



ARTICLE

HOTAIR lncRNA SNPs rs920778 and rs1899663 are associated with smoking, male gender, and squamous cell carcinoma in a Chinese lung cancer population

Cong Wang¹, Ying Li², Yong-wen Li², Hong-bing Zhang¹, Hao Gong¹, Yin Yuan¹, Wei-ting Li¹, Hong-yu Liu² and Jun Chen^{1,2}

The abnormal expression of the long noncoding RNA (lncRNA) HOX transcript intergenic antisense RNA (HOTAIR) plays an important role in the development of various cancers; however, single nucleotide polymorphisms (SNPs) in HOTAIR and their association with primary lung cancer susceptibility have not yet been reported. Here, we performed a case-control study including 262 primary lung cancer patients and 451 cancer-free control individuals to investigate the association between four haplotype-tagging SNPs (rs920778, rs12826786, rs4759314, and rs1899663) in the HOTAIR lncRNA and the risk of developing primary lung cancer. We found a significant association between the SNPs rs920778 and rs1899663 in the HOTAIR and primary lung cancer susceptibility ($P < 0.05$). Moreover, homozygous C/T (C/T + TT) for rs920778 (C > T) sites was significantly associated with gender, smoking history, and pathological type. In addition, linkage disequilibrium and haplotype analysis of HOTAIR gene polymorphisms for susceptibility to lung cancer revealed a high degree of linkage disequilibrium between the rs920778 and rs1899663 loci ($D' = 0.86$, $r^2 = 0.52$). The population of rs920778, rs1899663, and rs4759314 had a significantly increased risk of lung cancer ($P < 0.001$). In summary, the present study provides persuasive evidence that SNP rs920778 is closely correlated with susceptibility to primary lung cancer. Future studies are warranted to validate and expand these findings, and to further dissect the importance of these SNPs in the development of primary lung cancer.

Keywords: HOTAIR; lncRNA; lung cancer; single nucleotide polymorphisms

Acta Pharmacologica Sinica (2018) 39:1797–1803; <https://doi.org/10.1038/s41401-018-0083-x>

INTRODUCTION

Lung cancer exhibits the highest global morbidity and mortality rates [1]. Lung cancer is classified histologically as small cell (SCLC) or nonsmall cell lung cancer (NSCLC). NSCLC comprises 80–85% of all lung cancer cases, including adenocarcinoma, squamous cell carcinoma, and more rarely, large cell lung cancer [2]. Lung adenocarcinoma generally occurs in female and nonsmoking patients, and these cases typically involve drivers such as mutations in the *EGFR*, *KRAS*, and translocation of *EML4-ALK* genes. Squamous cell carcinoma generally occurs in male and smoking patients. However the driver genes involved remain unclear [3–6]. Efficacious targeted therapy was achieved in lung adenocarcinoma via the specific targeting of driver genes, such as *EGFR* and *ALK*. However, similar success remains to be achieved in lung squamous cell carcinoma [7].

Long noncoding RNAs (lncRNAs) are RNA molecules that have a length greater than 200 bases that do not encode any protein. A total of 4–9% of mammalian gene sequences produce lncRNA transcripts that participate in various important regulatory processes and are closely associated with the onset and progression of various diseases. More than 150 gene sites increase the susceptibility to cancer [3], and most of these sites are located

beyond the protein coding sequence. Sites such as enhancers or promoters can significantly influence gene transcription and protein expression [8–10]. Therefore, it is important to identify genes relevant to cancer and investigate their interactions with protein-encoding genes to increase our understanding of cancer pathogenesis.

The most intensively studied lncRNA is HOX transcript antisense intergenic RNA (HOTAIR), which encompasses 2158 nucleotides on chromosome 12q13.12. HOTAIR is transcribed from the antisense strand of the HOXC gene cluster, and it is a modular scaffold for two histone modification complexes [11]. Its 5' structural domain combines with polycomb repressive complex 2 (PRC2), and its 3' structural domain may be a scaffold for histone demethylated complex (a lysine-specific histone demethylase 1 (LSD1)). HOTAIR mediates epigenetically silenced target genes by binding these two complexes to specific gene sites [12], and it is a cancer-related gene. HOTAIR is associated with genetic susceptibility to various tumors, including esophageal cancer [13], gastric cancer [14], neuroglioma [15], and breast cancer [16] in different populations. Numerous studies verified that the expression of HOTAIR lncRNA is upregulated in lung cancer and participates in invasion and migration [17]. HOTAIR is closely associated with all TNM stages of

¹Department of Lung Cancer Surgery, Tianjin Medical University General Hospital, Tianjin 300052, China and ²Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute; Tianjin Medical University General Hospital, Tianjin 300052, China

Correspondence: Hong-yu Liu (LiuHongyu123@hotmail.com) or Jun Chen (huntercj2004@yahoo.com)

These authors contributed equally: Cong Wang, Ying Li, Yong-wen Li

Received: 13 February 2018 Accepted: 19 June 2018

Published online: 28 August 2018

lung cancer, including lymph node metastasis and poor survival prognosis, which indicates its potential as a novel biological marker for the diagnosis and monitoring of cancer onset and progression [18]. For instance, a single nucleotide polymorphism (SNP) located in intron 2 of the HOTAIR gene [cytosine to thymine (C > T) (rs920778)] was identified as a novel intronic HOTAIR enhancer that exhibits a specific genotypic effect on HOTAIR expression. Molecular epidemiological studies investigated the association between the HOTAIR rs920778 polymorphism and the risk of cancer, including esophageal squamous cell carcinoma (ESCC), breast cancer (BC), and gastric cancer (GC). However, the influence of HOTAIR SNPs on lung cancer has not yet been investigated.

To assess the influence of HOTAIR SNPs on genetic susceptibility to lung cancer, four SNP sites were selected: rs920778 C > T, rs12826786 C > T, rs4759314 A > G, and rs1899663 G > T. Compared with the CC genotype at the rs920778 site in the HOTAIR enhancer, the TT genotype increases the risk of esophageal squamous cell carcinoma [13]. The T allele at the rs12826786 site of the HOTAIR promoter increases the risk of gastric cardia adenocarcinoma. Genetic polymorphism rs4759314 from the AA genotype to the GG genotype is significantly associated with susceptibility to gastric cancer [19]. The TT genotype at the rs1899663 site increases the risk of breast cancer compared with the GG homozygotes in the Chinese population [20]. Therefore, these four sites are closely associated with the onset, progression, and poor prognosis of various cancers [21–24]. The present study examined whether these four sites were associated with genetic susceptibility to lung cancer using a database analysis of the relationship between HOTAIR expression levels and different factors in lung cancer patients.

MATERIALS AND METHODS

Patients and healthy controls

Study participants consisted of 262 patients who were diagnosed with primary lung cancer and underwent surgical resection between 2006 and 2010 at the Department of Lung Cancer Surgery, Tianjin Medical University General Hospital. The present study also included 451 healthy control individuals without a history of cancer who were enrolled from the Physical Examination Center of Tianjin Medical University General Hospital between 2006 and 2010. All enrolled subjects were from different areas in China. Written informed consent was obtained from all participants, and the Institutional Ethics Committee of Tianjin Medical University General Hospital approved the study. Inclusion criteria were the diagnosis of primary lung cancer (TNM staging, AJCC seventh) and the availability of outcome and follow-up data. The following information was collected from the patients' medical records: age, gender, clinical stage, pathological diagnosis, differentiation, lymph node status, metastasis, smoking status, and overall survival time. Survival was calculated from the day of resection until 1 April 2011.

Samples from the Cancer Genome Atlas (TCGA)

The TCGA database (<https://cancergenome.nih.gov/>) was searched, and 512 adenocarcinoma and 500 squamous cell carcinoma patients were selected. The clinical data were collected and arranged from building community resilience (BCR) data. All gene data were sourced from RNAseq expression data and were quoted after standardization through FPKM. Data in the TCGA database were updated in real time until 22 October 2017.

DNA extraction

Blood samples were collected in EDTA-coated tubes to prevent coagulation. Genomic DNA was isolated from peripheral blood lymphocytes of all participants using a DNA extraction kit (Qiagen).

Table 1. General information of all patients with NSCLC

Variable	Controls (n = 451)	Cases (n = 262)
Age, years		
≤57	273 (60.5%)	99 (37.8%)
>57	178 (39.5%)	163 (62.2%)
Gender		
Female	171 (37.9%)	74 (28.2%)
Male	280 (62.1%)	188 (71.8%)
Cigarette smoking		
Never	290 (64.3%)	85 (32.4%)
Ever	159 (35.3)	169 (64.5%)
Missing	2 (0.4%)	8 (3%)
Histology		
AC		99 (37.8%)
SCC		118 (44.7%)
LCLC		14 (5.3%)
SCLC		22 (8.4%)
Ade-squa		7 (2.7%)
Others		2 (0.8%)

AC adenocarcinoma, SCC squamous cell carcinoma, LCLC large cell lung cancer, Ade-Squa Adenosquamous carcinoma

Genotyping

Gene polymorphism analysis was performed using the TaqMan™ genotyping technique and a predesigned TaqMan™ SNP genotyping probe (Applied Biosystems, Foster City, CA, USA) or PCR primer, probe, and TaqMan™ master mix (Applied Biosystems). All genotyping was completed in 384-well plates (7900 HT Fast Real-Time PCR System) that included a reaction system (TaqMan™ genotyping master mix; 2 units, 2.5 μL), TaqMan™ SNP genotyping assay reagent (40 units, 0.125 μL), and template DNA (2.5 μL). A no-template (NTC) and positive control were included for each test. Plates were pre-read, and real-time PCR was performed under the following conditions: 50 °C for 2 min, 95 °C for 10 min, followed by 50 cycles at 95 °C for 15 s and 60 °C for 1 min. Each plate was re-read, and the results were analyzed using SDS 2.2 software.

Statistical analysis

Data were analyzed using SPSS 16.0 software for Windows. The intergroup mean was compared using a *t*-test, and *P* < 0.05 indicated statistical significance. The independence of risk factors was analyzed using logistic regression, and the risk rate is expressed as an odds ratio (OR) and 95% confidence interval (95% CI), with *P* < 0.05 indicating statistical significance. The observed and expected values were compared via calculation of the equilibrium frequency of the genotype, which underwent a χ^2 test. The analysis of linkage disequilibrium (LD) between different genetic polymorphic loci was performed using SHEsis online software at <http://analysis.bio-x.cn/myAnalysis.php>. LD analysis of the SNPs was performed to determine their nonrandom association in our population. The LD pattern between SNPs was measured using the correlation coefficient, r^2 , where $r^2 \geq 0.5$ was considered moderate to strong.

RESULTS

Research cohort

The present study cohort included 188 male and 74 female primary lung cancer patients aged 32–84 years old (median 62 years old) (Table 1). A total of 118 of these 262 patients had squamous cell carcinoma (SQCC), 99 patients had adenocarcinoma (ADC), 22

Table 2. Relevance analysis between gene polymorphisms of HOTAIR signal channel and the risk of lung cancer

SNP	Controls (n = 451)	Cases (n = 262)	OR(95% CI)	P value
rs920778				
CC	233 (51.7%)	110 (42.0%)	1	
C/T	192 (42.6%)	132 (50.4%)	1.467 (1.069–2.014)	0.018 (<0.05)
TT	26 (5.7%)	20 (7.6%)	1.629 (0.872–3.046)	0.126
C/T or TT	218 (48.3%)	152 (58%)	1.487 (1.094–2.020)	0.011 (<0.05)
rs12826786				
CC	355 (78.7%)	200 (79.0%)	1	
C/T	85 (18.8%)	49 (16.0%)	0.912 (0.619–1.343)	0.642
TT	11 (2.5)	13 (5.0%)	2.251 (0.969–5.227)	0.059
C/T or TT	96 (21.3%)	62 (21.0%)	1.042 (0.727–1.493)	0.822
rs4359714				
AA	412 (91.4%)	238 (90.8%)	1	
A/G	37 (8.2%)	21 (8.0%)	0.766 (0.446–1.315)	0.334
GG	2 (0.4%)	3 (1.1%)	2.571 (0.418–15.170)	0.314
A/G or GG	39 (8.6%)	24 (9.1%)	1.009 (0.592–1.720)	0.973
rs1899663				
GG	254 (56.3%)	178 (67.9%)	1	
G/T	189 (41.9%)	66 (25.2%)	0.771 (0.560–1.061)	0.111
TT	8 (1.8%)	18 (6.9%)	3.220 (1.272–8.150)	0.014 (<0.05)
G/T or TT	197 (43.7%)	84 (32.1%)	0.859 (0.630–1.170)	0.335

OR odds ratio, CI confidence interval. Data were calculated by logistic regression with adjustment for age, gender, smoking history, and pathological type

patients had small cell lung cancer (SCLC), 14 patients had large cell lung cancer (LCLC), seven patients had adenosquamous carcinoma, and two patients had other pathological types. A total of 169 patients had a smoking history, and 85 patients did not smoke. Table 1 presents detailed information of the lung cancer patient and healthy control groups.

Association between HOTAIR SNPs and lung cancer risk

The distribution frequency and corresponding OR value of the four SNP polymorphic sites in the HOTAIR gene in healthy and lung cancer patient groups are shown in Table 2. The OR value was calibrated according to gender, age, smoking history, and pathological type. Among the four SNP sites, rs920778 and rs1899663 were highly associated with lung cancer. In the healthy control and lung cancer groups, the distribution frequency varied with genotype at the rs920778 (C > T) site, and the distribution frequency of the C/T (C/T + TT) genotype in the lung cancer group was higher than that in the healthy control group (OR = 1.467, 95% CI: 1.069–2.014, $P < 0.05$). In addition, the distribution frequency of the TT genotype at the rs1899663 (G > T) site was significantly different from that of the GG genotype (OR = 3.220, 95% CI: 1.272–8.050, $P < 0.05$). In both groups, the distribution frequency of SNPs was not significantly different between the rs4759314 (A > G) or rs12826786 (C > T) sites.

The distribution frequency of the C > T genotype at the rs920778 site of HOTAIR is higher in male and smoking patients with squamous cell carcinoma

A stratified analysis was performed for each genotype at each HOTAIR site according to smoking history, gender, and pathological type. Subjects were divided into smoking and nonsmoking groups. At the rs920778 site, the OR in the smoking group varied with genotype. The C/T (1.822, 95% CI: 1.151–2.884, $P < 0.05$) and C/T + TT (1.864, 95% CI: 1.193–2.912, $P < 0.05$) sites were risk factors for primary lung cancer. In the nonsmoking group, the rs920778 site was not statistically significant.

In addition, subjects were divided into male and female groups. At the rs920778 site, the OR value in the male group varied with genotype. The C/T (1.865, 95% CI: 1.266–2.748, $P < 0.05$) and C/T + TT (1.760, 95% CI: 1.207–2.566, $P < 0.05$) genotypes increased genetic susceptibility to primary lung cancer. However, in females, the rs920778 site was not statistically significant. Moreover, at the rs1899663 site, the OR value of the TT genotype in males, but not in females, was statistically significant (3.304, 95% CI: 1.027–8.907, $P < 0.05$).

According to the pathological type of lung cancer, the subjects were divided into groups: SCLC, adenocarcinoma, and squamous cell carcinoma. At the rs920778 site, the OR of the C/T genotype in the SCLC group was 0.334 (95% CI: 0.120–0.927, $P < 0.05$). However, it could not be concluded that the C/T genotype played a protective role in primary lung cancer because only 22 patients with SCLC were recruited in the present study. In the squamous cell carcinoma group, the ORs of the C/T and C/T + TT genotype were 2.042 (95% CI: 1.327–3.143, $P < 0.05$) and 2.007 (95% CI: 1.317–3.060, $P < 0.05$), respectively, which indicates statistically significant risk factors for lung cancer. In the adenocarcinoma group, the results were not statistically significant (Tables 3 and 4).

Haplotype association analysis of SNPs in HOTAIR

Gene LD and haplotype analyses examined the association between HOTAIR gene polymorphisms and lung cancer risk. Genotype data at the rs12826786 site did not statistically conform to the Hardy–Weinberg equilibrium (HWE). Therefore, the three remaining gene sites were analyzed (rs920778, rs1899663, and rs4759314). At the rs920778 and rs1899663 sites, there was a high LD ($D' = 0.86$, $r^2 = 0.52$, Fig. 1). Eight possible haplotypes were investigated in the case-control group. In both the healthy control and lung cancer groups, $C_{(rs920778)}G_{(rs1899663)}A_{(rs4759314)}$ was the most common haplotype (68.2% and 61.4%, respectively), which reduced the risk of lung cancer (OR: 0.727, 95% CI: 0.576–0.918, $P = 0.007$). The risk of lung cancer was significantly decreased in the population with

Table 3. Stratified analyses of rs920778 for association with lung cancer risk

Variable	Controls	Cases	OR ^a (95% CI)	P value
Gender				
Male				
CC ^b	143 (31.7%)	69 (26.3%)	1	
C/T	120 (26.6%)	107 (40.8%)	1.865 (1.266–2.748)	*0.002
TT	17 (3.8%)	11 (4.2%)	1.341 (0.596–1.802)	3.017
C/T or TT	137 (30.4%)	118(45.4%)	1.760 (1.207–2.566)	*0.003
Female				
CC	90 (19.9%)	43 (16.4%)	1	
C/T	72 (16.0%)	23 (8.8%)	0.669 (0.369–1.211)	0.184
TT	9 (2.0%)	9 (3.4%)	2.093 (0.766–5.649)	0.145
C/T or TT	81 (18.0%)	32 (12.2%)	0.827 (0.478–1.429)	0.496
Smoking History				
Ever				
CC	76 (16.8%)	56 (21.4%)	1	
C/T	73 (16.2%)	98 (37.4%)	1.822 (1.151–2.884)	*0.01
TT	10 (2.2%)	16 (6.1%)	1.204 (0.917–5.143)	0.078
C/T or TT	83 (18.4%)	114 (43.5%)	1.864 (1.193–2.912)	*0.006
Never				
CC	155 (34.4%)	52 (19.8%)	1	
C/T	120 (26.6%)	36 (13.7%)	0.855 (0.525–1.392)	0.529
TT	17 (3.8%)	4 (1.5%)	0.727 (0.233–2.269)	0.583
C/T or TT	137 (30.4%)	40 (15.2%)	0.840 (0.524–1.346)	0.469
History type				
SCLC				
CC	233 (51.7%)	16 (6.1%)	1	
C/T	192 (42.6%)	5 (1.9%)	0.334 (0.120–0.927)	*0.035
TT	26 (5.7%)	1 (0.4%)		
C/T or TT	218 (48.3%)	6 (2.3%)	1.931 (0.788–4.730)	0.150
SCC				
CC	233 (51.7%)	41 (15.6%)	1	
C/T	192 (42.6%)	69 (26.3%)	2.042 (1.327–3.143)	*0.001
TT	26 (5.7%)	8 (3.1%)	1.749 (0.741–4.129)	0.202
C/T or TT	218 (48.3%)	77 (29.4%)	2.007 (1.317–3.060)	*0.001
ACC				
CC	233 (51.7%)	53 (19.8%)	1	
C/T	192 (42.6%)	39 (14.9%)	0.910 (0.576–1.437)	0.686
TT	26 (5.7%)	7 (2.7%)	1.206 (0.497–2.929)	0.678
C/T or TT	218 (48.3%)	46 (17.6%)	0.945 (0.610–1.465)	0.802

OR odds ratio, CI confidence interval

^aData were calculated by logistic regression with adjustment for age, gender, smoking history, and pathological type, where appropriate

^bNumber of patients/control subjects with genotype

*P < 0.05

Table 4. Stratified analyses of rs1899663 for association with lung cancer risk

Variable	Controls	Cases	OR ^a (95% CI)	P value
Gender				
Male				
GG ^b	160 (35.5%)	115 (43.9%)	1	
G/T	115 (25.5%)	61 (23.3%)	0.732 (0.494–1.803)	0.118
TT	5 (1.1%)	11 (4.2%)	3.034 (1.027–8.907)	*0.045
G/T or TT	120 (26.6%)	72 (26.5%)	0.828 (0.567–1.207)	0.326
Female				
GG	96 (21.3%)	43 (16.4%)	1	
G/T	73 (16.2%)	29 (11.1%)	0.887 (0.506–1.554)	0.675
TT	2 (0.4%)	3 (1.1%)	3.349 (0.540–20.722)	0.194
G/T or TT	75 (20.6%)	32 (12.2%)	0.953 (0.551–1.648)	0.862
Smoking History				
Ever				
GG	86 (19.1%)	109 (41.6%)	1	
G/T	72 (16.0%)	57 (21.8%)	0.064 (0.415–1.024)	0.652
TT	3 (0.7%)	11 (4.2%)	3.020 (0.816–11.178)	0.098
G/T or TT	75 (16.7%)	68 (26.0%)	0.747 (0.482–1.156)	0.190
Never				
GG	170 (37.7%)	51 (19.5%)	1	
G/T	116 (25.7%)	31 (11.8%)	0.891 (0.179–1.165)	0.654
TT	4 (0.9%)	3 (1.1%)	2.500 (0.542–11.537)	0.240
G/T or TT	120 (26.6%)	34 (12.9%)	0.944 (0.577–1.546)	0.820
History type				
SCLC				
GG	254 (56.3%)	14 (5.3%)	1	
G/T	189 (41.9%)	7 (2.7%)	0.511 (0.196–1.330)	0.169
TT	8 (1.8%)	1 (0.4%)		
G/T or TT	197 (43.7%)	8 (3.1%)	0.492 (0.189–1.281)	0.147
SCC				
GG	254 (56.3%)	72 (27.5%)	1	
G/T	189 (41.9%)	40 (15.3%)	0.757 (0.492–1.163)	0.203
TT	8 (1.8%)	6 (2.3%)	3.048 (0.993–9.353)	0.051
G/T or TT	197 (43.7%)	46 (17.6%)	0.839 (0.554–1.269)	0.405
ACC				
GG	254 (56.3%)	63 (24.0%)	1	
G/T	189 (41.9%)	31 (11.8%)	0.670 (0.419–1.071)	0.095
TT	8 (1.8%)	5 (1.9%)	2.902 (0.892–9.448)	0.077
G/T or TT	197 (43.7%)	36 (13.7%)	0.750 (0.478–1.176)	0.210

OR odds ratio, CI confidence interval

^aData were calculated by logistic regression with adjustment for gender, age, smoking history, and pathological type, where appropriate

^bNumber of patients/control subjects with genotype

*P < 0.05

the T_(rs920778)T_(rs1899663)A_(rs4759314) haplotype (OR: 2.223, 95% CI: 1.517–3.258, P < 0.001).

Analyses of HOTAIR expression, clinical characteristics, and prognostic conditions in lung cancer patients using the TCGA database

We previously observed that the C/T (C/T + TT) genotype at the rs920778 site increased the susceptibility to lung cancer in smokers. This SNP site is located within the enhancer of the

HOTAIR gene, and allele variation at this site influences the expression level and function of the HOTAIR gene. As a result, in individuals with the T allele, the expression level of the HOTAIR gene was higher than that in individuals with the CC genotype. To further investigate the relationship between the expression conditions of HOTAIR and the clinical characteristics of lung cancer, the differences in HOTAIR expression levels were analyzed between lung cancer and adjacent lung tissues from the TCGA

database. The expression levels of HOTAIR in squamous cell carcinoma and adenocarcinoma patients were significantly higher than that in the paired tissues adjacent to the cancer ($P < 0.001$, Fig. 2a) and highly correlated with the pathological type of lung cancer (Table 5, Fig. 2b). HOTAIR expression levels were significantly higher in squamous cell carcinoma patients than in adenocarcinoma patients. HOTAIR expression levels were also closely associated with smoking history (Table 5, Fig. 2d) and were

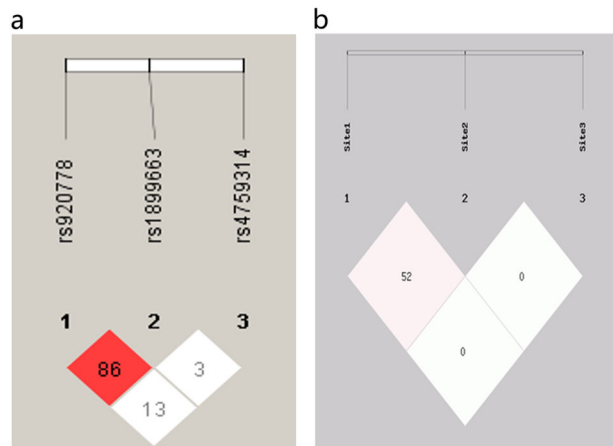


Fig. 1 Conditions of linkage disequilibrium for the three gene sites, rs920778, rs1899663, and rs4759314, analyzed using SHESIS software. **a** LD plot of the SNPs in HOTAIR; the significant SNPs are indicated by red boxes. **b** Numbers in squares indicate 100-fold r^2 values for each pair of SNPs

Variable	Total	Expression of HOTAIR ^a High	<i>P</i> value ^b Low	
Histological diagnosis				
AC	512	213	299	* <0.001
SCC	500	294	206	
Gender				
Male	605	313	292	0.223
Female	406	194	212	
Age, years				
<55	418	209	209	0.847
>55	566	287	279	
Smoking history				
Ever	827	429	398	*0.018
Never	185	78	107	
Clinical stage				
I–II	800	407	393	0.429
III–IV	200	95	105	
Metastasis				
No	754	386	368	0.243
Yes	250	117	133	

^aMedian HOTAIR level was used as the cut-off.
^b*p* values of Fisher's exact test are shown, * $P < 0.05$.

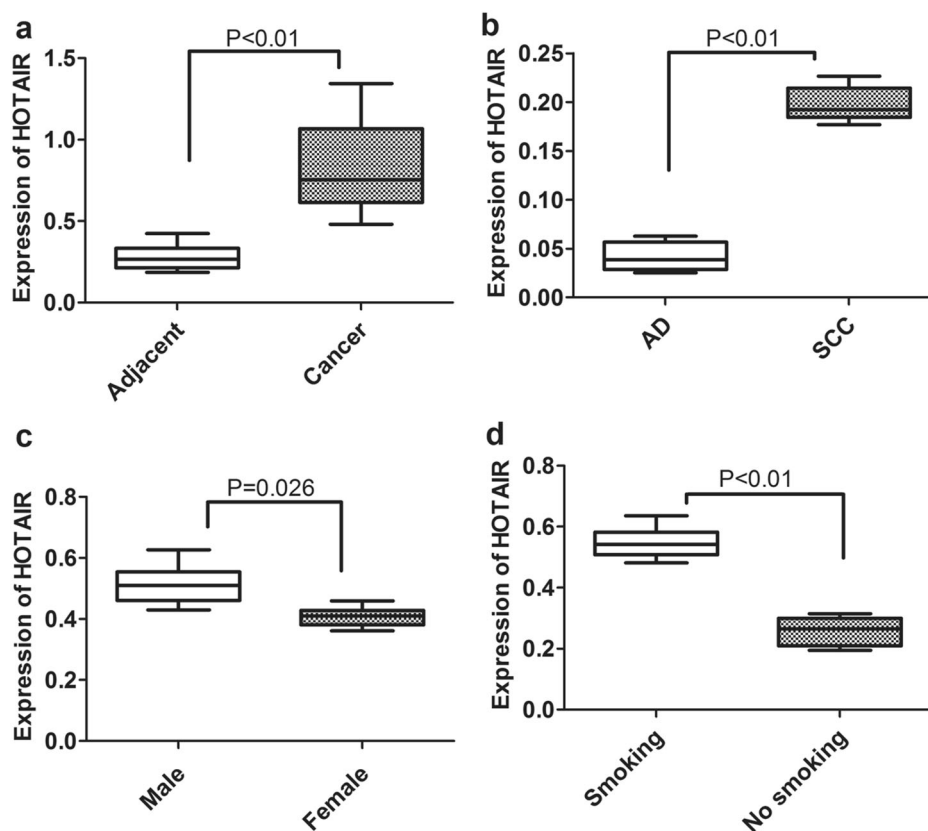


Fig. 2 Analysis of HOTAIR expression in different clinical characteristics. **a** Comparison between paired lung cancer and normal tissue. **b** Comparison between adenocarcinoma and squamous cell carcinoma patients. **c** Comparison between male and female patients. **d** Comparison between smoking and nonsmoking patients. AD adenocarcinoma, SCC squamous cell carcinoma

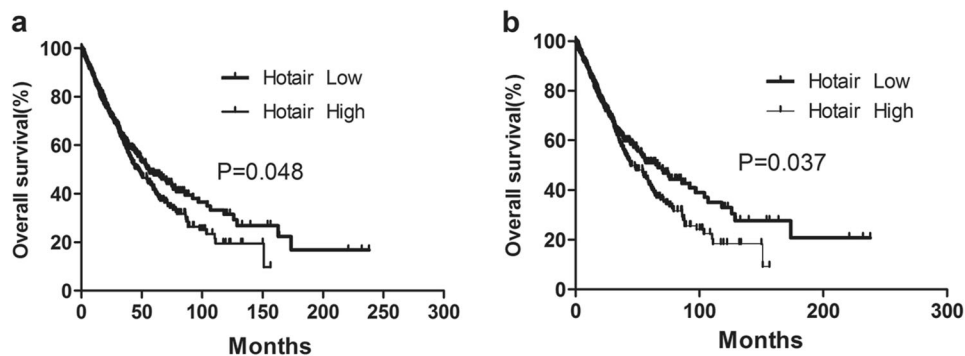


Fig. 3 Kaplan-Meier survival curves for lung cancer patients according to HOTAIR expression. **a** All cases. **b** Smoking cases

significantly increased in smoking lung cancer patients. The association between HOTAIR expression level and the clinicopathological characteristics (or survival prognostic conditions) of lung cancer patients was evaluated. The median HOTAIR expression level in 1012 lung cancer patients was assessed as high or low. The survival prognostic conditions in patients with low HOTAIR expression were significantly superior to those in patients with a high expression ($P = 0.048$, Fig. 3a) and superior to those in patients with high expression and smoking status ($P = 0.037$, Fig. 3b).

DISCUSSION

The morbidity rate of lung cancer in China is increasing; therefore, it is of great importance to elucidate the genetic factors associated with susceptibility to lung cancer. LncRNAs recently attracted research attention. Most functions of lncRNAs are not clear, but several intracellular roles, including a function in tumor progression, were identified. HOTAIR is a functional lncRNA that is derived from the HOXC gene cluster, and it is closely linked to the progression of lung cancer [17, 18, 25]. However, genetic variations in HOTAIR may regulate individual susceptibility to cancer and could influence HOTAIR expression and function [26–29]. Therefore, the present molecular epidemiology observational study examined whether the functional sites of HOTAIR were associated with genetic susceptibility to lung cancer. To the best of our knowledge, the present study is the first epidemiological study to examine the relationship between HOTAIR and susceptibility to lung cancer.

Many studies examined the influence of gene SNPs on susceptibility to lung cancer. For example, SNPs in miR-219-1 (rs213210, rs421446, and rs107822) significantly influence the susceptibility and prognosis of NSCLC [30]. In addition, compared with the AA genotype of ataxia telangiectasia-mutated (ATM), the TT genotype significantly increases the risk of lung cancer [31]. The present study examined whether SNPs in HOTAIR increased the susceptibility to lung cancer. Four sites were selected for statistical analyses: rs920778 (C > T), rs12826786 (C > T), rs4759314 (A > G), and rs1899663 (G > T). As expected, in the present hospital-based case-control study, the C/T and C/T + TT genotypes at the rs920778 site were significant risk factors for primary lung cancer compared with the CC genotype. The TT genotype at the rs1899663 site increased the susceptibility to lung cancer compared with the GG genotype. However, the actual increase may have been underestimated because of the small sample size. As shown by the stratified analysis of further factors (smoking, gender, and pathological type), the C/T and C/T + TT genotypes at the rs920778 site significantly increased the susceptibility to lung cancer in males but not in females. The C/T and C/T + TT genotypes at the rs920778 site also increased the risk of lung cancer in smokers. Therefore, as we expected, smoking-related

factors influenced HOTAIR expression via the induction of gene mutations within the gene itself. Cigarette smoke extract induces the inflammatory cell factor, IL-6, which activates the transcription factor STAT3 and directly interacts and increases the expression level of HOTAIR [32]. In the present study, the relationship between HOTAIR expression levels and smoking history in squamous cell carcinoma and adenocarcinoma patients was also analyzed in the TCGA database. In lung cancer patients, HOTAIR expression levels were strongly associated with smoking history ($P < 0.001$), and it was significantly higher in smoking patients than in nonsmoking patients. The results of the present study fully support the conclusions. However, the susceptibility factors for primary lung cancer were inferred from analyses of squamous cell carcinoma and adenocarcinoma data of NSCLC patients only. Therefore, all pathological types of primary lung cancer should be analyzed in future studies.

It has been shown that compared with the CC genotype at the rs920778 site, the TT genotype at the rs920778 site significantly increases the risk of esophageal squamous cell carcinoma, and HOTAIR expression level is increased in carriers of the T allele at the rs920778 site within the HOTAIR enhancer [13]. Another study reported that the expression level of HOTAIR was increased significantly in lung cancer patients, which was associated with the malignant phenotype and prognosis of lung cancer [33]. The C/T and C/T + TT genotypes at the rs920778 site increased the risk of squamous cell carcinoma, which indicates the possible association of HOTAIR SNPs with the pathological type of lung cancer [34]. Further analysis of lung cancer patients identified in the TCGA database revealed that the expression level of HOTAIR in squamous cell carcinoma patients was significantly higher than that in adenocarcinoma patients ($P = 0.018$), both of which were higher than that in the healthy population. These observations fully support our hypothesis that SNPs in the rs920778 site within HOTAIR significantly increase susceptibility to squamous cell carcinoma by increasing the expression level of HOTAIR. However, one limitation of our study is that HOTAIR expression levels were not evaluated in the lung cancer or control groups. Therefore, we cannot further verify whether the SNPs in HOTAIR influenced susceptibility to lung cancer by increasing HOTAIR expression levels. This analysis is planned in future studies.

In summary, gene polymorphisms in rs920778 and rs1899663 within HOTAIR significantly increase susceptibility to lung cancer. These two sites exhibit LD. In the population with a $T_{(rs920778)}C_{(rs1899663)}A_{(rs4759314)}$ haplotype, the risk of lung cancer is significantly increased. In addition, among NSCLC patients, the expression level of HOTAIR in squamous cell carcinoma is significantly higher than that in adenocarcinoma, and smoking influences HOTAIR expression levels. HOTAIR may be a novel predictor of the biological behavior of tumors and a novel target for the treatment of lung cancer.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Natural Science Foundation of China (to Jun Chen, No. 81773207; to Hong-yu Liu, No. 81372306), and the Tianjin key project of the Natural Science Foundation (to Jun Chen, 16JCZDJC34200), Special support program for High Tech Leader and Team of Tianjin (to Jun Chen, TJTZJHGCCXCXYTD-2-6), Tianjin Natural Science Foundation (to Hong-yu Liu, 13JCYBJC22600, 16PTSJYC00160) and the China Scholarship Council to Hong-yu Liu (201706940009).

AUTHOR CONTRIBUTIONS

CW, H-Y L, and JC wrote the manuscript and analyzed all of the data. CW, YL, Y-WL, H-BZ, HG, YY, and W-TL performed the experiments and collected the patients' information and data. H-YL and JC revised the manuscript.

ADDITIONAL INFORMATION

Competing interests: The authors declare no competing interests.

REFERENCES

1. Atmane A, Hammi S, Regragui A, Raoufi M, Marc K, Soualhi M, et al. Initial manifestation and atypical site for metastatic synovial sarcoma in an immunocompetent adult patient: about a case and literature review. *Pan Afr Med J.* 2017;28:103.
2. Anzai M, Morikawa M, Okuno T, Umeda Y, Demura Y, Sonoda T, et al. Efficacy and safety of nanoparticle albumin-bound paclitaxel monotherapy as second-line therapy of cytotoxic anticancer drugs in patients with advanced non-small cell lung cancer. *Medicine (Baltimore).* 2017;96:e9320.
3. Chung CC, Chanock SJ. Current status of genome-wide association studies in cancer. *Hum Genet.* 2011;130:59–78.
4. Liu CY, Stücker I, Chen C, Goodman G, McHugh MK, D'Amelio AM, et al. Genome-wide gene-asbestos exposure interaction association study identifies a common susceptibility variant on 22q13.31 associated with lung cancer risk. *Cancer Epidemiol Biomark Prev.* 2015;24:1564–73.
5. Hu WW, Tang CH, Sun Y, Lu TT, Jiang P, Wu YM, et al. Correlation between resistin gene polymorphism and clinical aspects of lung cancer. *Medicine (Baltimore).* 2017;96:e9485.
6. Park S, Ha S, Lee SH, Paeng JC, Keam B, Kim TM, et al. Intratumoral heterogeneity characterized by pretreatment PET in non-small cell lung cancer patients predicts progression-free survival on EGFR tyrosine kinase inhibitor. *PLOS ONE.* 2018;13:e0189766.
7. Liu S, Chen X, Chen R, Wang J, Zhu G, Jiang J, et al. Diagnostic role of Wnt pathway gene promoter methylation in non small cell lung cancer. *Oncotarget.* 2017;8:36354–67.
8. Sur I, Tuupanen S, Whittington T, Aaltonen LA, Taipale J. Lessons from functional analysis of genome-wide association studies. *Cancer Res.* 2013;73:4180–4.
9. Bond GL, Hu W, Bond EE, Robins H, Lutzker SG, Arva NC, et al. A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell.* 2004;119:591–602.
10. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, et al. TERT promoter mutations in familial and sporadic melanoma. *Science.* 2013;339:959–61.
11. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* 2007;129:1311–23.
12. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, et al. Long non-coding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* 2011;71:6320–6.
13. Zhang X, Zhou L, Fu G, Sun F, Shi J, Wei J, et al. The identification of an ESCC susceptibility SNP rs920778 that regulates the expression of lncRNA HOTAIR via a novel intronic enhancer. *Carcinogenesis.* 2014;35:2062–7.
14. Guo W, Dong Z, Bai Y, Guo Y, Shen S, Kuang G, et al. Associations between polymorphisms of HOTAIR and risk of gastric cardia adenocarcinoma in a population of north China. *Tumour Biol.* 2015;36:2845–54.
15. Xavier-Magalhães A, Oliveira AI, de Castro JV, Pojo M, Gonçalves CS, Lourenço T, et al. Effects of the functional HOTAIR rs920778 and rs12826786 genetic variants in glioma susceptibility and patient prognosis. *J Neurooncol.* 2017;132:27–34.
16. Bayram S, Sümbül AT, Batmacı CY, Genç A. Effect of HOTAIR rs920778 polymorphism on breast cancer susceptibility and clinicopathologic features in a Turkish population. *Tumour Biol.* 2015;36:3863–70.
17. Geng YJ, Xie SL, Li Q, Ma J, Wang GY. Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. *J Int Med Res.* 2011;39:2119–28.
18. Li N, Wang Y, Liu X, Luo P, Jing W, Zhu M, et al. Identification of circulating long noncoding RNA HOTAIR as a novel biomarker for diagnosis and monitoring of non-small cell lung cancer. *Technol Cancer Res Treat.* 2017;16:1060–66.
19. Lin Y, Guo W, Li N, Fu F, Lin S, Wang C. Polymorphisms of long non-coding RNA HOTAIR with breast cancer susceptibility and clinical outcomes for a southeast Chinese Han population. *Oncotarget.* 2018;9:3677–89.
20. Qi Q, Wang J, Huang B, Chen A, Li G, Li X, et al. Association of HOTAIR polymorphisms rs4759314 and rs920778 with cancer susceptibility on the basis of ethnicity and cancer type. *Oncotarget.* 2016;7:38775–84.
21. Miao Z, Ding J, Chen B, Yang Y, Chen Y. HOTAIR overexpression correlated with worse survival in patients with solid tumors. *Minerva Med.* 2016;107:392–400.
22. Lee M, Kim HJ, Kim SW, Park SA, Chun KH, Cho NH, et al. The long non-coding RNA HOTAIR increases tumour growth and invasion in cervical cancer by targeting the Notch pathway. *Oncotarget.* 2016;7:44558–71.
23. Sun W, Yang Y, Xu C, Xie Y, Guo J. Roles of long noncoding RNAs in gastric cancer and their clinical applications. *J Cancer Res Clin Oncol.* 2016;142:2231–7.
24. Shi YY, He L. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. *Cell Res.* 2005;15:97–8.
25. Herrera-Solorio AM, Armas-López L, Arrieta O, Zúñiga J, Piña-Sánchez P, Ávila-Moreno F. Histone code and long non-coding RNAs (lncRNAs) aberrations in lung cancer: implications in the therapy response. *Clin Epigenetics.* 2017;9:98.
26. Pan W, Liu L, Wei J, Ge Y, Zhang J, Chen H, et al. A functional lncRNA HOTAIR genetic variant contributes to gastric cancer susceptibility. *Mol Carcinog.* 2016;55:90–6.
27. Tian T, Li C, Xiao J, Shen Y, Lu Y, Jiang L, et al. Quantitative assessment of the polymorphisms in the HOTAIR lncRNA and cancer risk: a meta-analysis of 8 case-control studies. *PLOS ONE.* 2016;11:e0152296.
28. Bayram S, Ülger Y, Sümbül AT, Kaya BY, Rencüzoğulları A, Genç A, et al. A functional HOTAIR rs920778 polymorphism does not contribute to gastric cancer in a Turkish population: a case-control study. *Fam Cancer.* 2015;14:561–7.
29. Xue Y, Gu D, Ma G, Zhu L, Hua Q, Chu H, et al. Genetic variants in lncRNA HOTAIR are associated with risk of colorectal cancer. *Mutagenesis.* 2015;30:303–10.
30. Zheng C, Li X, Xia L, Fang X, Quan X, Yin Z, et al. Polymorphisms of pri-miR-219-1 are associated with the susceptibility and prognosis of non-small cell lung cancer in a Northeast Chinese population. *Oncotarget.* 2017;8:56533–41.
31. Myneni AA, Chang SC, Niu R, Liu L, Zhao B, Shi J, et al. Ataxia telangiectasia-mutated (ATM) polymorphisms and risk of lung cancer in a Chinese population. *Front Public Health.* 2017;5:102.
32. Liu Y, Luo F, Xu Y, Wang B, Zhao Y, Xu W, et al. Epithelial-mesenchymal transition and cancer stem cells, mediated by a long non-coding RNA, HOTAIR, are involved in cell malignant transformation induced by cigarette smoke extract. *Toxicol Appl Pharmacol.* 2015;282:9–19.
33. Zhuang Y, Wang X, Nguyen HT, Zhuo Y, Cui X, Fewell C, et al. Induction of long intergenic non-coding RNA HOTAIR in lung cancer cells by type I collagen. *J Hematol Oncol.* 2013;6:35.
34. Nakagawa T, Endo H, Yokoyama M, Abe J, Tamai K, Tanaka N, et al. Large non-coding RNA HOTAIR enhances aggressive biological behavior and is associated with short disease-free survival in human non-small cell lung cancer. *Biochem Biophys Res Commun.* 2013;436:319–24.