MALAT1 rs619586 A/G polymorphisms are associated with decreased risk of lung cancer

Ming Chen, MD*, Deng Cai, PhD, Haiyong Gu, MD, PhD, Jun Yang, MD, Liming Fan, MD

Abstract

Lung cancer is the leading cause of cancer-associated mortality worldwide. Genetic factors are reported to play important roles in lung carcinogenesis. To evaluate genetic susceptibility, we conducted a hospital-based case-control study on the effects of functional single nucleotide polymorphisms (SNPs) in long non-coding RNAs (lncRNAs) and microRNAs on lung cancer development. A total of 917 lung cancer cases and 925 control subjects were recruited. The MALAT1 rs619586A/G genotype frequencies between patient and control groups were significantly different (P < .001), specifically, 83.85% vs 75.88% (AA), 15.60% vs 21.79% (AG), and 0.55% vs 2.32% (GG). When the homozygous genotype MALAT1 rs619586 AA was used as the reference group, AG (AA vs AA: adjusted odds ratio [OR] 0.65, 95% confidence interval [CI] 0.51–0.83, P = .001) and GG genotypes were associated with significantly decreased risk of lung cancer (GG vs AA: adjusted OR 0.22, 95% CI 0.08–0.59, P = .003). In the dominant model, MALAT1 rs619586 AG/GG variants were also associated with a significantly decreased risk of lung cancer (adjusted OR 0.61, 95% CI 0.48–0.78, P < .001). In the recessive model, when MALAT1 rs619586 AA/AG genotypes were used as the reference genotype, the GG homozygous genotype was also associated with significantly decreased risk of lung cancer (adjusted OR 0.24, 95% CI 0.09–0.64, P = .004). Hsa-miR-34b/c rs4938723 T > C, pri-miR-124-1 rs531564 C > G and hsa-miR-423 rs6505162 C > A SNPs were not associated with lung cancer risk. Our collective data indicated that MALAT1 rs619586 A/G SNPs significantly reduced the risk of lung cancer. Large-scale studies on different ethnic populations and tissue-specific biological characterization are required to validate the current findings.

Abbreviations: CI = confidence interval, lncRNAs = long non-coding RNAs, miRNA = microRNA, OR = odds ratio, SNPs = single nucleotide polymorphisms.

Keywords: long non-coding RNAs, lung cancer, microRNA, molecular epidemiology, polymorphisms

1. Introduction

Long non-coding RNAs (lncRNAs) are non-protein coding transcripts ranging from 200 bases to 100 kb involved in all aspects of gene regulation and biological processes. Under different physiological and pathological conditions, lncRNAs have diverse functions. lncRNAs regulate gene expression processes, including chromatin modification,[11] transcription and posttranscriptional processing[2] at various levels. Several lncRNAs have modulatory effects on cell homeostasis and proliferation while others function in apoptosis.[3] More recent studies have demonstrated critical roles of lncRNAs in carcinogenesis. In cancer cells, lncRNAs regulate transcriptional, posttranscriptional and epigenetic levels, the important cellular signaling pathways.[4] Most lncRNAs are RNA polymerase (Pol) IV/Pol I-transcribing, while others transcribe RNA Pol III.[5] LncRNAs are involved in diverse cellular functions[6-7] as well as different mechanisms, with roles as decoys, guides and scaffolds.[8] Aberrant lncRNA expression contributes to progression of numerous tumors[9] and is considered an early event in some tumor types. A role of specific lncRNAs in glioma carcinogenesis has been reported based on data from microarray analysis.[10] LncRNAs additionally have important functions in lung, breast, and liver cancer development.[11] The well-characterized metastasis-associated lung adenocarcinoma transcript-1 (MALAT1) lncRNA is a nuclear-enriched abundant transcript expressed in the lungs, pancreas, nerve system and other healthy organs.[12] High expression of MALAT1 has also been detected in various cancer types, including lung cancer, endometrial stromal sarcoma, hepatocellular carcinoma, breast cancer and pancreatic cancer. Elevated expression of MALAT1 is associated with
hyperproliferation, metastasis, and poor prognosis. MALAT1 localizes to nuclear speckles, a subnuclear domain suggested to coordinate RNA polymerase II transcription, pre-mRNA splicing, and mRNA export.[13] Moreover, MALAT1 interacts with several pre-mRNA splicing factors including serine-arginine dipeptide-rich SR family splicing factors, such as SRSF1 (also known as ASF/ SF2), SC35 (SRSp2), and SRSF3. The IncRNA further induces the expression of cell cycle genes and controls alternative splicing of pre-mRNAs by modulating the intranuclear distribution of SR splicing factors.[14] Interestingly, knockdown of MALAT1 has no impact on the formation, size, and number of nuclear speckles but results in decreased nuclear speckle association of several pre-mRNA splicing factors, including SRSF1.[10]

MicroRNAs (miRNAs) are tiny non-coding RNAs that act as posttranscriptional gene regulatory elements.[15] MiRNAs exert their effects by binding to the 3’related of target genes and downregulating their expression[16] and are reported to be important players in carcinogenesis.[17]

Genetic factors, such as single nucleotide polymorphisms (SNPs), may contribute to carcinogenesis.[18] SNPs in genomic miRNA sequences could influence miRNA-dependent regulation, affect the final levels and functions of miRNAs, and consequently alter tumor susceptibility.[19]

Members of the miR-34 family are direct p53 targets induced in response to DNA damage or oncogenic stress.[20] Downregulation of mir-34b/c via methylation has been reported in colorectal cancer,[21] oral cancer,[22] and malignant melanoma.[23] Hsa-miR-34b/c rs4938723 SNP is located within the CpG island of pri-miR-34b/c and 423bp upstream from the transcription start site and is proposed to serve as the predicted binding site for GATA-X transcription factors.[24] The Hsa-miR-34b/c polymorphism is associated with risk of nasopharyngeal carcinoma,[25] hepatocellular carcinoma,[26] colorectal cancer[27] and breast cancer survival.[28] A polymorphism is reported to confer reduced blood sample collection, genomic DNA isolation and SNP genotyping were conducted using the ligation detection reaction (LDR) method with technical support from Shanghai Biowing Applied Biotechnology Company, as described previously.[14] The quality of genotyping for MALAT1 rs619586 A/G, miR-34b/c rs4938723 T>C, pri-miR-124-1 rs531564 C>G and hsa-miR-423 rs6505162 C>A was high. For quality control, repeated analyses were conducted using 184 (10%) randomly selected samples with high DNA quality.

2.2. Isolation of DNA and genotyping using ligation detection reaction

Blood sample collection, genomic DNA isolation and SNP genotyping were performed using the ligation detection reaction (LDR) method with technical support from Shanghai Biowing Applied Biotechnology Company, as described previously.[14] The quality of genotyping for MALAT1 rs619586 A/G, miR-34b/c rs4938723 T>C, pri-miR-124-1 rs531564 C>G and hsa-miR-423 rs6505162 C>A was high. For quality control, repeated analyses were conducted using 184 (10%) randomly selected samples with high DNA quality.

2.3. Statistical analyses

Student t test and χ2 test were performed to assess the differences in distribution of selected variables, demographic characteristics, and genotypes for the 4 SNPs between lung cancer cases and controls. Using logistic regression analyses, the correlations between the 4 SNPs and risk of lung cancer were evaluated by calculating the crude odds ratio (ORs), adjusted ORs and corresponding 95% confidence intervals [CIs]. Hardy-Weinberg equilibrium (HWE) in controls was tested with an online calculator (http://ihg.gsf.de/cgi-bin/hw/hwa1.pl). Statistical analyses were performed with SAS (v 9.1.3) software (SAS Institute, Cary, NC, USA).

3. Results

3.1. Characteristics of the study population

The characteristics of cases and control subjects recruited for study are summarized in Table 1. In terms of age and sex, cases and controls appeared adequately matched (P = .467 and P = .095, respectively), as determined with the χ2 test. No significant difference was detected in smoking rate (P = .263) and drinking status (P = .284) between the 2 groups, as shown in Table 1. Primary information on MALAT1 rs619586 A/G SNPs is provided in Table 2. The genotyping success rate for MALAT1 rs619586 A/G was 98.15% in all 1842 samples. The concordance rates of repeat analyses were 100% for both SNPs. Minor allele frequencies (MAFs) in our controls were similar to MAFs of these SNPs recorded in the Chinese database (Table 2). The observed genotype frequency for MALAT1 rs619586 A/G polymorphisms was 0.132 in the controls in HWE (P = .131).

3.2. Associations between MALAT1 rs619586 A/G polymorphisms and the risk of lung cancer

The genotype frequencies of MALAT1 rs619586 A/G were 83.85% (AA), 15.60% (AG), and 0.55% (GG) in the patient

2. Materials and methods

2.1. Isolation of DNA and genotyping

This case-control study was approved the Ethical Committee on Human Studies, Shanghai Chest Hospital (Shanghai, China). Written informed consent was provided by the participants. Subjects were selected from Shanghai Chest Hospital. Between April 2015 and October 2016, 917 non-small cell lung cancer patients were recruited consecutively, including 801 adenocarcinoma and 116 squamous cell carcinoma cases. All lung cancer cases were diagnosed using pathological methods. Exclusion criteria were as follows: patients previously diagnosed with cancer, small-cell lung cancer, any metastasized cancer and radiotherapy or chemotherapy. The study included 917 lung cancer cases and 923 cancer-free controls. Demographic data were collected from each subject using a pre-tested questionnaire, including sex, age at diagnosis, race, and related risk factors (including tobacco smoking and alcohol consumption).
group and 75.88% (AA), 21.79% (AG), and 2.32% (GG) in the control group, which were significantly different (P < .001). When the MALAT1 rs619586 AA homozygous genotype was used as the reference group, the AG genotype was associated with significantly decreased risk of lung cancer (AG vs AA: adjusted OR: 0.65, 95% CI: 0.51–0.83, P = .001) as well as the GG genotype (GG vs AA: adjusted OR: 0.22, 95% CI: 0.08–0.59, P = .003). In the dominant model, MALAT1 rs619586 AG/GG variants were associated with significantly decreased risk of lung cancer, compared with the MALAT1 rs619586 AA genotype (adjusted OR: 0.61, 95% CI: 0.48–0.78, P < .001). In the recessive model, when MALAT1 rs619586 AA/AG genotypes were used as the reference group, the GG homozygous genotype was also associated with significantly decreased risk of lung cancer (adjusted OR: 0.24, 95% CI: 0.09–0.64, P = .004) (Table 3).

### Table 1

**Distribution of selected demographic variables and risk factors in lung cancer cases and controls.**

| Variable                      | Cases (n=917) | Controls (n=925) | P*  |
|-------------------------------|---------------|------------------|-----|
| Age (yrs)                     | n  | %      | n  | %      |     |
| <60                           | 379 | 41.2   | 397 | 42.9   | .467|
| ≥60                           | 530 | 58.8   | 528 | 57.1   |     |
| Age, yrs, mean ± SD           | 59.78 (±10.88)| 60.06 (±7.58) | .521|
| Sex                           |     |        |     |        | .095|
| Men                           | 517 | 56.4   | 557 | 60.2   |
| Women                         | 400 | 43.6   | 368 | 39.8   |
| Tobacco use                   |     |        |     |        | .263|
| Never                         | 666 | 72.6   | 650 | 70.3   |
| Ever                          | 251 | 27.4   | 275 | 29.7   |
| Alcohol use                   |     |        |     |        | .284|
| Never                         | 674 | 73.5   | 700 | 75.7   |
| Ever                          | 243 | 26.5   | 225 | 24.3   |
| Cancer pathology types        |     |        |     |        |     |
| Adenocarcinoma                | 801 | 87.4   |     |        |
| Squamous cell carcinoma       | 116 | 12.6   |     |        |

3.3. Associations between hsa-miR-34b/c rs4938723 T > C, pri-miR-124-1 rs531564 C > G and hsa-miR-423 rs6505162 C > A polymorphisms and the risk of lung cancer

The genotype distributions of hsa-miR-34b/c rs4938723 T > C, pri-miR-124-1 rs531564 C > G and hsa-miR-423 rs6505162 C > A in cases and control subjects are shown in Table 3. In single locus analyses, the genotype frequencies of hsa-miR-34b/c rs4938723 T > C were 44.4% (TT), 45.1% (TC), and 10.5% (CC) in patients and 42.2% (TT), 43.4% (TC), and 10.5% (CC) in control subjects. The difference between the 2 groups was not statistically significant (P = .746). In the recessive model, when hsa-miR-34b/c rs4938723 TT/TC genotypes were used as the reference group, the CC homozygous genotype was not associated with risk of lung cancer (CC vs TT/TC: adjusted OR = 1.17, 95% CI: 0.86–1.59, P = .352). In the recessive model, when hsa-miR-34b/c rs4938723 TT/TC genotypes were used as the reference group, the CC homozygous genotype was also associated with significantly decreased risk of lung cancer (CC vs TT/TC: adjusted OR = 0.50, 95% CI: 0.32–0.78, P = .004) (Table 3).

### Table 2

**Primary information for MALAT1 rs619586 A/G, pri-miR-124-1 rs531564 C > G, hsa-miR-34b/c rs4938723 T > C and hsa-miR-423 rs6505162 C > A polymorphisms.**

| Genotyped SNPs | MALAT1 rs619586 A/G | pri-miR-124-1 rs531564 C > G | hsa-miR-34b/c rs4938723 T > C | hsa-miR-423 rs6505162 C > A |
|----------------|---------------------|-----------------------------|-----------------------------|-----------------------------|
| Chromosome     | 11                  | 8                           | 11                          | 17                          |
| Gene Official Symbol | ncRNA             | MIR124-1                    | MIR34B/C                    | MIR423                      |
| Function       | 65430968             | ncRNA                       | ncRNA                       | ncRNA                       |
| Chr Pos (Genome Build 36.3) | 4                 | 9798109                     | 11087775                    | 25468309                    |
| Regulome DB Score† | Y                      | 5                           | 5                           | 1f                          |
| TFBS†          | —                   | Y                           | Y                           | Y                           |
| Splicing (ESE or ESS) | —                   | —                           | —                           | —                           |
| MAF‡ for Chinese in database | 0.123              | 0.135                       | 0.325                       | 0.187                       |
| MAF in our controls (n=925) | 0.132              | 0.154                       | 0.322                       | 0.198                       |
| P value for HWE§ test in our controls | 0.131              | 0.091                       | 0.880                       | 0.412                       |
| Genotyping method¶ | LDR                 | LDR                         | LDR                         | LDR                         |
| % Genotyping value | 98.15%              | 98.26%                      | 99.40%                      | 96.63%                      |

* http://www.regulomedb.org/.
† TFBS = Transcription Factor Binding Site (http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm).
‡ MAF = minor allele frequency, from gnomAD-Exomes Asian.
§ HWE = Hardy–Weinberg equilibrium.
¶ LDR = Ligation Detection Reaction.
Moreover, no association was observed between MALAT1 rs619586 A/G polymorphisms and the risk of lung cancer (Table 3). In this hospital-based case-control study, we investigated the potential correlations of MALAT1 rs619586 A/G, pri-miR-124-1 rs531564 C > G, hsa-miR-34b/c rs4938723 T > C and hsa-miR-423 rs6505162 C > A polymorphisms with susceptibility to lung cancer. Data from our multivariable logistic analyses supported the association of MALAT1 rs619586 A/G polymorphisms with a decreased risk of lung cancer.

**4. Discussion**

In this hospital-based case-control study, we investigated the potential correlations of MALAT1 rs619586 A/G, pri-miR-124-1 rs531564 C > G, hsa-miR-34b/c rs4938723 T > C, pri-miR-124-1 rs531564 C > G and hsa-miR-423 rs6505162 C > A polymorphisms with susceptibility to lung cancer. Data from our multivariable logistic analyses supported the association of MALAT1 rs619586 A/G polymorphisms with a decreased risk of lung cancer.

LncRNAs serve as precursors of small non-coding RNAs to produce microRNAs (miRNA) and endogenous small interfering RNAs or as a "miRNA sponge" to inhibit miRNA activity. LncRNAs also act as scaffolds during the formation of cellular substructures or protein complexes. Several LncRNAs have been shown to function as oncogenes or tumor suppressors. Previous research suggests that LncRNAs play integral roles in control of cellular growth, division and differentiation and use various mechanisms to control the cancer state. Perturbation of LncRNA expression can contribute to the development and progression of cancer. MALAT1 is a nuclear-enriched abundant transcript expressed in the lung, pancreas, nerve system and other healthy organs. Elevated expression of highly conserved lncRNA can contribute to the development and progression of cancer. MALAT1 is a nuclear-enriched abundant transcript expressed in the lung, pancreas, nerve system and other healthy organs. Elevated expression of highly conserved lncRNA can contribute to the development and progression of cancer.

**Table 3**

Logistic regression analyses of associations between MALAT1 rs619586 A/G, pri-miR-124-1 rs531564 C > G, hsa-miR-34b/c rs4938723 T > C and hsa-miR-423 rs6505162 C > A polymorphisms and risk of lung cancer.

| Genotype       | Cases (n=917) | Controls (n=925) | Crude OR (95% CI) | P     | Adjusted OR* (95% CI) | P     |
|----------------|--------------|------------------|------------------|-------|----------------------|-------|
| MALAT1 rs619586 A/G |              |                  |                  |       |                      |       |
| AA             | 758          | 686              | 75.88            | 1.00  | 1.00 (reference value) |      |
| AG             | 141          | 197              | 21.79            | 0.65  | 0.51–0.82            | <.001 |
| GG             | 5            | 21               | 2.32             | 0.22  | 0.08–0.58            | .002  |
| GG vs AG vs AA |              |                  |                  |       |                      |       |
| A/G/G          | 146          | 218              | 24.12            | 0.61  | 0.48–0.77            | <.001 |
| AA/AG          | 899          | 883              | 97.68            | 1.00  | 1.00 (reference value) |      |
| GG             | 5            | 21               | 2.32             | 0.23  | 0.09–0.62            | .004  |
| G allele       | 151          | 239              | 13.22            |       | 0.24 (0.09–0.64)     | .004  |

| pri-miR-124-1 rs531564 C > G |              |                  |                  |       |                      |       |
| CC             | 672          | 648              | 72.2             | 1.00  | 1.00 (reference value) |      |
| CG             | 214          | 221              | 24.6             | 0.93  | 0.75–1.16            | .535  |
| GG             | 27           | 28               | 3.1              | 0.93  | 0.54–1.60            | .792  |
| GG vs CG vs CC |              |                  |                  |       |                      |       |
| CG/GG          | 241          | 249              | 27.8             | 0.93  | 0.76–1.15            | .514  |
| CC + CG        | 896          | 869              | 96.9             | 1.00  | 0.55–1.62            | .869  |
| GG             | 27           | 28               | 3.1              | 0.95  | 0.55–1.62            | .839  |
| G allele       | 268          | 277              | 15.4             |       | 0.96 (0.56–1.64)     | .899  |

| hsa-miR-34b/c rs4938723 T > C |              |                  |                  |       |                      |       |
| TT             | 406          | 422              | 46.1             | 1.00  | 1.00 (reference value) |      |
| TC             | 413          | 398              | 43.4             | 1.08  | 0.89–1.31            | .444  |
| CC             | 96           | 96               | 10.5             | 1.04  | 0.76–1.42            | .809  |
| CC vs TC vs TT |              |                  |                  |       |                      |       |
| TC + CC        | 509          | 494              | 53.9             | 1.07  | 0.89–1.29            | .465  |
| TT + TC        | 819          | 820              | 89.5             | 1.00  | 1.00                  |      |
| CC             | 96           | 96               | 10.5             | 1.00  | 0.74–1.35            | .994  |
| C allele       | 605          | 590              | 32.2             |       | 1.00 (0.74–1.35)     | .999  |

| hsa-miR-423 rs6505162 C > A |              |                  |                  |       |                      |       |
| CC             | 573          | 571              | 63.9             | 1.00  | 1.00 (reference value) |      |
| CA             | 277          | 291              | 32.6             | 0.95  | 0.78–1.16            | .607  |
| AA             | 37           | 31               | 3.5              | 1.19  | 0.79–1.64            | .490  |
| AA vs CA vs CC |              |                  |                  |       |                      |       |
| CA + AA        | 314          | 322              | 36.1             | 0.97  | 0.80–1.18            | .772  |
| CC + CA        | 850          | 862              | 96.5             | 1.00  | 0.80–1.18            | .763  |
| AA             | 37           | 31               | 3.5              | 1.21  | 0.74–1.97            | .443  |
| A allele       | 351          | 353              | 19.8             |       | 1.24 (0.76–2.01)     | .396  |

*Adjusted for age, sex, smoking and drinking status.
Moreover, owing to the moderate sample sizes evaluated, our findings, with larger samples and detailed individual information should be undertaken. Finally, because lung cancer risk is affected by multiple environmental factors, gene-gene and gene-environment interactions, MALAT1, hsa-miR-34b/c, pri-miR-124-1 and hsa-miR-423 may be associated with differential degrees of genetic risk in different ethnicities and upon exposure to diverse environment-related risk factors.

In summary, our results provide evidence that MALAT1 rs619586 A/G functional polymorphisms may serve as susceptibility loci for lung cancer. Further studies are required to validate or refute the results of this preliminary study.

Acknowledgments

We appreciate all patients who participated in this study.

Author contributions

Conceptualization: Haiyong Gu, Ming Chen, Jun Yang.

Data curation: Haiyong Gu, Deng Cai, Jun Yang, Liming Fan.

Formal analysis: Deng Cai.

Investigation: Deng Cai.

Methodology: Ming Chen, Liming Fan.

Project administration: Liming Fan.

Writing – original draft: Haiyong Gu, Ming Chen, Liming Fan.

Writing – review & editing: Jun Yang.
Ye Y, Wang KK, Gu J, et al. Genetic variations in microRNA-related genes are novel susceptibility loci for esophageal cancer risk. Cancer Prev Res (Phila) 2008;1:460–9.

Smith RA, Jedlinski DJ, Gabrovska PN, et al. A genetic variant located in miR-423 is associated with reduced breast cancer risk. Cancer Genomics Proteomics 2012;9:115–8.

Xing J, Wan S, Zhou F, et al. Genetic polymorphisms in pre-microRNA genes as prognostic markers of colorectal cancer. Cancer Epidemiol Biomarkers Prev 2012;21:217–27.

Gong WJ, Yin JY, Li XP, et al. Association of well-characterized lung cancer lncRNA polymorphisms with lung cancer susceptibility and platinum-based chemotherapy response. Tumour Biol 2016;37:8349–58.

Chen ZJ, Zhao H, He L, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. Nat Genet 2011;43:55–9.

Gutschner T, Dedeken S. The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol 2012;9:703–19.

Wang J, Liu X, Wu H, et al. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res 2010;38:5366–83.

Chen et al. Medicine (2021) 100:12

Clemson CM, Hutchinson JN, Sara SA, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell 2009;33:717–26.

Vogt M, Munding J, Gruner M, et al. Frequent concomitant inactivation of miR-34a and miR-34b/c by CpG methylation in colorectal, pancreatic, mammary, ovarian, urothelial, and renal cell carcinomas and soft tissue sarcomas. Virchows Arch 2011;458:313–22.

He L, He X, Lim LP, et al. A microRNA component of the p53 tumour suppressor network. Nature 2007;447:1130–4.

He L, He X, Lowe SW, et al. microRNAs join the p53 network–another piece in the tumour-suppression puzzle. Nat Rev Cancer 2007;7:819–22.

Kim NH, Kim HS, Kim NG, et al. p53 and microRNA-34 are suppressors of canonical Wnt signaling. Sci Signal 2011;4:ra71.

Tazawa H, Tsuchiya N, Izumiya M, et al. Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. Proc Natl Acad Sci U S A 2007;104:15472–7.

Mathe E, Nguyen GH, Funamizu N, et al. Inflammation regulates microRNA expression in cooperation with p53 and nitric oxide. Int J Cancer 2012;131:760–5.