

Impact of COL6A4P2 gene polymorphisms on the risk of lung cancer in a Chinese Han population

Ying Duan

Xi'an Jiaotong University Medical College First Affiliated Hospital

Gaowen Liu

Xianyang Central Hospital

Fanglin Niu

Northwest University

Jing Li

Northwest University

Mengdan Yan

Northwest University

Tianbo Jin

Northwest University

Mingwei Chen (✉ mingwei_chen@163.com)

Xi'an Jiaotong University

Research article

Keywords: Lung cancer, COL6A4P2 gene, Polymorphisms, Susceptibility

Posted Date: January 20th, 2020

DOI: <https://doi.org/10.21203/rs.2.21226/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

The aim of this study was to investigate the effects of COL6A4P2 polymorphisms on lung cancer (LC) in Chinese Han population.

Methods

To examine whether variants of COL6A4P2 contribute to LC, five single nucleotide polymorphisms (SNPs) of COL6A4P2 were genotyped by Agena MassARRAY in 510 LC patients and 495 controls. Odds ratio (OR) and 95% confidence intervals (CIs) were calculated by logistic regression.

Results

We found that COL6A4P2 rs34445363 significantly increased the risk of LC in the alleles model (OR = 1.26, 95%CI: 1.01 - 1.58, $p = 0.038$). And rs34445363 also increased the LC risk under the log-additive model (OR = 1.26, 95%CI: 1.01 - 1.58, $p = 0.041$) with the multigene model analysis. Further stratification analysis showed that rs34445363 increased the LC risk under the log-additive model (OR = 1.42, 95%CI: 1.03 - 1.95, $p = 0.033$) in people aged ≤ 61 ; and rs61733464 was associated with a decreased LC risk in the log-additive model (OR = 0.72, 95%CI: 0.52 - 0.99, $p = 0.048$) in people aged ≤ 61 . We also found that the mutations of rs34445363 and rs77941834 were associated with increased LC risk in the codominant model (rs34445363, GA vs. GG, OR = 1.73, 95%CI: 1.04 - 2.86, $p = 0.034$; rs77941834, TA vs. TT, OR = 1.88, 95%CI: 1.06 - 3.34, $p = 0.032$) in females.

Conclusions

This study provided an evidence for polymorphisms of COL6A4P2 gene associated to the development of LC, also a new insight into etiology of LC.

1. Introduction

Lung cancer (LC) is the malignant tumor with the fastest growth in morbidity and mortality and the greatest threat to people's health and life (1). According to the database of Global Cancer Observatory (<http://gco.iarc.fr/>) (2), there are 2,093,876 new cases of LC worldwide, accounting for 11.6% of all cancers; the number of people who died of LC is 1,761,007, accounting for 17.9% of all cancer deaths in 2018. Among them, the incidence and mortality of LC in female were 13.1% and 6.9%, respectively. LC has become the most malignant tumor with the highest incidence and mortality (3–5). In China, LC also has high incidence and mortality, and men's morbidity and mortality are more than twice than that of women (6). Most studies suggested that the occurrence of LC is related to environmental (smoke, occupational exposure, and air pollution) and genetic factors (7, 8), especially genetic factors play important role. Li et al. (9) revealed that LC susceptibility in Chinese Han population is related to HOTAIR gene mutation. Dimitrakopoulos et al. (10) believed that NF- κ B2 gene mutation is significantly associated with LC risk. However, the correlation between COL6A4P2 gene polymorphisms and LC susceptibility has not been reported.

COL6A4P2 (Collagen Type VI Alpha 4 Pseudogene 2), also named as COL6A4, located on Chr.3q22 in humans. COL6A4 gene expressed type VI collagen (COL6), which is an extracellular matrix protein that plays an important role in maintaining the integrity of lung tissue. Chiu et al. (11) showed by quantitative secretion cleavage that COL6 is a protein involved in tumor metastasis. Voiles et al. (12) demonstrated that the expression of COL6 protein in LC is up-regulated. Thus, we suspect that the COL6A4 gene may be associated with LC.

It is reported that COL6A4 transcribe as an unprocessed pseudogene due to the presence of multiple stop codons on the gene sequence (13). Many studies have been made to show that pseudogenes play an important role in the development of cancer. Cheng et al. (14) found that pseudogenes affect the occurrence and development of cancer by forming lncRNA-pseudogene-mRNA competitive triples. Lynn et al. (15) confirmed that the polymorphisms of the MYLK1 pseudogene is associated with an increased risk of colon cancer. Wei et al. (16) found that the pseudogene DUXAP10 promotes the invasiveness of LC. Therefore, we speculated that the COL6A4P2 gene may play a role in cancer development.

In this study, we first explored the association of COL6A4P2 gene and LC risk, revealing the relationship between COL6A4P2 gene polymorphism and LC susceptibility in Chinese Han population.

2. Materials And Methods

2.1 Study participants

Using a case-control design, 510 LC patients (mean age: 60.78 ± 9.96 years) and 495 controls (mean age: 61.94 ± 7.72 years) were enrolled. All patients were recruited from Shaanxi Provincial Cancer Hospital (Xi'an City, Shaanxi, China). Patient inclusion criteria: 1) patients with newly diagnosed LC, 2) histopathological LC diagnosed by an experienced pathologist, 3) no previous radiation therapy or chemical therapy, 4) no history of cancer and metastatic carcinoma. Patients with asthma, bronchitis, pneumonia, lung abscess, tuberculosis and other lung diseases, autoimmune diseases, trauma or other tumors were excluded from the study. After that, we investigated and collected clinical indicators of LC patients, including gender, age, histological classification, tumor stage, and the status of lymph node metastasis.

The controls were healthy volunteers from Shaanxi Provincial Cancer Hospital (Xi'an, Shaanxi, China) recruited during the same period. Inclusion criteria of control group included no medical or family history of cancer or any pulmonary disease. At the time of recruitment, each subject was personally interviewed by trained personnel using a structured questionnaire to obtain information regarding demographic characteristics.

2.2 Data collection

This study was approved by the ethics committee of the Shaanxi Provincial Cancer Hospital, and conformed to the ethical principles for medical research involving humans of the World Medical Association Declaration of Helsinki. All participants signed informed consent forms before participating in this study. Subsequently, a sample of approximately 5 mL of venous blood was obtained from each participant and collected into tubes containing ethylenediamine tetra-acetic acid for anticoagulation. Genomic DNA was extracted from peripheral blood samples using a Whole-Blood Genomic DNA Extraction Kit (GOLDMAG, Xi'an, China) according to the manufacturer's instructions. The purity and concentration of the DNA samples were evaluated using a NanoDrop 2000C system (Thermo Scientific, Waltham, MA, USA). Isolated DNA was stored at -80 °C until analysis.

2.3 SNP genotyping

Five candidate SNPs in the COL6A4P2 gene were selected with a minor allele frequency (MAF) >0.05 from global population in the 1,000 Genome Projects (<http://www.internationalgenome.org/>). Then we used HaploReg v4.1 (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>) to predict the possible functions of SNPs. The primers for amplification and single-base extension were designed using the Assay Design Suite, V2.0 (<https://agenacx.com/online-tools/>). Genotyping of the five SNPs was carried out on MassARRAY iPLEX (Agena Bioscience, San Diego, CA, USA) platform using matrix-assisted laser desorption ionization-time of flight mass spectrometry (17). Genotyping results were generated using Agena Bioscience TYPER software, version 4.0. Genotyping was carried out by laboratory personnel in a double-blinded fashion.

2.4 Analysis of COL6A4P2 and SNPs expression

Data regarding the expression of COL6A4P2 in LC was obtained from the UALCAN online database (<http://ualcan.path.uab.edu/analysis.html>), a web server providing customizable functions. Tumors and normal samples in the UALCAN database were derived from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) projects. Then predicted the effect of COL6A4P2 gene expression on LC prognosis use OncoLnc database (<http://www.oncolnc.org/>). We also predicted the expression of SNPs in the COL6A4P2 gene in normal lung tissues by GTEx database (<https://gtexportal.org/home/>).

2.5 Statistical analyses

An independent sample t-test was used to assess differences in population and clinical characteristics of study participants. Fisher's exact tests for HWE were performed by comparing the observed and expected genotype frequencies to calculate the genotype frequencies among the controls. Pearson's χ^2 test was used to compare the allelic and genotype frequencies of each SNP between LC patients and controls. Multiple genetic model analyses (codominant, dominant, recessive, and log-additive) were applied using PLINK software (<http://zzz.bwh.harvard.edu/plink/ld.shtml>) to assess the association between SNPs and LC risk. Furthermore, we calculated stratification factors using gender and age to adjust for possible confounders. Finally, we used Haploview software (version4.2) to construct haplotype and to estimate the pairwise linkage disequilibrium, the SHEsis software platform (<http://analysis.bio-x.cn/myAnalysis.php>) was used to estimate the correlation between haplotype and LC risk. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using logistic regression analyses adjusted for gender and age (18), with the wild-type allele used as a reference. Statistical analyses were performed using SPSS software (version 21.0, IBM Corporation, Armonk, NY, USA). All p-values of statistical tests were two-sided, and p < 0.05 was considered indicative of statistical significance.

3. Results

3.1 Characteristics of cases and controls

The basic clinical information of LC patients and controls were shown in Table 1. 510 patients presented with different distribution, according to age (age ≤ 61, 266 cases; age > 61, 244 cases), gender (male, 355 cases; female, 155 cases), pathological type (lung squamous cell carcinoma [LUSC], 120 patients; lung adenocarcinoma [LUAD], 188 patients), tumor stage (I-II, 84 cases; III-IV, 261 cases), and lymph node metastasis (LNM) status (positive, 215 cases; negative, 84 cases).

Table 1
The comparison of basic characteristics between cases and controls

Characteristics		Case (n = 510)	Control (n = 495)
Age	≤ 61	266	224
	> 61	244	271
	Mean ± SD	60.78 ± 9.96	61.94 ± 7.72
Gender	Male	355	346
	Female	155	149
Pathological type	LUSC	120	
	LUAD	188	
	Unknown	202	
Tumor stage	I-II	84	
	III-IV	261	
	Unknown		
LNM	Positive	215	
	Negative	84	
	Unknown		
LUSC = lung squamous cell carcinoma; LUAD = lung adenocarcinoma; LNM = lymph node metastasis			

Basic information and allele frequencies of the COL6A4P2 gene polymorphisms were presented in Table 2. The genotype distribution of all SNPs in control subjects met the HWE ($p > 0.05$). HaploReg function annotation results revealed that SNPs associated with LC risk were successfully predicted to have biological functions. The correlation between COL6A4P2 polymorphisms and LC risk under the allele model was shown in Table 2, the results showed that rs34445363 is associated with an increased LC risk (OR = 1.26, 95%CI: 1.01–1.58, $p = 0.038$), and there were no differences between other four SNPs (rs7625942, rs77941834, rs61733464, and rs11914893) in the COL6A4P2 gene and LC risk ($p > 0.05$).

Table 2
Basic Information about SNPs in COL6A4P2 and association with risk of lung cancer in allele model

Gene	SNP ID	Chr.	Alleles(A/B)	Frequency (MAF)		p-value for HWE	OR (95% CI)	p	Function
				Case	Control				
COL6A4P2	rs34445363	3q22.1	A/G	0.217	0.180	0.879	1.26 (1.01–1.58)	0.038	Selected eQTL hits
COL6A4P2	rs7625942	3q22.1	A/G	0.223	0.225	0.608	0.98 (0.80–1.21)	0.915	Motifs changed, Selected eQTL hits
COL6A4P2	rs77941834	3q22.1	A/T	0.122	0.097	0.798	1.29 (0.97–1.71)	0.086	Motifs changed, Selected eQTL hits
COL6A4P2	rs61733464	3q22.1	A/G	0.186	0.213	0.346	0.85 (0.68–1.06)	0.146	DNase, Motifs changed, Selected eQTL hits
COL6A4P2	rs11914893	3q22.1	A/C	0.108	0.115	0.825	0.93 (0.70–1.23)	0.620	Motifs changed, GRASP QTL hits
SNP = single nucleotide polymorphism; Chr. = chromosome; A/B = minor/major, MAF = minor allele frequency; HWE = Hardy Weinberg equilibrium.									
$p < 0.05$ indicates statistical significance.									
Bold values indicate a significant difference.									

3.2 Association between the COL6A4P2 gene and the risk of LC

Genetic models (codominant, dominant, recessive, and log-additive) and genotype frequencies were used to further identify any associations between the SNPs and the risk of LC. The results showed that rs34445363 in the COL6A4P2 gene significantly increased the LC risk under the log-additive model (adjusted by age and gender, OR = 1.26, 95%CI: 1.01–1.58, $p = 0.041$, Table 3), and no significant difference was found for the other SNPs between cases and controls (all $p > 0.05$).

Table 3

Distribution of genotypes of COL6A4P2 polymorphism depicting their association with lung cancer risk and its histological subtypes

SNP ID	Model	Genotype	Control	LC			LSCC			LUAD		
				Case	OR (95%CI)	p	Case	OR (95%CI)	p	Case	OR (95%CI)	p
rs34445363	Codominant	GG	329	313	1.00	72	1.00	112	1.00			
		GA	146	173	1.25(0.10–1.64)	0.102	43	1.27(0.82–1.96)	0.278	66	1.39(0.96–2.00)	0.082
		AA	15	24	1.63(0.84–3.17)	0.151	5	1.52(0.52–4.46)	0.442	10	1.88(0.81–4.36)	0.144
	Dominant	GG	329	313	1.00	72	1.00	112	1.00			
		GA/AA	161	197	1.29(0.99–1.67)	0.056	48	1.29(0.85–1.97)	0.229	76	1.43(1.01–2.04)	0.046
	Recessive	GG/GA	475	486	1.00	115	1.00	178	1.00			
		AA	15	24	1.51(0.78–2.92)	0.220	5	1.40(0.48–4.06)	0.533	10	1.68(0.73–3.87)	0.223
	Log-additive	–	–	–	1.26(1.01–1.58)	0.041	–	1.26(0.88–1.80)	0.212	–	1.38(1.02–1.86)	0.034
rs61733464	Codominant	GG	310	340	1.00	82	1.00	133	1.00			
		GA	158	150	0.86(0.66–1.13)	0.278	33	0.79(0.50–1.24)	0.299	46	0.65(0.44–0.96)	0.031
		AA	26	20	0.70(0.38–1.28)	0.246	5	0.75(0.27–2.07)	0.581	9	0.76(0.34–1.69)	0.504
	Dominant	GG	310	340	1.00	82	1.00	133	1.00			
		GA/AA	184	170	0.84(0.65–1.09)	0.181	38	0.78(0.51–1.21)	0.265	55	0.66(0.46–0.96)	0.031
	Recessive	GG/GA	468	490	1.00	115	1.00	179	1.00			
		AA	26	20	0.73(0.40–1.34)	0.310	5	0.81(0.30–2.21)	0.683	9	0.87(0.39–1.91)	0.724
	Log-additive	–	–	–	0.85(0.68–1.05)	0.139	–	0.82(0.57–1.18)	0.285	–	0.74(0.55–1.01)	0.059

SNP = single nucleotide polymorphism; LC = lung cancer; LUAD = lung adenocarcinoma; LSCC = lung squamous cell carcinoma; OR = odds ratio; 95%CI = 95% confidence interval.

p < 0.05 indicates statistical significance.

Bold values indicate a significant difference.

Furthermore, we identified by pathological analysis that rs34445363 locus variation significantly increased the risk of LUAD in the dominant model (adjusted by age and gender, GA/AA vs. GG, OR = 1.43, 95%CI: 1.01–2.04, p = 0.046) and log-additive model (adjusted by age and gender, OR = 1.38, 95%CI: 1.02–1.86, p = 0.034); However, mutations of rs61733464 in the COL6A4P2 gene have a lower incidence of LUAD with the GA genotype in the codominant model (adjusted by age and gender, GA vs. GG, OR = 0.65, 95%CI: 0.44–0.96, p = 0.031) and under the dominant model (adjusted by age and gender, GA/AA vs. GG, OR = 0.66, 95%CI: 0.46–0.96, p = 0.031).

3.3 Relationship between COL6A4P2 polymorphism and clinicopathological features

To evaluate association of COL6A4P2 SNPs on various clinic-pathological features, we segregated patients according to clinical stage (I–II vs. III–IV), and status of LNM (positive vs. negative). There was no significant correlation between LNM status and COL6A4P2 gene polymorphism variation (Supplementary table 1). However, for rs77941834 variant, the codominant model (adjusted by age and gender, TA vs. TT, OR = 0.52, 95%CI: 0.29–0.94, p = 0.030), dominant model (adjusted by age and gender, TA/AA vs. TT, OR = 0.49, 95%CI: 0.28–0.86, p = 0.013), and log-additive model (adjusted by age and gender, OR = 0.55, 95%CI: 0.34–0.87, p = 0.011) were significantly decrease the LC risk in patients with I–II as compared to patients with III–IV tumor stage (Table 4). No statistically significant association was observed for tumor staging and the other four SNPs (rs34445363, rs7625942, rs61733464 and rs11914893).

Table 4
Relationship between COL6A4P2 polymorphism and tumor staging of lung cancer

SNP ID	Model	Genotype	Control	Case	OR (95%CI)	p
rs77941834	Codominant	TT	57	210	1.00	
		TA	23	46	0.52 (0.29–0.94)	0.030
		AA	4	5	0.33 (0.09–1.31)	0.116
	Dominant	TT	57	210	1.00	
		TA/AA	27	51	0.49 (0.28–0.86)	0.013
	Recessive	TT/TA	80	256	1.00	
		AA	4	5	0.39 (0.10–1.49)	0.167
	Log-additive				0.55 (0.34–0.87)	0.011
	SNP = single nucleotide polymorphism; OR = odds ratio; 95%CI = 95% confidence interval.					
p < 0.05 indicates statistical significance.						
Bold values indicate a significant difference.						

3.4 Stratification analysis of age and gender

Multiple inheritance model analysis showed that age and gender were significantly affect the correlation between COL6A4P2 SNPs and LC risk. We found that rs34445363 was related to a higher incidence of LC in people aged ≤ 61 with the AA genotype in the codominant model (adjusted by gender, AA vs. GG, OR = 2.62, 95%CI: 1.00–6.85, p = 0.049) and in the log-additive model (adjusted by gender, OR = 1.42, 95%CI: 1.03–1.95, p = 0.033); rs61733464 was associated with a decreased LC risk under the dominant model (adjusted by gender, GA/AA vs. GG, OR = 0.68, 95%CI: 0.46–0.99; p = 0.048) and log-additive model (adjusted by gender, OR = 0.72, 95%CI: 0.52–0.99, p = 0.048) in people aged ≤ 61 years (Table 5).

Table 5

Distribution of COL6A4P2 polymorphisms in populations of different ages and genders and its relationship with risk of lung cancer

SNP ID	Model	Genotype	Age > 61				Age ≤ 61			
			Control	Case	OR (95%CI)	p	Control	Case	OR (95%CI)	p
rs34445363	Codominant	GG	179	152	1.00		150	161	1.00	
		GA	82	86	1.24 (0.85–1.81)	0.254	64	87	1.29 (0.87–1.92)	0.210
		AA	9	6	0.76 (0.26–2.19)	0.606	6	18	2.62 (1.00–6.85)	0.049
	Dominant	GG	179	152	1.00		150	161	1.00	
		GA/AA	91	92	1.20 (0.83–1.72)	0.340	70	105	1.41 (0.96–2.06)	0.079
	Recessive	GG/GA	261	238	1.00		214	248	1.00	
		AA	9	6	0.70 (0.24–2.02)	0.513	6	18	2.41 (0.93–6.24)	0.070
	Log-additive	–	–	–	1.11 (0.81–1.53)	0.524	–	–	1.42 (1.03–1.95)	0.033
	rs61733464	Codominant	GG	174	159	1.00		136	181	1.00
GA			81	74	0.98 (0.67–1.44)	0.923	77	76	0.70 (0.47–1.03)	0.073
AA			15	11	0.82 (0.36–1.86)	0.636	11	9	0.58 (0.23–1.46)	0.249
Dominant		GG	174	159	1.00		136	181	1.00	
		GA/AA	96	85	0.96 (0.66–1.38)	0.812	88	85	0.68 (0.46–0.99)	0.048
Recessive		GG/GA	255	233	1.00		213	257	1.00	
		AA	15	11	0.83 (0.37–1.86)	0.642	11	9	0.66 (0.26–1.63)	0.365
Log-additive		–	–	–	0.95 (0.70–1.28)	0.713	–	–	0.72 (0.52–0.99)	0.048
SNP ID		Model	Genotype	Male				Female		
	Control			Case	OR (95%CI)	p	Control	Case	OR (95%CI)	p
rs34445363	Codominant	GG	225	220	1.00		104	92	1.00	
		GA	110	118	1.10 (0.80–1.52)	0.547	36	55	1.73 (1.04–2.86)	0.034
		AA	11	17	1.47 (0.67–3.24)	0.334	4	7	1.98 (0.56–6.98)	0.289
	Dominant	GG	225	220	1.00		104	92	1.00	
		GA/AA	121	135	1.14 (0.84–1.55)	0.411	40	62	1.75 (1.08–2.85)	0.024
	Recessive	GG/GA	335	338	1.00		140	147	1.00	

SNP = single nucleotide polymorphism; OR = odds ratio; 95%CI = 95% confidence interval.

p < 0.05 indicates statistical significance.

Bold values indicate a significant difference.

SNP ID	Model	Genotype	Age > 61				Age ≤ 61			
			Control	Case	OR (95%CI)	p	Control	Case	OR (95%CI)	p
		AA	11	17	1.43 (0.65–3.11)	0.372	4	7	1.67 (0.18–5.82)	0.42
	Log-additive	–	–	–	1.15 (0.88–1.49)	0.314	–	–	1.60 (1.05–2.44)	0.028
rs77941834	Codominant	TT	279	284	1.00		124	112	1.00	
		TA	63	61	0.94 (0.64–1.39)	0.763	23	39	1.88 (1.06–3.34)	0.032
		AA	4	10	2.42 (0.75–7.84)	0.141	1	2	2.21 (0.20–24.76)	0.519
	Dominant	TT	279	284	1.00		124	112	1.00	
		TA/AA	67	71	1.03 (0.71–1.50)	0.878	24	41	1.89 (1.07–3.33)	0.027
	Recessive	TT/TA	342	345	1.00		147	151	1.00	
		AA	4	10	2.44 (0.76–7.91)	0.136	1	2	1.94 (0.17–21.66)	0.590
	Log-additive	–	–	–	1.11 (0.80–1.53)	0.547	–	–	1.81 (1.06–3.08)	0.030

SNP = single nucleotide polymorphism; OR = odds ratio; 95%CI = 95% confidence interval.

p < 0.05 indicates statistical significance.

Bold values indicate a significant difference.

Also, we found that gender significantly affects the association between SNPs of the COL6A4P2 gene and LC risk (Table 5). The mutation of COL6A4P2 rs34445363 in females could significantly increase the LC risk with the GA genotype under the codominant model (adjusted by age, GA vs. GG, OR = 1.73, 95%CI: 1.04–2.86, p = 0.034), dominant model (adjusted by age, GA/AA vs. GG, OR = 1.75, 95%CI: 1.08–2.85, p = 0.024) and log-additive model (adjusted by age, OR = 1.60, 95%CI: 1.05–2.44, p = 0.028); Women with rs77941834 mutation have a higher incidence of LC with the TA genotype under the codominant model (adjusted by age, TA vs. TT, OR = 1.88, 95%CI: 1.06–3.34, p = 0.032), in dominant model (adjusted by age, TA/AA vs. TT, OR = 1.89, 95%CI: 1.07–3.33, p = 0.027) and log-additive model (adjusted by age, OR = 1.81, 95%CI: 1.06–3.08, p = 0.030).

3.5 Association of COL6A4P2 haplotypes with the risk of LC

SNPs in the current study were in linkage disequilibrium for the study population (Fig. 1). Unfortunately, there was no statistically significant difference among any of the COL6A4P2 haplotype frequencies in cases and controls (Supplementary table 2).

3.6 Expression of COL6A4P2 and SNPs

Database analysis showed that compared with healthy subjects, expression of the COL6A4P2 gene was significantly higher in LUAD (p = 1.62 × 10⁻¹²), and expression of the COL6A4P2 gene was significantly higher in LUSC (p = 2.44 × 10⁻⁷, Fig. 2A and 2C). OncoLnc database analysis showed that expression of the COL6A4P2 gene was significantly correlated with survival rate in LUAD patients (Fig. 2B, p = 4.25 × 10⁻³). However, the expression of the COL6A4P2 gene had no significant effect on the prognosis of LUSC (p = 3.00 × 10⁻¹, Fig. 2D). Furthermore, GTEx database prediction results showed that four SNPs (rs34445363, p = 5.80 × 10⁻¹⁴; rs7625942, p = 8.90 × 10⁻⁸; rs77941834, p = 1.60 × 10⁻⁵; rs61733464, p = 1.00 × 10⁻⁹) on the COL6A4P2 gene are significantly expressed in normal lung tissues (Fig. 3).

4. Discussion

In this study we have analyzed the association of the COL6A4P2 gene polymorphisms towards susceptibility for LC. We identified that rs34445363 in COL6A4P2 gene was associated with an increased risk of LC. And our results also suggested that rs34445363 site mutations increase the risk of LUAD, while the mutation of rs61733464 significantly decrease the LUAD risk. Those suggested an association between genetic polymorphism of COL6A4P2 and the susceptibility of LC.

Numerous studies have shown that collagen levels play an important role in the development of LC (19, 20). Naveen et al. (21) identified collagen VI as a potential biomarker for early diagnosis of LC by proteomic analysis, suggesting that LC is associated with collagen-encoding genes. The COL6A4P2 gene is a pseudogene formed by the chromosomal break of the collagen-encoding gene COL6A4 (13, 22), so we speculate that the COL6A4P2 gene may be related to LC. Our results suggested that the rs34445363 mutation on the COL6A4P2 gene significantly increases the risk of LC, validating our conjecture, and consistent with previous studies.

Our results also found that the relationship between the COL6A4P2 gene polymorphism and LC risk was influenced by gender and age. A retrospective analysis of Oh et al. (23) assessed the important effects of gender and age in the development of LC. Aareleid et al. (24) revealed that LC has different incidence rates in different genders and ages. These studies were consistent with our results and enhance the credibility of our findings.

Further, we predicted the differential expression of COL6A4P2 gene in normal lung tissues and LC tissues through a database. Voiles et al. (12) found that collagen VI protein levels increased in tumor lung tissue, speculated that the expression of COL6A4P2 gene in tumor lung tissue is variable. This coincides with our predictions. Fagerberg et al. (25) found that the COL6A4P2 gene is specifically expressed in human lung tissue by genome-wide integration analysis of transcriptomics and antibody proteomics. These findings suggested the important research significance of COL6A4P2 gene in the development of LC, prompting that the COL6A4P2 gene deserves further study.

In conclusion, the present study is the first to investigate the relationship between COL6A4P2 gene and LC, and pointed out that COL6A4P2 gene polymorphism is associated with LC risk in Chinese Han population. However, further studies are warranted on larger patients from other ethnic groups to confirm our results.

Declarations

Ethics approval and consent to participate

This study was approved by the ethics committee of the Shaanxi Provincial Cancer Hospital, and conformed to the ethical principles for medical research involving humans of the World Medical Association Declaration of Helsinki. All participants signed informed consent forms before participating in this study.

Consent to publish

All the authors agreed to publish the manuscript.

Availability of data and materials

The datasets used and analyzed in the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no conflict of interest.

Funding

No

Authors' Contributions

Conceptualization, Mingwei Chen; methodology, Ying Duan; software, Gaowen Liu and Fanglin Niu; data curation, Jing Li and Mengdan Yan; writing, review and editing, Ying Duan and Tianbo Jin.

Acknowledgments

We thank all authors for their contributions and supports. We are also grateful to all participants for providing blood samples.

References

1. Kou Y, Li G, Shao J, Liu C, Wu J, Lu J, et al. Genome-Wide Profiling Reveals That Herbal Medicine Jinfukang-Induced Polyadenylation Alteration Is Involved in Anti-Lung Cancer Activity. *Evid Based Complement Alternat Med.* 2017;2017:5326909.
2. Albarqouni L, Elezzi K, Abu-Rmeileh NME. A comparison between health research output and burden of disease in Arab countries: evidence from Palestine. *Health Res Policy Syst.* 2018;16(1):25.
3. de Groot PM, Wu CC, Carter BW, Munden RF. The epidemiology of lung cancer. *Transl Lung Cancer Res.* 2018;7(3):220-33.
4. Sun KX, Zheng RS, Zeng HM, Zhang SW, Zou XN, Gu XY, et al. [The incidence and mortality of lung cancer in China, 2014]. *Zhonghua Zhong Liu Za Zhi.* 2018;40(11):805-11.
5. Bhimji SS, Wallen JM. *Cancer, Lung, Adenocarcinoma. StatPearls. Treasure Island (FL)2018.*
6. Cao M, Chen W. Epidemiology of lung cancer in China. *Thorac Cancer.* 2018.
7. Mohamed S, Bayoumi H, El-Aziz NA, Mousa E, Gamal Y. Prevalence, risk factors, and impact of lung Cancer on outcomes of idiopathic pulmonary fibrosis: a study from the Middle East. *Multidiscip Respir Med.* 2018;13:37.
8. Liu D, Wen H, He J, Gao S, Li S, Liu L, et al. Society for Translational Medicine Expert Consensus on the preoperative assessment of circulatory and cardiac functions and criteria for the assessment of risk factors in patients with lung cancer. *J Thorac Dis.* 2018;10(9):5545-9.
9. Li H, Yang Z, Li J, Lv X, Gao M, Bi Y, et al. Genetic variants in lncRNA HOTAIR are associated with lung cancer susceptibility in a Chinese Han population in China: a case-control study. *Cancer Manag Res.* 2018;10:5209-18.

10. Dimitrakopoulos FD, Antonacopoulou AG, Kottorou AE, Maroussi S, Panagopoulos N, Koukourikou I, et al. NF- κ B Genetic Variations are Significantly Associated with Non-Small Cell Lung Cancer Risk and Overall Survival. *Sci Rep*. 2018;8(1):5259.
11. Chiu KH, Chang YH, Wu YS, Lee SH, Liao PC. Quantitative secretome analysis reveals that COL6A1 is a metastasis-associated protein using stacking gel-aided purification combined with iTRAQ labeling. *J Proteome Res*. 2011;10(3):1110-25.
12. Voiles L, Lewis DE, Han L, Lupov IP, Lin TL, Robertson MJ, et al. Overexpression of type VI collagen in neoplastic lung tissues. *Oncol Rep*. 2014;32(5):1897-904.
13. Fitzgerald J, Rich C, Zhou FH, Hansen U. Three novel collagen VI chains, alpha4(VI), alpha5(VI), and alpha6(VI). *J Biol Chem*. 2008;283(29):20170-80.
14. Wu C, Wei Y, Zhu Y, Li K, Zhu Y, Zhao Y, et al. Identification of cancer-related potential biomarkers based on lncRNA-pseudogene-mRNA competitive networks. *FEBS Lett*. 2018;592(6):973-86.
15. Lynn H, Sun X, Ayshiev D, Siegler JH, Rizzo AN, Karnes JH, et al. Single nucleotide polymorphisms in the MYLK1 pseudogene are associated with increased colon cancer risk in African Americans. *PLoS One*. 2018;13(8):e0200916.
16. Wei CC, Nie FQ, Jiang LL, Chen QN, Chen ZY, Chen X, et al. The pseudogene DUXAP10 promotes an aggressive phenotype through binding with LSD1 and repressing LATS2 and RRAD in non small cell lung cancer. *Oncotarget*. 2017;8(3):5233-46.
17. Huang CY, Xun XJ, Wang AJ, Gao Y, Ma JY, Chen YT, et al. CHRNA5 polymorphisms and risk of lung cancer in Chinese Han smokers. *Am J Cancer Res*. 2015;5(10):3241-8.
18. Dai ZJ, Liu XH, Ma YF, Kang HF, Jin TB, Dai ZM, et al. Association Between Single Nucleotide Polymorphisms in DNA Polymerase Kappa Gene and Breast Cancer Risk in Chinese Han Population: A STROBE-Compliant Observational Study. *Medicine (Baltimore)*. 2016;95(2):e2466.
19. Shea MP, O'Leary KA, Wegner KA, Vezina CM, Schuler LA. High collagen density augments mTOR-dependent cancer stem cells in ERalpha+ mammary carcinomas, and increases mTOR-independent lung metastases. *Cancer Lett*. 2018;433:1-9.
20. Li Y, Chen Y, Ma Y, Nenkov M, Haase D, Petersen I. Collagen prolyl hydroxylase 3 has a tumor suppressive activity in human lung cancer. *Exp Cell Res*. 2018;363(1):121-8.
21. Chaudhary N, Bhatnagar S, Malik S, Katare DP, Jain SK. Proteomic analysis of differentially expressed proteins in lung cancer in Wistar rats using NNK as an inducer. *Chem Biol Interact*. 2013;204(2):125-34.
22. Gara SK, Grumati P, Urciuolo A, Bonaldo P, Kobbe B, Koch M, et al. Three novel collagen VI chains with high homology to the alpha3 chain. *J Biol Chem*. 2008;283(16):10658-70.
23. Oh JY, Lee YS, Min KH, Hur GY, Lee SY, Kang KH, et al. Presence of lung cancer and high gender, age, and physiology score as predictors of acute exacerbation in combined pulmonary fibrosis and emphysema: A retrospective study. *Medicine (Baltimore)*. 2018;97(31):e11683.
24. Aareleid T, Zimmermann ML, Baburin A, Innos K. Divergent trends in lung cancer incidence by gender, age and histological type in Estonia: a nationwide population-based study. *BMC Cancer*. 2017;17(1):596.
25. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. 2014;13(2):397-406.

Figures

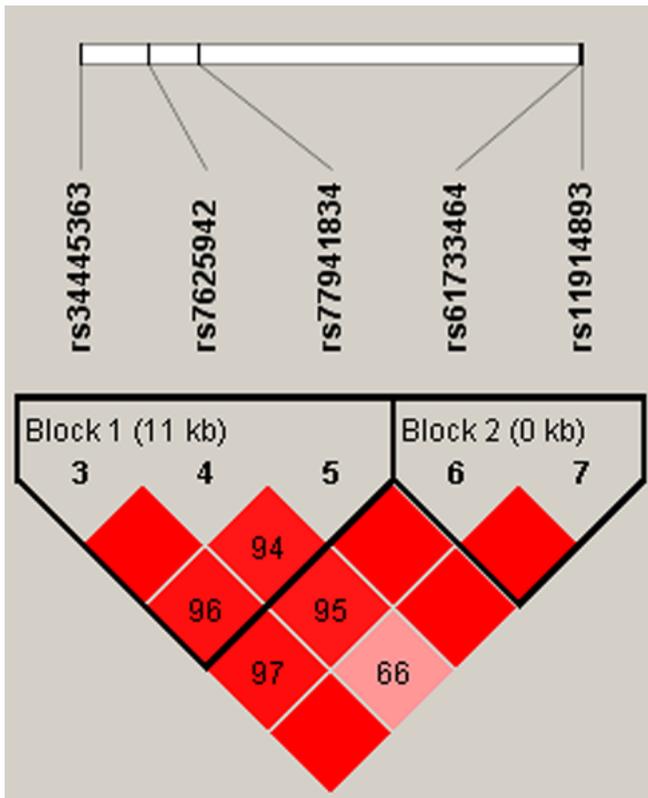


Figure 1

Haplotype block map for SNPs of the COL6A4P2 gene. Linkage disequilibrium plots containing 5 SNPs from COL6A4P2. Red squares display statistically significant associations between a pair of SNPs, as measured by D' ; darker shades of red indicate higher D' .

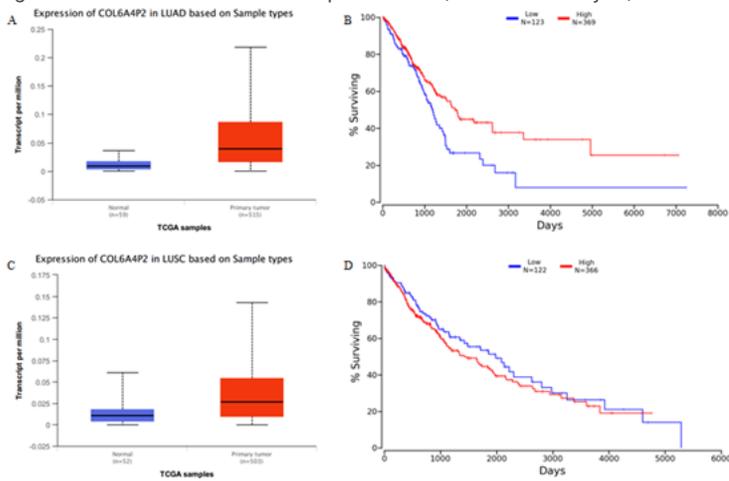


Figure 2

Prediction of the expression and prognosis of COL6A4P2 gene in LUAD and LUSC. (A) Expression of COL6A4P2 in LUAD and normal tissues ($p = 1.62 \times 10^{-12}$). (B) Effect of COL6A4P2 gene expression on survival rate ($p = 4.25 \times 10^{-3}$). (C) Expression of COL6A4P2 in LUSC and normal tissues ($p = 2.44 \times 10^{-7}$). (D) Effect of COL6A4P2 gene expression on survival rate ($p = 3.00 \times 10^{-1}$).

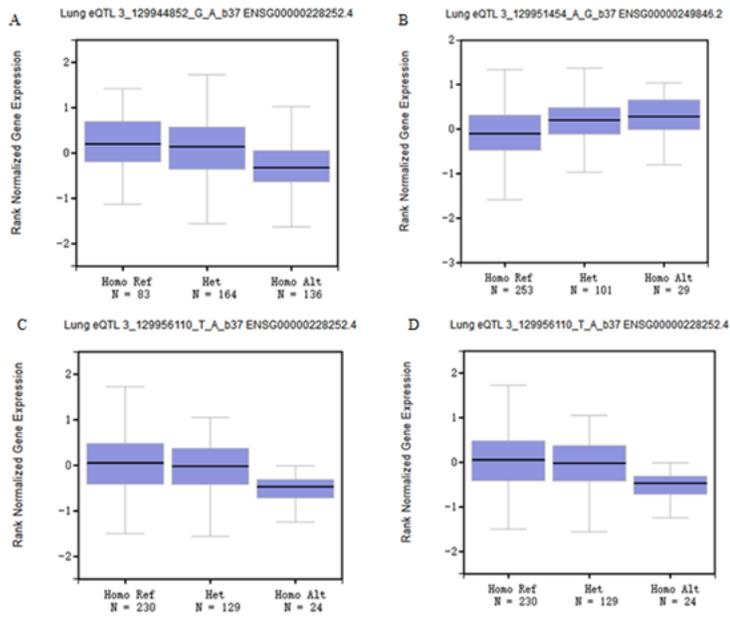


Figure 3

The expression of COL6A4P2 SNPs in lung tissues. A indicates the expression of rs34445363 genotype in lung ($p = 5.80 \times 10^{-14}$); B indicates the expression of rs7625942 genotype in lung ($p = 8.90 \times 10^{-8}$); C indicates the expression of rs77941834 genotype in lung ($p = 1.60 \times 10^{-5}$); D indicates the expression of rs61733464 genotype in lung ($p = 1.00 \times 10^{-9}$)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarytable2.docx](#)
- [Supplementarytable1.docx](#)