Cancer (NSCLC) in Croatian and Norwegian Populations", PLOS ONE, vol. 10, no. 4, p. e0122651, 2015. Available: 10.1371/journal.pone.0122651.

[12] G. Li, J. Luo, Q. Xiao, C. Liang, P. Ding and B. Cao, "Predicting MiRNA-Disease Associations Using Network Topological Similarity Based on DeepWalk", IEEE Access, vol. 5, pp. 24032-24039, 2017. Available: 10.1109/access.2017.2766758.

[13] P. Kumar et al., "Circulating miRNA Biomarkers for Alzheimer's Disease", PLoS ONE, vol. 8, no. 7, p. e69807, 2013. Available: 10.1371/journal.pone.0069807.

[14] J. Wang et al., "Identification of a Circulating MicroRNA Signature for Colorectal Cancer Detection", PLoS ONE, vol. 9, no. 4, p. e87451, 2014. Available: 10.1371/journal.pone.0087451.

[15] Y. Zhu et al., "Distinctive microRNA signature is associated with the diagnosis and prognosis of acute leukemia", Medical Oncology, vol. 29, no. 4, pp. 2323-2331, 2011. Available: 10.1007/s12032-011-0140-5.

[16] B. Chen, T. Gao, W. Yuan, W. Zhao, T. Wang and J. Wu, "Prognostic Value of Survival of MiRNAs Signatures in Non-small Cell Lung Cancer", Journal of Cancer, vol. 10, no. 23, pp. 5793-5804, 2019. Available: 10.7150/jca.30336.

[17] W. Zhu et al., "Diagnostic Value of Serum miR-182, miR-183, miR-210, and miR-126 Levels in Patients with Early-Stage Non-Small Cell Lung Cancer", PLOS ONE, vol. 11, no. 4, p. e0153046, 2016. Available: 10.1371/journal.pone.0153046.

[18] J. Jiang, "Real-time expression profiling of miRNA precursors in human cancer cell lines", Nucleic Acids Research, vol. 33, no. 17, pp. 5394-5403, 2005. Available: 10.1093/nar/gki863.

[19] S. Volinia et al., "A miRNA expression signature of human solid tumors defines cancer gene targets", Proceedings of the National Academy of Sciences, vol. 103, no. 7, pp. 2257-2261, 2006. Available: 10.1073/pnas.0510565103.

[20] H. Arora, R. Qureshi, S. Jin, A. Park and W. Park, "miR-9 and let-7g enhance the sensitivity to ionizing radiation by suppression of NFκB1", Experimental and Molecular Medicine, vol. 43, no. 5, p. 298, 2011. Available: 10.3858/emm.2011.43.5.031.

[21] M. Cheng et al., "A miRNA-135a/b binding polymorphism in CD133 confers decreased risk and favorable prognosis of lung cancer in Chinese by reducing CD133 expression", Carcinogenesis, vol. 34, no. 10, pp. 2292-2299, 2013. Available: 10.1093/carcin/bgt181.

[22] Z. Huang et al., "HMDD v3.0: a database for experimentally supported human miRNA–disease associations", Nucleic Acids Research, vol. 47, no. 1, pp. D1013-D1017, 2018. Available: 10.1093/nar/gky1010.

[23] G. Davey Smith and S. Ebrahim, "'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*", International Journal of Epidemiology, vol. 32, no. 1, pp. 1-22, 2003. Available: 10.1093/ije/dyg070.

[24] Z. Lei et al., "MiR-142-3p represses TGF-β-induced growth inhibition through repression of TGFβR1 in non-small cell lung cancer", The FASEB Journal, vol. 28, no. 6, pp. 2696-2704, 2014. Available: 10.1096/fj.13-247288

[25] P. Sekula, F. Del Greco M, C. Pattaro and A. Köttgen, "Mendelian Randomization as an Approach to Assess Causality Using Observational Data", Journal of the American Society of Nephrology, vol. 27, no. 11, pp. 3253-3265, 2016. Available: 10.1681/asn.2016010098.

[26] P. Royston, "Multiple Imputation of Missing Values", The Stata Journal: Promoting communications on statistics and Stata, vol. 4, no. 3, pp. 227-241, 2004. Available: 10.1177/1536867x0400400301.

[27] T. Le et al., "Inferring microRNA–mRNA causal regulatory relationships from expression data", Bioinformatics, vol. 29, no. 6, pp. 765-771, 2013. Available: 10.1093/bioinformatics/btt048.

[28] Z. Yang et al., "dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers", Nucleic Acids Research, vol. 45, no. 1, pp. D812-D818, 2016. Available: 10.1093/nar/gkw1079.

[29] B. Xie, Q. Ding, H. Han and D. Wu, "miRCancer: a microRNA-cancer association database constructed by text mining on literature", Bioinformatics, vol. 29, no. 5, pp. 638-644, 2013. Available: 10.1093/bioinformatics/btt014.

[30] A. Ruepp et al., "PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes", Genome Biology, vol. 11, no. 1, p. R6, 2010. Available: 10.1186/gb-2010-11-1-r6.

[31] Y. Gao, K. Jia, J. Shi, Y. Zhou and Q. Cui, "A Computational Model to Predict the Causal miRNAs for Diseases", Frontiers in Genetics, vol. 10, 2019. Available: 10.3389/fgene.2019.00935.

[32] Z. Lei et al., "MiR-142-3p represses TGF-β-induced growth inhibition through repression of TGFβR1 in non-small cell lung cancer", The FASEB Journal, vol. 28, no. 6, pp. 2696-2704, 2014. Available: 10.1096/fj.13-247288

[33] X. Peng and W. Liu, "MiR-142-3p functions as a potential tumor suppressor directly targeting HMGB1 in non-small-cell lung carcinoma", International journal of clinical and experimental pathology, vol. 8, no. 9, pp. 10800-108007, 2015.

[34] M. Wang et al., "Downregulation of microRNA-182 inhibits cell growth and invasion by targeting programmed cell death 4 in human lung adenocarcinoma cells", Tumor Biology, vol. 35, no. 1, pp. 39-46, 2013. Available: 10.1007/s13277-013-1004-8.

[35] J. Lu et al., "MicroRNA expression profiles classify human cancers", Nature, vol. 435, no. 7043, pp. 834-838, 2005. Available: 10.1038/nature03702.

[36] C. Zhu et al., "MicroRNA-183 promotes migration and invasion of CD133+/CD326+ lung adenocarcinoma initiating cells via PTPN4 inhibition", Tumor Biology, vol. 37, no. 8, pp. 11289-11297, 2016. Available: 10.1007/s13277-016-4955-8.

[37] L. Tanoue, "Systematic review of the relationship between family history and lung cancer risk", Yearbook of Medicine, vol. 2007, pp. 265-266, 2007. Available: 10.1016/s0084-3873(08)70179-2.

[38] J. Wei, F. Li, J. Yang, X. Liu and W. Cho, "MicroRNAs as regulators of airborne pollution-induced lung inflammation and carcinogenesis", Archives of Toxicology, vol. 89, no. 5, pp. 677-685, 2015. Available: 10.1007/s00204-015-1462-4.

[39] W. Cho, A. Chow and J. Au, "MiR-145 inhibits cell proliferation of human lung adenocarcinoma by targeting EGFR and NUDT1", RNA Biology, vol. 8, no. 1, pp. 125-131, 2011. Available: 10.4161/rna.8.1.14259.

[40] W. Cho, A. Chow and J. Au, "Restoration of tumour suppressor hsa-miR-145 inhibits cancer cell growth in lung adenocarcinoma patients with epidermal growth factor receptor mutation", European Journal of Cancer, vol. 45, no. 12, pp. 2197-2206, 2009. Available: 10.1016/j.ejca.2009.04.039.

[41] M. Sherafatian and F. Arjmand, "Decision tree-based classifiers for lung cancer diagnosis and subtyping using TCGA miRNA expression data", Oncology Letters, 2019. Available: 10.3892/ol.2019.10462.

[42] J. Xue, J. Yang, M. Luo, W. Cho and X. Liu, "MicroRNA-targeted therapeutics for lung cancer treatment", Expert Opinion on Drug Discovery, vol. 12, no. 2, pp. 141-157, 2016. Available: 10.1080/17460441.2017.1263298.

[43] F. Wang, F. Meng, L. Wang, S. Wong, W. Cho and L. Chan, "Associations of mRNA:microRNA for the Shared Downstream Molecules of EGFR and Alternative Tyrosine Kinase Receptors in Non-small Cell Lung Cancer", Frontiers in Genetics, vol. 7, 2016. Available: 10.3389/fgene.2016.00173.

[44] B. He et al., "The Association between Four Genetic Variants in MicroRNAs (rs11614913, rs2910164, rs3746444, rs222832) and Cancer Risk: Evidence from Published Studies", PLoS ONE, vol. 7, no. 11, p. e49032, 2012. Available: 10.1371/journal.pone.0049032.

[45] W. Cho, "Role of miRNAs in lung cancer", Expert Review of Molecular Diagnostics, vol. 9, no. 8, pp. 773-776, 2009. Available: 10.1586/erm.09.57.