Collagen type VI $\alpha_5$ gene variations may predict the risk of lung cancer development in Chinese Han population

Ying Duan$^1$, Gaowen Liu$^2$, Yao Sun$^3$, Jiamin Wu$^3$, Zichao Xiong$^3$, Tianbo Jin$^3$ & Mingwei Chen$^1*$$^1$

The abundant expression of collagen type VI $\alpha_5$ (COL6A5) exists in lung tissue, and its role in lung cancer is still unknown. We performed a genetic association study with an attempt to detect the relationships between single nucleotide polymorphisms (SNPs) in COL6A5 and lung cancer predisposition in Chinese Han population. We finally selected six tag-SNPs to determine their genotypes among 510 lung cancer patients and 495 healthy controls with the MassARRAY platform. The associations of SNPs and lung cancer risk were estimated by logistic regression method with adjustment for confounding factors. Two available databases were used for gene expression and prognosis analysis. COL6A5 rs13062453, rs1497305, and rs77123808 were significantly associated with the risk of lung cancer in the whole population or stratified subgroups ($p < 0.05$). Among them, COL6A5 rs13062453 and rs1497305 were also linked to the susceptibility of lung adenocarcinoma. Additionally, rs1497305 was found to be strongly related to the TNM staging under five genetic models ($p < 0.05$). Results from databases suggested the important role of COL6A5 in lung cancer development. COL6A5 polymorphisms rs13062453, rs1497305 and rs77123808 were associated with lung cancer risk in Chinese Han population. These findings first yield new insight of COL6A5 in lung cancer.

The incidence and mortality rates of lung cancer rank first worldwide, with 2,093,876 (11.6% of total cancer patients) new cases and 1,761,007 (18.4% of total cancer deaths) deaths occurred globally in 2018$^1$. Considering the sex-distribution difference, lung cancer is the leading cause of cancer incidence (1,368,524 new cases) and death (1,184,947 deaths) among males, accounting for approximately 14.4% new cases and 22.0% deaths respectively, and the third prevalent cancer (725,352 new cases, 8.4%) with second fatality rate (576,060 deaths, 13.8%) among females$^1$. The morbidity and mortality of lung cancer is very high in China, even worse, continue rising in recent years$^2$. Previous epidemiological studies have reported a dramatical increase of crude incidence of lung cancer among Chinese from 2000 to 2014, which reached 57.13 per 100,000 in 2014 (74.31 per 100,000 in males and 39.08 per 100,000 in females)$^2$. In the same year, the estimated mortality rate of lung cancer deaths is 45.80 per 100,000 (61.10 per 100,000 in males and 29.71 per 100,000 in females)$^3$. Multiple etiological analyses have found that external risk factors such as smoking, air pollution and occupational exposure contribute to the lung tumorigenesis$^4$$^5$. Moreover, researchers have emphasized that genetic factors play critical roles in the development of lung cancer as well. Nowadays, accumulating evidence has elucidated that genetic polymorphisms are risk markers that can mediate individual susceptibility to lung cancer$^7$$^8$. These findings also provide multiple promising indicators that pave the road for individual cancer risk assessment. Therefore, exploring these novel factors is a vital step which not only reveal the underlying genetic mechanisms of lung cancer, but facilitate cancer risk evaluation among populations with diverse genetic backgrounds.

Collagen type VI has been identified as a extracellular matrix protein with functional role and ubiquitous distribution in numerous connective tissues, where it forms microfibrils organized into extend microfilamentous network$^{12}$. It has been well established for a long time that collagen VI is comprised of three genetically distinct $\alpha$-chains, namely collagen VI $\alpha_1$, $\alpha_2$ and $\alpha_3$, which is encoded by COL6A1, COL6A2, and COL6A3.

$^1$Department of Respiratory Medicine, The First Affiliated Hospital of School of Medicine of Xi’an Jiaotong University, Xi’an, Shaanxi, 710061, China. $^2$Xi’an Central hospital, Xi’anyang, Shaanxi, 712000, China. $^3$Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education. School of Medicine, Northwest University, Xi’an, Shaanxi, 710069, China. *email: chenmw36@163.com
respectively\(^{13}\). Subsequently, three new collagen VI chains were identified, and designated \(\alpha_4\), \(\alpha_5\) and \(\alpha_6\) chains that are encoded by distinct genes, \(COL6A4\), \(COL6A5\), and \(COL6A6\)^{14}. The aberrant expression of collagen VI genes has been reported in several malignant tumors, suggesting the potential effects of collagen VI in cancer development\(^{15–17}\). According to previous publications, the interaction between collagen VI and other components will lead to the remodeling of extracellular matrix, and stimulate the signaling pathway implicated in cell proliferation, migration, differentiation and apoptosis\(^{18–21}\). Collagen VI exerts important functions in maintaining the integrity of lung tissue\(^{22}\). And \(COL6A5\) (Collagen type VI \(\alpha_5\)), which is also known as \(COL29A1\), was highly expressed in lung tissue as well as skin, whereas less is reported on lung cancer\(^{23}\). Accordingly, we hypothesized that the single nucleotide polymorphisms (SNPs) in \(COL6A5\) might be associated with lung cancer risk, and performed a genetic association study to investigate the relationships between the \(COL6A5\) variations and lung cancer predisposition among Chinese Han population.

### Materials and Methods

#### Ethics statement and consent to subjects.

This study was performed with the approval of the ethics committee of The First Affiliated Hospital of School of Medicine of Xi’an Jiaotong University and all the procedures conformed to the 1964 Helsinki declaration and its later amendments. All the subjects were informed verbally and in writing of the protocols in this work. Signed informed consent was obtained prior to participation.

#### Study participants.

The present study was carried out based on two groups of sample: 510 patients with diagnosed lung cancer and 495 healthy individuals. All the participants were recruited from Shaanxi Provincial Cancer Hospital. Cases were hospitalized from 2015 to 2018 and histopathological diagnosed as primary lung carcinoma by at least two experienced pathologists. Electronic bronchoscopy, percutaneous lung puncture, spu-tum exfoliation and pleural effusion cytology examination were supplementary methods for diagnosis. Patients were further searched in dbSNP database (http://www.bioinfo.org.cn/relative/dbSNP%20Home%20Page.htm) to obtain detailed genetic information. A volume of 3–5 ml peripheral blood was donated from the participants and stored in ethylene diamine tetraacetic acid (EDTA) anticoagulant tubes at \(-80^\circ C\). GoldMag-Mini Whole Blood Genomic DNA Purification Kit (GoldMag Co. Ltd, Xi’an City, China) was used for genomic DNA extraction from each sample according to the standard protocol. The purified and concentrated DNA was evaluated by NanoDrop spectrophotometer (Thermo Scientific, Waltham, Massachusetts, USA) at wavelength of 260 nm and 280 nm. The qualified DNA was used as the template for the primary polymerase chain reaction (PCR), and followed by a unique base extension. Using the Agena MassARRAY method (Agena Bioscience, San Diego, CA, USA), the genotypes of the six variants were assayed with products of the last PCR step. The genotype data was read out by the Agena Bioscience TYPER software, version 4.0.
Statistical analysis. The distribution differences of the age and gender among cases and controls were assessed by independent samples t-test and Pearson's chi-squared test, respectively. Genotype frequency of the variations in the two study groups were compared with Pearson's chi-squared test. HWE p value was also calculated in controls using Pearson's chi-squared test. Furthermore, logistic regression analysis with adjustment was conducted to estimate the associations of the candidate SNPs with lung cancer susceptibility in genetic models (allele model, genotype model, dominant model, recessive model, and additive model) via available odds ratio (OR) values and 95% confidence intervals (CIs). The minor allele of each variation was hypothesized as the risk factor in our research. In statistical analysis, \( p = 0.05 \) was considered significant threshold. Stratified analyses by age, gender, pathological type, and TNM stage were carried out as well. SPSS version 20.0 (IBM SPSS, Inc., Chicago, IL, USA) and PLINK package, version 2.1.7 were used for basic characteristic statistics and risk evaluation.

Expression and prognosis analyses in databases. In order to further investigate the role of \( \text{COL6A5} \) gene in lung tumorigenesis, we performed a large sample-based expression detection in Gene Expression Profiling Interactive Analysis (GEPIA) database (http://gepia.cancer-pku.cn/) and overall survival analysis in Kaplan-Meier Plotter (Lung cancer, http://kmplot.com/analysis/index.php?p=service&cancer=lung). The expression levels of \( \text{COL6A5} \) were determined in lung cancer (including lung adenocarcinoma and lung squamous cell carcinoma) tissues and normal samples from healthy individuals. The prognostic differences between patients with high and low expression of \( \text{COL6A5} \) were assessed by hazard ratio (HR) and log-rank \( p \) value.

Results

Expression and prognosis patterns of \( \text{COL6A5} \) gene. With the GEPIA database, we detected the expression levels of \( \text{COL6A5} \) in lung tumors and normal tissues. Overall, \( \text{COL6A5} \) was down-regulated in both lung adenocarcinoma and lung squamous cell carcinoma (Fig. 2). Subsequently, prognosis analysis indicated that patients with low expression of \( \text{COL6A5} \) had worse overall survival (Fig. 3, \( \text{HR} = 0.67, 95\% \text{ CI} = 0.56–0.79, p < 0.001 \)).

Study subjects. The basic characteristics of the 510 lung cancer patients and 495 healthy individuals were compared and showed in Table 1. The case group consisted of 355 males and 155 females, with a mean age of 60.78 ± 9.96. Moreover, pathological type, TNM classification and lymph node metastasis status of the most lung cancer patients were also collected and involved in further stratified analysis. As for the controls, 346 males and 149 females were enrolled, which were matched to the cases in an approximately 1:1 fashion (\( p = 0.920 \)). The mean age of the healthy subjects was 61.94 ± 7.72 years in this study.

Basic information of the candidate SNPs. Specific information and minor allele frequency of the six candidate SNPs are listed in Table 2, including rs77123808 (C < A), rs2034664 (T < A), rs10212241 (T < C), rs13062453 (A < G), rs1497305 (A < G), and rs2403340 (A < G). The call rate for each SNP was over 99% in the current study. All the selected variants are located in \( \text{COL6A5} \), chromosome 3, with MAF > 0.05 in both cases and controls. From Table 2, we found that the frequency of minor allele “T” of rs10212241 (0.438 in cases and 0.438 in controls) and “A” of rs2403340 (0.175 in cases and 0.174 in controls) were similar among lung cancer patients and healthy subjects. Genotype distributions of the six \( \text{COL6A5} \) variants in lung cancer cases and healthy controls.
are showed in Supplementary Table S1. Pearson’s chi-squared test p value of rs1497305 indicated that there was difference in genotype distribution between the two groups (p = 0.020). In control group, the HWE p value of each variant was calculated (rs77123808: 1.00, rs2034664: 0.04, rs10212241: 1.00, rs13062453: 0.15, rs1497305: 0.09, rs2403340: 0.04). Because of the lower MAF in our population, rs2034664 and rs2403340 exhibited marginal HWE p values and were excluded from further analysis.

**Association analysis of COL6A5 variants with lung cancer risk.** We first investigated the relationships between rs77123808, rs10212241, rs13062453, rs1497305 and lung cancer risk in the Chinese Han population. As presented in Table 3, the heterozygous “GA” genotype of rs13062453 was associated with an enhanced risk of lung cancer when compared with the homozygous “GG” genotype (adjusted OR = 1.33, 95% CI = 1.02–1.74, p = 0.036, power = 61.77%). Genetic model analysis revealed statistical significance of this variants in dominant
Table 2. Descriptive information of the selected SNPs in COL6A5. SNP: Single nucleotide polymorphism; Chr: Chromosome.

| SNP          | Chromosome | Position     | Functional Consequence | Alleles A < B | Minor Allele Frequency | Call Rate |
|--------------|------------|--------------|------------------------|----------------|------------------------|-----------|
| rs77123808   | Chr 3      | 130368023    | intron variant         | C < A         | 0.338                  | 1.00      |
| rs2034664    | Chr 3      | 13037023     | intron variant         | T < A         | 0.183                  | 0.999     |
| rs10212241   | Chr 3      | 130385219    | transcript variant     | T < C         | 0.438                  | 0.999     |
| rs13062453   | Chr 3      | 130419022    | intron variant         | A < G         | 0.365                  | 0.999     |
| rs1497305    | Chr 3      | 130428544    | intron variant         | A < G         | 0.289                  | 0.999     |
| rs2403340    | Chr 3      | 130439499    | intron variant         | A < G         | 0.173                  | 0.999     |

| SNP          | Allele/Genotype | Control N (%) | Case N (%) | OR (95% CI) | p value |
|--------------|-----------------|---------------|------------|-------------|---------|
| rs77123808   | A               | 638 (64.44%)  | 675 (66.18%) | 1.00        | 0.415   |
|              | C               | 352 (35.56%)  | 345 (33.82%) | 0.93 (0.77–1.11) |         |
|              | AA              | 205 (41.41%)  | 221 (43.33%) | 1.00        |         |
|              | AC              | 228 (46.06%)  | 233 (45.69%) | 0.96 (0.73–1.24) | 0.735   |
|              | CC              | 62 (12.53%)   | 56 (10.98%)  | 0.85 (0.56–1.27) | 0.423   |
| rs10212241   | C               | 555 (56.17%)  | 573 (56.18%) | 1.00        | 0.999   |
|              | T               | 433 (43.83%)  | 447 (43.82%) | 1.00 (0.84–1.19) |         |
|              | CC              | 156 (31.58%)  | 163 (31.96%) | 1.00        |         |
| rs13062453   | G               | 664 (67.21%)  | 646 (63.46%) | 1.00        | 0.078   |
|              | A               | 324 (32.79%)  | 372 (36.54%) | 1.18 (0.98–1.42) |         |
|              | GG              | 230 (46.56%)  | 203 (39.88%) | 1.00        |         |
|              | GA              | 204 (41.30%)  | 240 (47.15%) | 1.33 (1.02–1.74) | 0.036*  |
|              | AA              | 60 (12.15%)   | 66 (12.97%)  | 1.21 (0.81–1.81) | 0.346   |
| rs1497305    | G               | 710 (72.45%)  | 722 (71.06%) | 1.00        | 0.492   |
|              | A               | 270 (27.55%)  | 294 (28.94%) | 1.07 (0.88–1.30) |         |
|              | GG              | 265 (54.08%)  | 247 (48.62%) | 1.00        |         |
|              | GA              | 180 (36.73%)  | 228 (44.88%) | 1.37 (1.05–1.78) | 0.019*  |
|              | AA              | 45 (9.18%)    | 33 (6.50%)   | 0.76 (0.47–1.24) | 0.269   |

Table 3. Evaluation of the correlation between COL6A5 variants and lung cancer susceptibility among Chinese Han population. SNP: Single nucleotide polymorphism; OR: Odds ratio; 95% CI: 95% confidence interval. The OR and 95% CI were calculated by logistic regression analysis. p value was obtained by Wald test with adjustment for age and gender. *Statistically significant results are in bold (p < 0.05).

Association analysis in stratified subgroups. Furthermore, stratified analysis was performed in order to explore the effects of these SNPs on lung cancer in different subgroups. First, the correlations between the genetic variations and lung cancer risk were evaluated among people with different age and gender, and the significant COL6A5 variants were summarized in Table 4 (allele and genotype analysis) and Supplementary Table S3 (Genetic model analysis). Among the individuals aged younger than 61 years, the rs1497305 “GA” carriers harbored an increased risk of lung cancer when compared with “GG” carriers (adjusted OR = 1.30, 95% CI = 1.01–1.68, p = 0.039, power = 54.43%; Supplementary Table S2). Likewise, remarkable correlation also existed between the “GA” genotype of rs13062453 and increased lung cancer susceptibility among Chinese Han individuals according to an adjusted OR of 1.37 (95% CI = 1.05–1.78, power = 69.11%) and significant p value of 0.019 (Table 3). However, rs77123808 and rs10212241 did not show any associations with lung cancer in our study cohort (p > 0.05).
Table 4. Significant COL6A5 variants associated with lung cancer susceptibility in stratified analysis for age or gender. SNP: Single nucleotide polymorphism; OR: Odds ratio; 95% CI: 95% confidence interval. The OR and 95% CI were calculated by logistic regression analysis. *Statistically significant results are in bold (p < 0.05).

Discussion

The current study evaluated the correlations between polymorphisms in COL6A5 gene and lung cancer susceptibility among Chinese Han individuals. We found that COL6A5 rs13062453, rs1497305, and rs77123808 were associated with the risk of lung cancer in Chinese Han population. Additionally, COL6A5 rs1497305 was a potential predictor for TNM staging. Basing on the expression patterns presented by databases, we found a down-regulation of COL6A5 in lung tumors when compared with the normal tissues. The low expression level was linked to the worse overall survival of lung cancer patients. These results preliminarily elucidated the potential genetic determinants in COL6A5 of lung cancer, and first uncovered the involvement of COL6A5 gene in lung carcinogenesis.

Collagen family consists of 28 types of extracellular matrix proteins with α chains, which can facilitate cell growth and maintain the mechanical elasticity of connective tissue44. Collagen type VI plays a prominent role in both tumorigenesis and metastasis55. Collagen type VI α1, α2 and α3 are three major extracellular matrix protein of collagen type VI subgroup, which can promote tumor growth55. Collagen type VI α4, α5 and α6 are three members that are identified lately, with a N-terminal region containing seven vWF-A modules, a collagen triple helical region, and a C-terminal region made of two or three vWF-A modules as well as specific sequence46,29. However, it is largely unknown how collagen type VI α4, α5 and α6 are implicated in tumor development. COL6A6 has been reported to be down-regulated in breast cancer27,29. Significant down-regulation of COL6A5 has been proved in esophageal squamous cell carcinoma30 and fibrous epulis31. Wang and the colleagues has found that collagen type VI may interacts with ROBO2 in esophageal squamous cell carcinoma development, and ROBO2 serves as a tumor suppressor during malignant process32. Collagen type VI serves as an indispensable protein in extracellular matrix, and mainly deposited in lung fibrosis33. Pulmonary fibrosis, combined with emphysema, has been fully demonstrated to be strongly associated with the development of lung cancer34,35. However, the characteristics of COL6A5 gene in lung cancer remains incompletely understood. Bioinformatics analyses showed a decrease of COL6A5 level in lung tumor, and the low expression resulted in poor prognosis.
In our study, \textit{COL6A5} found any significant relationships between variants and asthma or chronic obstructive pulmonary disease for age and gender. Statistically significant results are in bold (\(* p < 0.05\)).

Table 5. Significant \textit{COL6A5} variants in stratified analysis for pathological type. SNP: Single nucleotide polymorphism; OR: Odds ratio; 95% CI: 95% confidence interval. The OR and 95% CI were calculated by logistic regression analysis. \(* p < 0.05\).

| SNP       | Allele/Genotype | Control N (%) | Case N (%) | OR (95% CI) | \(p\)-value |
|-----------|-----------------|---------------|------------|-------------|-------------|
| rs13062453| G               | 233 (61.97%)  | 157 (65.97%)| 1.00        | 0.715       |
| A         | 324 (32.79%)    | 143 (38.03%)  | 81 (34.03%)| 1.06 (0.78–1.43) |             |
| GG        | 230 (46.56%)    | 70 (37.23%)   | 52 (43.70%)| 1.00        |             |
| GA        | 204 (41.30%)    | 93 (49.47%)   | 53 (44.54%)| 1.18 (0.76–1.83) | 0.456       |
| AA        | 60 (12.15%)     | 25 (13.30%)   | 11 (6.18%) | 1.07 (0.55–2.09) | 0.847       |

| SNP       | Allele/Genotype | Control N (%) | Case N (%) | OR (95% CI) | \(p\)-value |
|-----------|-----------------|---------------|------------|-------------|-------------|
| rs1497305 | G               | 261 (69.41%)  | 168 (70.59%)| 1.00        | 0.366       |
| A         | 270 (72.55%)    | 116 (39.47%)  | 70 (29.41%)| 1.10 (0.80–1.50) |             |
| GG        | 265 (54.08%)    | 93 (46.73%)   | 59 (49.58%)| 1.00        |             |
| GA        | 180 (36.73%)    | 97 (47.84%)   | 50 (42.02%)| 1.24 (0.81–1.91) | 0.323       |
| AA        | 45 (9.18%)      | 9 (4.52%)     | 10 (4.80%) | 1.08 (0.50–2.31) | 0.848       |

| SNP       | Allele/Genotype | Control N (%) | Case N (%) | OR (95% CI) | \(p\)-value |
|-----------|-----------------|---------------|------------|-------------|-------------|
| rs1497305 | G               | 377 (73.78%)  | 261 (69.41%)| 1.00        | 0.011*      |
| A         | 141 (27.22%)    | 63 (37.50%)   | 60 (37.50%)| 0.62 (0.43–0.90) |             |
| GG        | 134 (51.74%)    | 32 (38.10%)   | 52 (12.15%)| 1.00        |             |
| GA        | 109 (42.08%)    | 41 (48.81%)   | 50 (42.02%)| 0.32 (0.13–0.76) | 0.010*      |
| AA        | 16 (6.18%)      | 11 (13.10%)   | 10 (4.52%) | 0.63 (0.37–1.07) | 0.089       |

We thus speculated that \textit{COL6A5} functions as a tumor suppressor gene in oncogenesis of lung. More convincing experimental evidence is needed to confirm these findings.

Single nucleotide substitutions in genes have been validated to be the risk factors in diverse diseases, which are proposed as predictors for individual risk assessment. Susceptible SNPs of \textit{COL6A5} have been identified and demonstrated in familial neuropathic chronic itch in previous research. Eight variants located in \textit{COL6A5} gene have been demonstrated to be related to the susceptibility to atopic dermatitis among Europeans. A longitudinal exome-wide association study performed among Japanese has showed that \textit{COL6A5} SNP rs11917356 was significantly correlated with hypertension risk through affect systolic blood pressure. Furthermore, it has been reported that variations in \textit{COL6A5} might play causal roles for asthma together with environmental exposures. But a recent study has not found any significant relationships between \textit{COL6A5} variants and asthma or chronic obstructive pulmonary disease in German. In our study, \textit{COL6A5} polymorphisms rs13062453, rs1497305, and rs77123808 were genetic polymorphic markers that confer susceptibility to lung cancer among Chinese Han individuals. All the significant SNPs were resided in the intron region of \textit{COL6A5} gene, which possesses regulatory functions in pre-mRNA processing. We considered that alternation at these polymorphic sites might modulate the expression efficiency of \textit{COL6A5} mRNA, contributing to the abnormal pattern of \textit{COL6A5} and thereby influencing the individual susceptibility to lung cancer. Additionally, \textit{COL6A5} rs1497305 was found to be associated with TNM staging, which harbors the potential to become new target for lung cancer clinical stage evaluation in future.

There are several limitations in our research that should be acknowledged. First, all the cases and controls are recruited from one hospital, and the possible selection bias may exist. The random subject selection could help us reduce the selection bias to some degree. Second, the expression detection and prognosis analysis of \textit{COL6A5} gene were carried out by databases, which should be verified with well-designed experimental study. Third, the underlying mechanism of the promising SNPs and \textit{COL6A5} gene in lung cancer development is still unclear, thus suggesting the nest step for the further research.

In summary, this is the first research that discussed the relationships between \textit{COL6A5} polymorphisms and lung cancer risk. We uncovered significant variants that were associated with lung cancer susceptibility in Chinese Han population. These findings also provided supporting evidence for the involvement of \textit{COL6A5} gene in the occurrence of lung cancer.

**Data availability**
The data generated or analyzed during this study are available from the corresponding author on reasonable request.

Received: 31 December 2018; Accepted: 17 February 2020;
Published online: 19 March 2020
References

1. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians 68, 394–424, https://doi.org/10.3322/caac.21492 (2018).

2. Chen, J. et al. Epidemiology of lung cancer in China. Thoracic Cancer 6, 209–215 (2015).

3. Chen, W. et al. Cancer incidence and mortality in China, 2014. Chinese journal of cancer research = Chung-kuo yen cheng yen chiu 30, 1–12, https://doi.org/10.1111/tjcr.12114 (2018).

4. Chen, W., Zhang, S. & Zou, X. Evaluation on the incidence, mortality and tendency of lung cancer in China. Thoracic Cancer 1, 35–40, https://doi.org/10.1111/1759-7714.10001.x (2010).

5. Zhang, W., Li, F. & Gao, W. Traffic-related air pollution and lung cancer: A meta-analysis. Thoracic Cancer 8, 546, https://doi.org/10.1111/1759-7714.12440 (2017).

6. Williams, R. F. & Brian, L. W. Occupational and environmental causes of lung cancer. Clinics in Chest Medicine 33, 681–703 (2012).

7. Bousse, Y. & Amos, C. I. A decade of GWAS results in lung cancer. Cancer Epidemiol Biomarkers Prev, cebp.0794.2016 (2017).

8. Weissfeld, J. L. et al. Lung Cancer Risk Prediction Using Common SNPs Located in GWAS-Identified Susceptibility Regions. Journal of Thoracic Oncology 10, 1538–1545 (2015).

9. Lei, G. et al. Polymorphisms in the TERT gene are associated with lung cancer risk in the Chinese Han population. European Journal of Cancer Prevention 23, 497–501 (2014).

10. Hsu, Y. et al. Genetic variation in the TP63 gene is associated with lung cancer risk in the Han population. Tumour Biology the Journal of the International Society for Onco developmental Biology & Medicine 35, 1863–1866 (2014).

11. Huang, C. Y. et al. CHRNA5 polymorphisms and risk of lung cancer in Chinese Han smokers. Am J Cancer Res 5, 3241–3248 (2015).

12. Sabatelli, P. et al. Expression of the collagen VI alpha5 and alpha6 chains in normal human skin and in skin of patients with collagen VI-related myopathies. The Journal of investigative medicine 131, 99–107, https://doi.org/10.1038/jid.2010.284 (2011).

13. Fitzgerald, J., Holdren, P. & Hansen, U. The expanded collagen VI family: new chains and new connections. Connective tissue research 54, 345–350, https://doi.org/10.1023/A:1008257020138 (2005).

14. Fitzgerald, J., Rich, C., Zhou, F. H. & Hansen, U. Three novel collagen VI chains, alpha4(VI), alpha5(VI), and alpha6(VI). The Journal of biological chemistry 283, 20170–20180, https://doi.org/10.1074/jbc.M710139200 (2008).

15. Han, J., Daniel, J. C. & Pappas, G. D. Expression of type VI collagen during glioblastoma cell invasion in brain tissue cultures. Cancer Lett 88, 127–132 (1995).

16. Daniels, K. J. et al. Expression of type VI collagen in uveal melanoma: its role in pattern formation and tumor progression. Laboratory Investigation 75, 55–66 (1996).

17. Iyengar, P. et al. Adipocyte-derived collagen VI affects early mammary tumor progression in vivo, demonstrating a critical interaction in the tumor/stroma microenvironment. The Journal of clinical investigation 115, 1163–1176, https://doi.org/10.1172/jci32424 (2005).

18. Sherman-Baust, C. et al. Remodeling of the extracellular matrix through overexpression of collagen VI contributes to cisplatin resistance in ovarian cancer cells. Cancer Cell 3, 377–386 (2003).

19. Emmanuelle, T., Blandine, G., Robert, G. & Stallcup, W. B. NG2 proteoglycan mediates beta1 integrin-independent cell adhesion and spreading on collagen VI. Journal of Cellular Biochemistry 280, 762–763 (2002).

20. Naugle, J. E. et al. Type VI collagen induces cardiac myofibrillar differentiation: implications for postinfarction remodeling. American journal of physiology. Heart and circulatory physiology 290, H323–H330, https://doi.org/10.1152/ajpheart.00321.2005 (2006).

21. Ruhl, M. et al. Soluble Collagen VI Drives Serum-starved Fibroblasts through S Phase and Prevents Apoptosis via Down-regulation of Bax. Journal of Biological Chemistry 274, 34361–34368 (1999).

22. Larry, V. et al. Overexpression of type VI collagen in neoplastic lung tissues. Oncology Reports 32, 1897–1904 (2014).

23. Sabatelli, P. et al. Expression of collagen VI alpha5 and alpha6 chains in human muscle and in Duchenne muscular dystrophy-related muscle fibrosis. Matrix biology: journal of the International Society for Matrix Biology 31, 187–196, https://doi.org/10.1016/j.matbio.2011.12.003 (2012).

24. Sorushanova, A. et al. The Collagen Suprafamily: From Biosynthesis to Advanced Biomaterial Development. Advanced materials (Deerfield Beach, Fla.) 31, e1801651, https://doi.org/10.1002/adma.201801651 (2019).

25. Chen, P., Cescon, M. & Bonald, P. Collagen VI in cancer and its biological mechanisms. Trends in molecular medicine 19, 410–417, https://doi.org/10.1016/j.molmed.2013.04.001 (2013).

26. Gara, S. K. et al. Three novel collagen VI chains with high homology to the alpha3 chain. The Journal of biological chemistry 283, 10658–10670, https://doi.org/10.1074/jbc.M709540200 (2008).

27. Yeh, M. H. et al. Extracellular Matrix-receptor Interaction Signaling Genes Associated with Inherited Breast Cancer Survival. Anticancer research 38, 4593–4605, https://doi.org/10.21873/anticancer.12764 (2018).

28. Makouki, J. et al. Gene expression profiling of breast cancer in Lebanese women. Scientific reports 6, 36639, https://doi.org/10.1038/srep36639 (2016).

29. Soure, M. K. et al. Gene expression comparison between primary triple-negative breast cancer and paired axillary and sentinel lymph node metastasis. The breast journal https://doi.org/10.1111/tbj.13684 (2019).

30. Li, J. et al. The clinical significance of collagen family gene expression in esophageal squamous cell carcinoma. PeerJ 7, e7705, https://doi.org/10.7717/peerj.7705 (2019).

31. Jiang, Y., Fang, B., Xu, B. & Chen, L. The RAS-PI3K-AKT-NF-kappaB pathway transcriptionally regulates the expression of BCL2 family and IAP family genes and inhibits apoptosis in fibrosarcoma cells. Journal of clinical laboratory analysis, e23102, https://doi.org/10.1002/jcla.23102 (2019).

32. Trifonov, V., Pasqualello, L., Dalla Fava, R. & Rabdan, R. MutComFocal: an integrative approach to identifying recurrent and focal genomic alterations in tumor samples. BMC systems biology 7, 25, https://doi.org/10.1186/s12918-019-0573-9 (2013).

33. Specks, U., Nerlich, A., Colby, T. V., Wiest, I. & Timpf, R. Increased expression of type VI collagen in lung fibrosis. American Journal of Respiratory and Critical Care Medicine 151, 1956–1964 (1995).

34. Li, Y. et al. Effect of emphysema on lung cancer risk in smokers: a computed tomography-based assessment. Cancer Prevention Research 4, 43–50 (2011).

35. Usui, K., Tanai, C., Tanaka, Y., Noda, H. & Ishihara, T. The prevalence of pulmonary fibrosis combined with emphysema in patients with lung cancer. Respiratory medicine 16, 326–331 (2007).

36. Martinelli-Boneschi, E. et al. COL6A5 variants in familial neuropathic chronic itch. Brain: a journal of neurology 140, 555–567, https://doi.org/10.1093/brain/aww343 (2017).

37. Soderhall, C. et al. Variants in a Novel Epidermal Collagen Gene (COL29A1) Are Associated with Atopic Dermatitis. Vol. 5 (2007).

38. Vauschek, Y. et al. Longitudinal exome-wide association study to identify genetic susceptibility loci for hypertension in a Japanese population. Experimental & Molecular Medicine 49, e409 (2017).

39. Bornelov, S. et al. Rule-Based Models of the Interplay between Genetic and Environmental Factors in Childhood Allergy. Plos One 8, e80800 (2013).

40. Castro-Giner, F. et al. A pooling-based genome-wide analysis identifies new potential candidate genes for atopy in the European Community Respiratory Health Survey (ECRHS). BMC Medical Genetics 128–128 (2009).

41. Monika, H. et al. Variation in the COL29A1 gene in German patients with atopic dermatitis, asthma and chronic obstructive pulmonary disease. Journal of Dermatology 37, 740–742 (2010).

42. Cherev, M. & Carmel, L. The function of introns. Frontiers in genetics 3, 55, https://doi.org/10.3389/fgene.2012.00055 (2012).
Acknowledgements
The authors would like to thank all the patients and control individuals for their participation, and the support of our cooperated team from Oncology department, Shaanxi Provincial Cancer Hospital. We also appreciate the reviewers and editors for their patience to this work.

Author contributions
Y.D., G.L. and M.C. conceived this study. Y.D., G.L., Y.S. and J.W. collected the blood samples and performed the genotyping experiment. Z.X., T.J. and M.C. conducted the statistical analysis. Y.D., G.L., T.J. and M.C. wrote the manuscript and all the authors revised article and approved its final version.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-61614-x.

Correspondence and requests for materials should be addressed to M.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access
This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020