Original Article

MUC16 in non-small cell lung cancer patients affected by familial lung cancer and indoor air pollution: clinical characteristics and cell behaviors

Ying Chen[1], Yunchao Huang[1], Madiha Kanwal[2], Guangjian Li[1], Jiapeng Yang[1], Huatao Niu[3], Zhenhui Li[4], Xiaojie Ding[5]

[1]Department of Thoracic Surgery I, the Third Affiliated Hospital of Kunming Medical University (Yunnan Cancer Hospital, Yunnan Cancer Center), Kunming 650106, China; [2]The Laboratory of Cancer Cell Biology, Institute of Molecular Genetics, ASCR, Videnska, Prague, Czech Republic; [3]Department of Neurosurgery, [4]Department of Radiology, [5]The Key Laboratory of Lung Cancer Research, the Third Affiliated Hospital of Kunming Medical University (Yunnan Cancer Hospital, Yunnan Cancer Center), Kunming 650106, China

Contributions: (I) Conception and design: Y Chen, X Ding; (II) Administrative support: Y Huang; (III) Provision of study materials or patients: Y Chen, Y Huang, G Li, J Yang; (IV) Collection and assembly of data: Y Chen, M Kanwal, H Niu, Z Li; (V) Data analysis and interpretation: Y Chen, M Kanwal, G Li, J Yang, X Ding; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Xiaojie Ding. The Key Laboratory of Lung Cancer Research, The Third Affiliated Hospital of Kunming Medical University (Yunnan Cancer Hospital), 519 Kunzhou Road, Kunming 650106, China. Email: dingxj2001@163.com.

Background: Inherited susceptibility and environmental carcinogens are crucial players in lung cancer etiology, and both exhibit population heterogeneity. *MUC16* is overexpressed in various cancers and often associated with poor prognosis. Present work was to investigate the clinical significance of *MUC16* in non-small cell lung cancer patients affected by familial lung cancer (FLC) and indoor air pollution caused by coal use.

Methods: Clinicopathologic characteristics and *MUC16* expression were analyzed and evaluated in our subject population. Vectors were constructed for *MUC16* gene knockout and overexpression, then we examined how *MUC16* affected lung cancer cell behaviors, including proliferation, migration, invasion and chemoresistance.

Results: FLC showed significant association with early-onset (P<0.01) and later stage (P<0.01). Indoor air pollution was associated with younger age (P<0.01), later stage (P<0.05) and AD histology type (P<0.05). Interestingly, two age peaks were observed in our FLC and sporadic group respectively, possibly suggesting multiple major contributors to lung cancer in our subject population. *MUC16* overexpression was significantly associated with FLC (P<0.05), indoor air pollution (P<0.01) and later stage (P<0.01), additionally more metastasis cases were observed in patients with up-regulated *MUC16* (18.1% *vs.* 10.3%). Taken together, elevated *MUC16* may potentially be one molecular character of FLC in local residents. Intriguingly, patients with more *MUC16* up-regulation seemed to have a lower number of white blood cells, especially neutrophils, this reflected *MUC16*'s role in immune regulation. In cell behavior experiments, high *MUC16* level could contribute to lung cancer cell proliferation, migration, invasion and chemoresistance, but there were variations among cell lines.

Conclusions: *MUC16* plays crucial roles in lung cancer pathogenesis, progression and chemoresistance. Interestingly, its association with FLC and indoor air pollution highlights the complexity of lung cancer etiology. Our findings provide useful information to study the intricate interaction between environmental carcinogens and population genetic background.

Keywords: Indoor air pollution; familial lung cancer (FLC); *MUC16*; chemoresistance

Submitted Jan 31, 2019. Accepted for publication Jul 11, 2019.
doi: 10.21037/tlcr.2019.07.10
View this article at: http://dx.doi.org/10.21037/tlcr.2019.07.10

Introduction

Lung cancer is the leading cancer diagnosed worldwide (1.8 million/year) and has a mortality rate higher than the next three cancers combined (158,080 *vs.* 115,760 deaths) (1). Unfortunately, lung cancer survival remains poor, owing to diagnosis at advanced stage and resistance to standard chemotherapy (2). In order to study the complex elements in lung cancer development, we recruited lung cancer patients from China's Yunnan Province, including Xuanwei/Fuyuan region, which reported some of the highest lung cancer rates in the world (3-6). Importantly, this subject population has special features: indoor air pollution caused by coal combustion and family aggregation of lung cancer (3-6).

Familial lung cancer (FLC) exhibits special characteristics when compared with its sporadic counterpart. Previous work revealed heterogeneity in different FLC populations. Many reported an increased lung cancer risk in FLC populations (7-12); some supported that FLC has a bigger effect in certain ethnic groups (7,8); others suggested female relatives have a higher risk than male relatives (9,10). Actually, lung cancer susceptibility could be inherited in complex patterns through generations, and there can be unique characteristics within each group or subpopulation.

MUC16 is a large transmembrane glycoprotein (20–25 mD) with 22,152 amino acid residues (13-15). *MUC16* is overexpressed and associated with poor prognosis in various cancers, including lung cancer (14-17). Some studies showed that *MUC16* could be potential therapy target for cancer patients (13,18,19). One study based on Cancer Genome Atlas reported that *MUC16* was among the top mutated genes (*TP53*, *USH2A*, *TTN*, *MUC16*) in different cancers, including lung cancer (20). *MUC16* has been shown to be associated with enhanced cancer cell growth, metastasis and chemoresistance (16,21-26), which are typical features of increased cancer aggressiveness.

Present work was designed to investigate the expression and clinical significance of *MUC16* in non-small cell lung cancer (NSCLC) patients, affected by FLC and indoor air pollution caused by coal use, in China's Yunnan Province; furthermore, to evaluate the role of *MUC16* in the proliferation, migration, invasion and chemosensitivity of lung cancer cells.

Methods

Patients and tissue samples

Present study was designed to investigate the clinical significance of *MUC16* in NSCLC patients affected by FLC and indoor air pollution in Yunnan, China. Patients were selected from those enrolled in Department of Thoracic Surgery I of Yunnan Cancer Hospital from Sep. 2015 to Jun. 2017. Subjects were selected based on the following criteria: (I) The case population was mainly composed of residents from Xuanwei/Fuyuan region of Yunnan Province, who primarily use coal for heating or cooking for more than 10 years; (II) the control subjects were patients from other areas in the same province, who reported no history of occupational or domestic coal use. In total, 185 cases and 92 controls were enrolled; (III) subjects with FLC were defined as individuals with three or more first-degree relatives affected by lung cancer. There were 51 patients classified as having FLC. All the information was based on self-report and confirmed by personal medical records.

Clinicopathologic data were documented in hospital cooperated databank (https://www.linkdoc.com). The TNM stage was reviewed according to the 8th edition of The International Association for the Study of Lung Cancer (IASLC) staging system. Clinicopathologic data were shown in *Table 1* and *Table S1*, majority of patients enrolled had adenocarcinoma (AD) and squamous cell carcinoma (SCC). The study was approved by the Ethical Committees of Yunnan Cancer Hospital (No. KY2019.57). All patients provided informed consent.

Tissue sample pairs including cancer and adjacent nonmalignant tissue of the same patient were stored in RNAlater (Sigma, St. Louis, MO, USA) immediately after surgery. A slide was cut from every sample for HE stain. Those containing >60% cancer tissue and <15% necrosis were used for study. Absolute neutrophil count (ANC) and absolute lymphocyte count (ALC) value were obtained before surgery or major treatment.

Quantitative real-time polymerase chain reaction (q-PCR)

Total RNA was extracted from tissue or cultured cells using TRIZOL reagent (Invitrogen, Carlsbad, CA, USA) Reverse transcription was performed using Promega reverse transcription kit (Promega, Madison, WI, USA). q-PCR

Table 1 Clinical characteristics of 277 NSCLC patients divided by familial lung cancer or indoor air pollution exposure

Variables	Total	Familial lung cancer, n (%)			Indoor air pollution, n (%)		
		Positive	Negative	P values[a]	Positive	Negative	P values[a]
Total number of patients	277	51	226		185	92	
Gender				0.293			0.762
Male	127	20 (39.2)	107 (47.3)		86 (46.5)	41 (44.6)	
Female	150	31 (60.8)	119 (52.7)		99 (53.5)	51 (55.4)	
Average age (range 30–82 years)				0.000398			0.00103
<54 years	150	39 (76.5)	111 (49.1)		113 (61.1)	37 (40.2)	
≥54 years	127	12 (23.5)	115 (50.9)		72 (38.9)	55 (59.8)	
Histology type				0.136			0.0351
Adenocarcinoma	264	51 (100.0)	213 (94.2)		180 (97.3)	84 (91.3)	
Squamous cell carcinoma	13	0 (0.0)	13 (5.8)		5 (2.7)	8 (8.7)	
Metastasis				0.526			0.0532
Negative	236	42 (82.4)	194 (85.8)		163 (88.1)	73 (79.3)	
Positive	41	9 (17.6)	32 (14.2)		22 (11.9)	19 (20.7)	
Stage				2.0608E-7			0.0364
I–II	123	6 (11.8)	117 (51.8)		74 (40.0)	49 (53.3)	
III–IV	154	45 (88.2)	109 (48.2)		111 (60.0)	43 (46.7)	
Average ANC				0.425			0.971
$<3.89\times10^9$/L	160	32 (62.7)	128 (56.6)		107 (57.8)	53 (57.6)	
$\geq3.89\times10^9$/L	117	19 (37.3)	98 (43.4)		78 (42.2)	39 (42.4)	
Average ALC				0.905			0.0362
$<1.99\times10^9$/L	150	28 (54.9)	122 (54.0)		92 (49.7)	58 (63.0)	
$\geq1.99\times10^9$/L	127	23 (45.1)	104 (46.0)		93 (50.3)	34 (37.0)	
Average ANC + ALC				0.951			0.890
$<5.88\times10^9$/L	164	30 (58.8)	134 (59.3)		109 (58.9)	55 (59.8)	
$\geq5.88\times10^9$/L	113	21 (41.2)	92 (40.7)		76 (41.1)	37 (40.2)	

[a], P value calculated by chi-square test or Fisher's exact test, when there is at least one cell with expected count less than 5. NSCLC, non-small cell lung cancer; ANC, absolute neutrophil count; ALC, absolute lymphocyte count.

was performed using SYBR Green Master Mix (Thermo Fisher, waltham, MA, USA)

Vectors construction for MUC16 gene knockout and overexpression

CRISPR-Cas9 vectors were constructed for *MUC16* gene knockout as described in (27). In order to effectively knockout *MUC16* gene, two sgRNA were combined to target the first exon of *MUC16* (PX459-*MUC16*-sgRNA-1 and PX459-*MUC16*-sgRNA-2). Lenti-CRISPR-dCas9 system was used for *MUC16* overexpression, three sgRNA were used simultaneously to increase activation efficiency, the vector construction and lentivirus packaging followed protocols in (28). PX459 and Lenti-CRISPR-dCas9 system were gift from Feng Zhang (*Table S2*). The sgRNA sequences were designed using CRISPRdirect (29) (http://crispr.dbcls.jp/) and listed in *Tables S3*,*S4*. More

information is in the Supplementary File.

Cell culture, plasmid transfection and lentivirus infection

The cell lines used in this study were kindly provided by Cell Bank and Stem Cell Bank, Chinese Academy of Sciences. H23 and H838 lung cancer cells were cultured in RPMI medium supplemented with 10% FBS and antibiotics. Similarly, lentivirus packaging cell line 293T was cultured in DMEM medium with the above-mentioned supplements. The cells were incubated at 37 ℃ with 5% CO_2.

MUC16-knockout vectors (PX459-*MUC16*-sgRNA-1; PX459-*MUC16*-sgRNA-2) were transfected into target cells using Lipofectamine 2000 (Thermo Fisher, waltham, MA, USA) according to the manufacturer's instruction, empty vector was used as control. Lentivirus infection were carried out as mentioned in (28) with empty virus as control. Transfection and infection were performed freshly for each cell behavior experiment, *MUC16* levels were monitored by q-PCR, cell populations with more than 60% *MUC16* decrease and more than 3 times *MUC16* increase were immediately used for the behavior experiments.

Immunoblot analysis

Cells were grown for 48 h after transfection or infection, then lysed using RIPA buffer (TIANGAN, Beijing, China), and the protein contents were measured using BCA Kit (TIANGAN). An amount of 60 μg protein from each sample was subjected to SDS-PAGE gel (5%) for electrophoresis, then transferred to PVDF membrane (Millipore, Bedford, MA, USA) and blocked in skim milk (5%) for 1 h. The membranes were incubated with primary antibody: mouse anti-*MUC16* (Abcam, Cambridge, MA, USA) 1:500 in 1% BSA for 2 h at 37 ℃; for loading control: mouse anti–β-*actin* (Santa Cruz Biotechnology, Santa Cruz, CA, USA) 1:1,000 in 1% BSA for 2 h at 37 ℃. After washing, the membranes were incubated with secondary antibody: goat anti-mouse IgG peroxidase labeled (KPL, Gaithersburg, MD, USA). Proteins were detected by X-ray film (kodak) in a dark room using Immobilon Western Chemiluminescent HRP Substrate (Millipore).

Proliferation assay, cell migration and invasion assay

For growth kinetics analyses, cells were seeded in 96-well plates at a density of 100–500 cells/well, each group had 6 plates. MTT assay kit (TIANGAN) was used to reflect the number of viable cells present (metabolic activity growth). Cells in one plate were measured every 24 h for 6 days. The growth amount was determined as the relative absorbance.

For Migration assay, 1×10^6 cells were plated in the trans-well chamber (8 μm, Millipore) with serum-free medium, then inserted into 24-well plate containing 10% FBS in medium, and incubated for 15–20 hours. The inserts were fixed with methanol and HE stained, cells that did not migrate were removed. The insert membranes were scanned and analyzed using NIH image software (https://imagej.nih.gov/ij/), and the cell density is measured as pixel intensity.

For Invasion assay, the same trans-well chambers were first coated with Matrigel (BD Bioscience), other steps were the same as described in migration assay.

Cytotoxicity assay

Cells were plated at 10,000–20,000 cells/well in 96-well plates, next day cells (confluence ~70%) were treated with increasing concentrations of cisplatin or paclitaxel (Sigma) for 72 h. The drug concentrations were listed in *Table S5*. Cell viability was measured by MTT assay kit (TIANGAN) following the manufacturer's instruction. The percentage of cell survival was defined as the relative absorbance of treated versus untreated wells. All assays were performed in triplicate.

Data analysis

Statistical significance was evaluated by Student-t test, Chi-square test or Fischer's exact test using SPSS 17.0 (SPSS Institute, Chicago, IL, USA). $P<0.05$ (two-sided P value) was considered to be significant.

Results

Lung cancer patients affected by family history and indoor air pollution: younger age and later stage at diagnosis

In total 277 subjects, 185 reported indoor air pollution from coal use, 92 were negative for coal burning exposure, and 51 were classified having FLC. The characteristics were shown in *Table 1*, *Table S1* and *Figure 1*. FLC showed

strong association with early-onset (P<0.01) (*Figure 1A*) and later stage (P<0.01). Indoor air pollution was associated with younger age (P<0.01) (*Figure 1B*), AD histology type (P<0.05) and later stage (P<0.05). FLC subjects had an average age of 50 (range, 36–70) and majority (45 cases, 88.2%) had stage III–IV disease, oppositely the average age of non-FLC subjects was 55 (range, 30–82), only 48.2% (109/226) of them were in stage III–IV. Divided by indoor air pollution exposure, positive group had an average age at 52 (range, 32–76), with 60.0% (111/185) patients in stage III–IV, comparatively negative group's average age was at 57 (range, 30–82), with 46.7% (43/92) patients in stage III–IV.

Interestingly, evaluated by relative ratio of patient age, FLC group showed double peaks (44 *vs.* 53 years), sporadic group also had two peaks (53 *vs.* 63 years), but both were much later than their FLC counterpart (*Figure 1A*). If divided by indoor air pollution exposure, positive group had a clear peak around age 52, while the negative group showed a much flatten curve (*Figure 1B*).

In addition, patients' absolute neutrophil (ANC) and lymphocyte (ALC) count were also analyzed. Even not significant, FLC subjects tended to have slightly lower ANC (average: 3.70×10^9/L *vs.* 3.94×10^9/L) and ALC (average: 1.88×10^9/L *vs.* 2.01×10^9/L) than non-FLC subjects (*Figure 1C,D*). Indoor air pollution was associated with higher ALC value (P<0.05) (average: 2.05×10^9/L *vs* 1.87×10^9/L), but no obvious ANC value difference was found for indoor air pollution exposure (average: 3.90×10^9/L *vs.* 3.89×10^9/L) (*Figure 1E,F*).

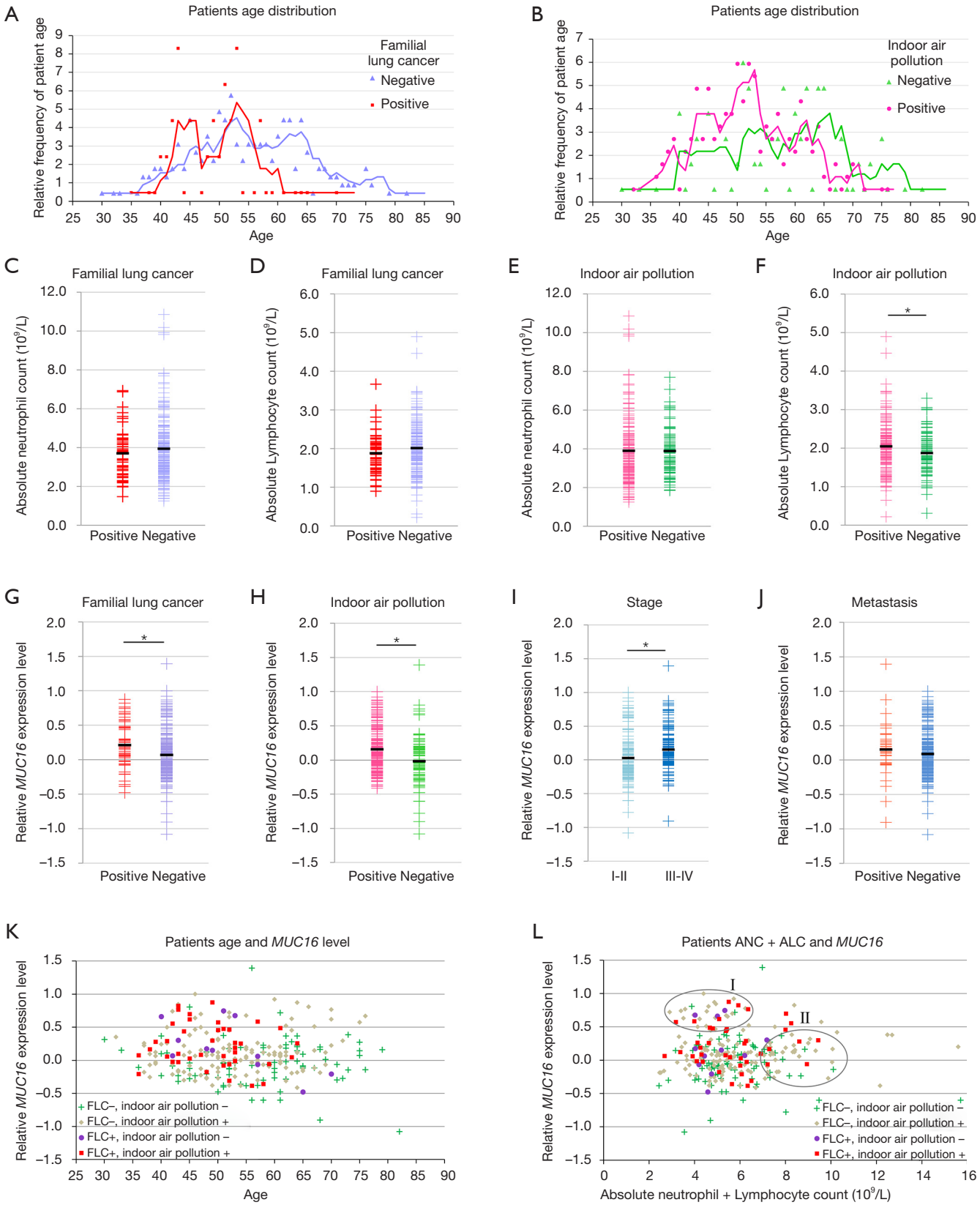
Overexpression of MUC16 in lung carcinoma show association with FLC and indoor air pollution

To investigate the significance of *MUC16* in our subject population, we examined *MUC16* expression in the above 277 cases by q-PCR. *MUC16* overexpression was significantly associated with FLC (P<0.05) (*Figure 1G*) and indoor air pollution (P<0.01) (*Figure 1H*), and also later stage (P<0.01) (*Figure 1I*) (*Table 2*). FLC subjects had nearly doubled rate of *MUC16* up-regulation (23.1% *vs.* 12.0%), and patients exposed to indoor air pollution were more likely to overexpress *MUC16* (75.0% *vs.* 55.6%), stage III–IV patients showed much higher ratio of *MUC16* overexpression (67.5% *vs.* 39.3%). Although not significant (P=0.0685), more metastasis events were observed in *MUC16*-upregulated group (18.1% *vs.* 10.3%) (*Figure 1J*). No apparent association was found for other parameters. However, one study suggested that *MUC16* overexpression rate was higher in AD compared to SCC (26). It may be explained by the special characteristics and also the size of our subject population.

All patients were further divided into four subgroups (*Table S1*) (*Figure 1K*). (I) Subjects affected by both FLC and indoor air pollution developed lung cancer much earlier, together with more frequently up-regulated *MUC16*. (II) FLC + and indoor air pollution – group had only a few individuals, mainly younger with up-regulated *MUC16*. (III) FLC—but indoor air pollution + group also had relatively more young patients, some with clearly increased *MUC16*. (IV) Subjects negative for both FLC and indoor air pollution seemed to have more even age distribution, with slightly less *MUC16* overexpression in young patients.

Previous reports indicated that *MUC16* could suppress immune response (16,30), so the relationship between *MUC16* level and patients' absolute neutrophil (ANC)/ lymphocyte (ALC) count was also analyzed. To better reflect immune reaction, the ANC, ALC values were analyzed separately and also combined (*Figure 1L,M,N*). In *Figure 1L,M*, two groups of subjects seemed drifting outside the main population, in *MUC16*-ALC (*Figure 1N*), subgroups positive for FLC or indoor air pollution had relatively higher *MUC16* (already confirmed), but no obvious unbalance was found in ALC distribution. Therefore, threshold values were set to isolate those individuals in *Figure 1L, M* for further study (*Table S6*). Group I: *MUC16* increase >2.8-fold (apparently elevated) and ANC $<4.8\times10^9$/L (average-to-low). Group II: *MUC16* level ≤2.8-fold increase (included those with less *MUC16* increase, no-change and down-regulated) and ANC $>5.5\times10^9$/L (higher-than-majority). The circles in *Figure 1L,M* covered major members to represent the group. Threshold standard is in the supplementary material (*Figure S1*).

Both groups were mostly composed of patients with either FLC or indoor air pollution or double positive. Group I included 38 individuals, with middle-to-low ANC + ALC value and clearly higher *MUC16*. Group II had 33 subjects, with higher-than-majority ANC + ALC value but lower *MUC16*. It indicated that patients with higher *MUC16* overexpression seemed to have a lower number of white blood cells, for ANC alone (*Figure 1M*) and ANC + ALC (*Figure 1L*). Intriguingly, group I had nearly doubled rate for FLC (31.6% *vs.* 15.2%) and was also significantly higher for in indoor air pollution exposure (P<0.05), suggesting the two factors were not only associated with *MUC16* up-regulation but also possibly suppressed immune reaction.



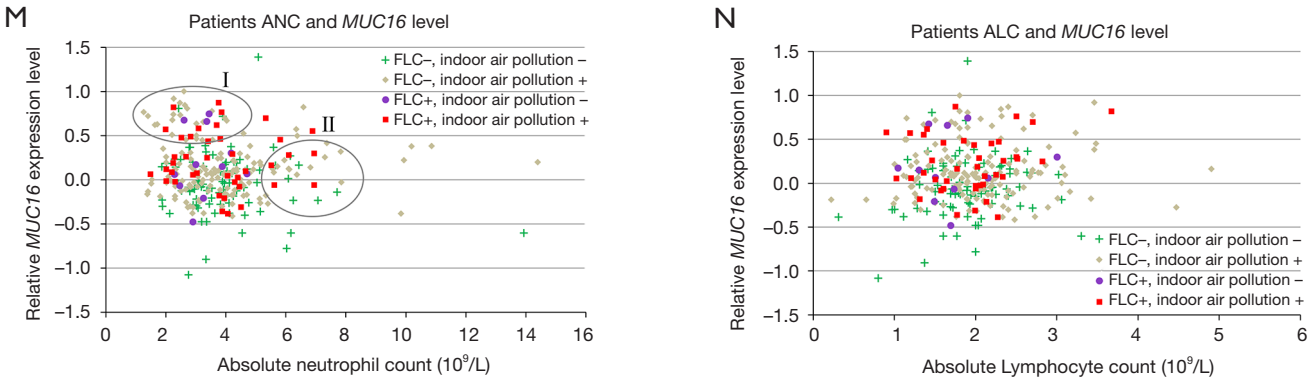


Figure 1 Clinical characteristics of NSCLC patients affected by familial lung cancer and indoor air pollution. (A) Familial lung cancer (FLC) patients developed lung cancer much earlier. FLC group had double peaks (44 *vs.* 53 years), and the sporadic group also had two peaks (53 *vs.* 63 years). (B) Patients exposed to indoor air pollution were also younger. Patients' absolute neutrophil (ANC) and lymphocyte (ALC) count were also analyzed. Even not significant, FLC subjects tended to have slightly lower ANC and ALC value (C,D); while indoor air pollution group had significantly higher ALC value (P<0.05) but not ANC value (E,F). *MUC16* overexpression was significantly associated with familial lung cancer (P<0.05) (G) and indoor air pollution (P<0.01) (H), and also later stage (P<0.01) (I). Although not significant, patients had metastasis were found with generally higher *MUC16* (J). Black bar in the middle is the average value of that group. (K) Divided into 4 subgroups: subjects affected by FLC and indoor air pollution developed lung cancer much earlier, together with more frequently up-regulated *MUC16*; FLC + and indoor air pollution – group had only a few individuals, mainly younger with up-regulated *MUC16*; FLC – but indoor air pollution + group also had relatively more younger patients, some with clearly increased *MUC16*; subjects negative for both FLC and indoor air pollution seemed to have more even age distribution, with slightly less *MUC16* overexpression in young patients. To better reflect immune reaction, the ANC, ALC values were analyzed separately and combined (L,M,N). In (L), (M), there were two groups of subjects drifting outside the main population, (L) group I had middle-to-low ANC + ALC value and clearly higher *MUC16*, while group II showed higher-than-majority ANC + ALC value but lower *MUC16*. The similar distribution could be found in (M). (N) Subgroups positive for FLC or indoor air pollution had relatively higher *MUC16* (already confirmed in previous results), but no obvious unbalance was found in ALC distribution. *P<0.05.

MUC16 gene knockout and overexpression in human lung cancer cell line

To examine the function of *MUC16* in lung cancer, we performed gene knockout (two different sgRNA targets) and overexpression (three different sgRNA targets) of *MUC16* in human lung cancer cell line H23 and H838. The *MUC16* up/down regulation could be detected at both mRNA (*Figure 2A,B*) and protein level (*Figure 2C,D,E,F*). Since there was no further cell selection after vector transfection, protein down-regulation was around 25% (H23) and 50% (H838), but *MUC16* up-regulation had apparent effect in both cell lines (~200%) (*Figure 2C,D*). Compared with ectopic overexpression of *MUC16*-Cter (the cytoplasmic tail region of *MUC16*) in several studies (21,22,24), the Lenti-CRISPR-dCas9 system can up-regulate *MUC16* whole protein level by directly activating its transcription from promoter region (28), thus better for examining the influence of *MUC16* overexpression.

High MUC16 level promote lung cancer cell proliferation and migration/invasion

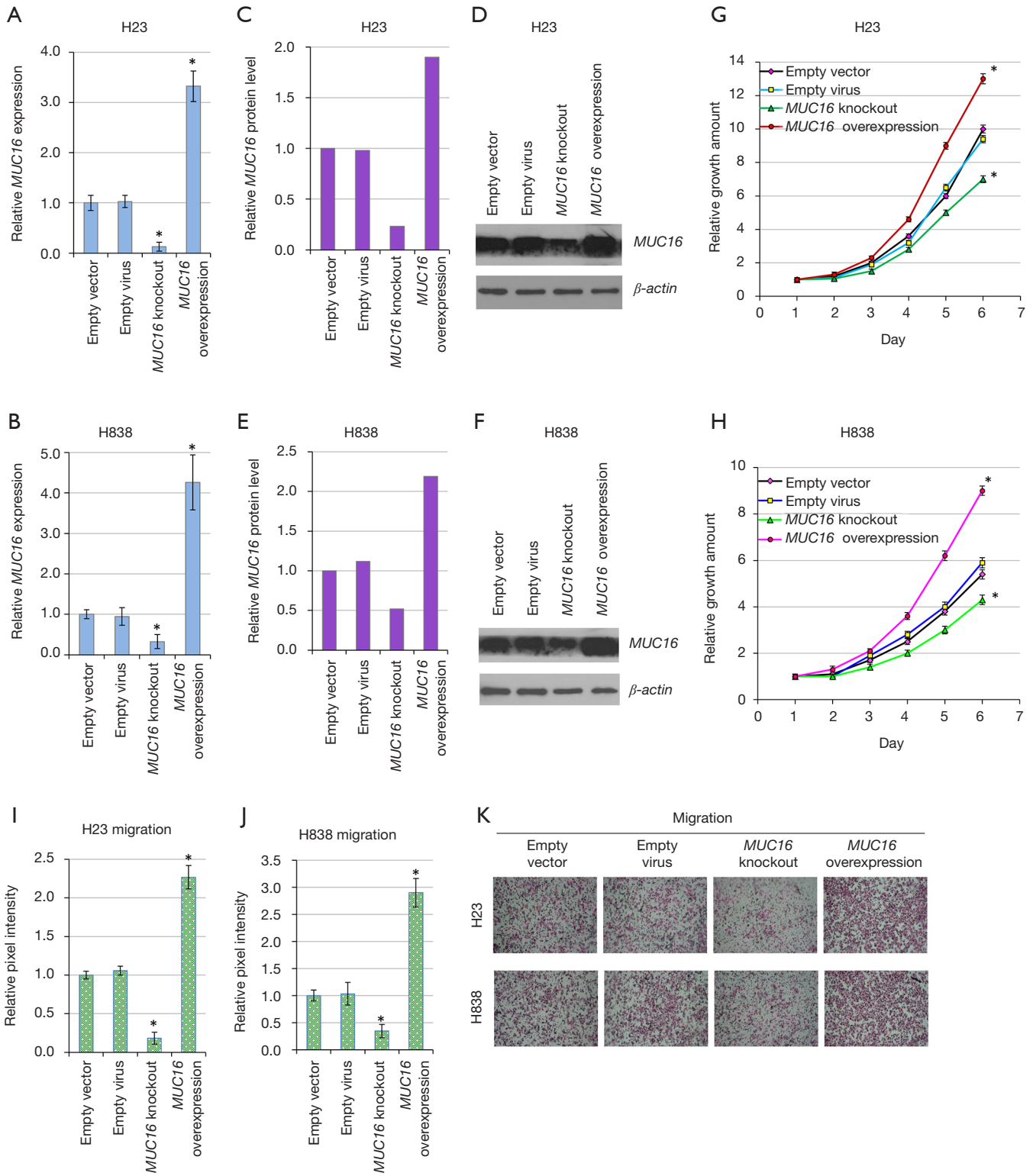
In growth kinetics assay, *MUC16* knockout cells had clearly decreased growth rate (P<0.05) compared to empty vector groups, on the other hand, *MUC16* overexpressed cells had significantly higher growth rate compared to empty virus infected cells (P<0.05) (*Figure 2G,H*). The growth rate increase seemed more apparent in H838 than H23, possibly reflecting cell line variation. These results indicated that *MUC16* could play positive role in lung cancer cell proliferation.

Migration assay showed that *MUC16* knockout cells had decreased migratory capacity (P<0.05), in the opposite, *MUC16* overexpressed cells had increased migratory capacity than empty virus infected cells (P<0.05) (*Figure*

Table 2 *MUC16* expression detected by q-PCR in 277 NSCLC patients

Variables	Total	*MUC16* expression, n (%)		
		Up-regulated	Down-regulated or no-change	P values[a]
Total number of patients	277	160	117	
Gender				0.412
Male	127	70 (43.8)	57 (48.7)	
Female	150	90 (56.2)	60 (51.3)	
Average age: 54 years (range, 30–82 years)				0.412
<54 years	150	90 (56.2)	60 (51.3)	
≥54 years	127	70 (43.8)	57 (48.7)	
Histology type				0.391
Adenocarcinoma	264	151 (94.4)	113 (96.6)	
Squamous cell carcinoma	13	9 (5.6)	4 (3.4)	
Metastasis				0.0685
Negative	236	131 (81.9)	105 (89.7)	
Positive	41	29 (18.1)	12 (10.3)	
Stage				0.000003
I–II	123	52 (32.5)	71 (60.7)	
III–IV	154	108 (67.5)	46 (39.3)	
Smoking history				0.979
Yes (current or ex-smoker)	85	49 (30.6)	36 (30.8)	
Never	192	111 (69.4)	81 (69.2)	
Indoor air pollution (solid fuel use)				0.000689
Present	185	120 (75.0)	65 (55.6)	
Absent	92	40 (25.0)	52 (44.4)	
Familial lung cancer				0.0179
Present	51	37 (23.1)	14 (12.0)	
Absent	226	123 (76.9)	103 (88.0)	
Average ANC: 3.89×10^9/L				0.551
<3.89×10^9/L	160	90 (56.3)	70 (59.8)	
≥3.89×10^9/L	117	70 (43.7)	47 (40.2)	
Average ALC: 1.99×10^9/L				0.875
<1.99×10^9/L	150	86 (53.8)	64 (54.7)	
≥1.99×10^9/L	127	74 (46.3)	53 (45.3)	
Average ANC + ALC: 5.88×10^9/L				0.356
<5.88×10^9/L	164	91 (56.9)	73 (62.4)	
≥5.88×10^9/L	113	69 (43.1)	44 (37.6)	

[a]P value calculated by chi-square test or Fisher's exact test, when there is at least one cell with expected count less than 5. NSCLC, non-small cell lung cancer; q-PCR, quantitative real-time polymerase chain reaction; ANC, absolute neutrophil count; ALC, absolute lymphocyte count.



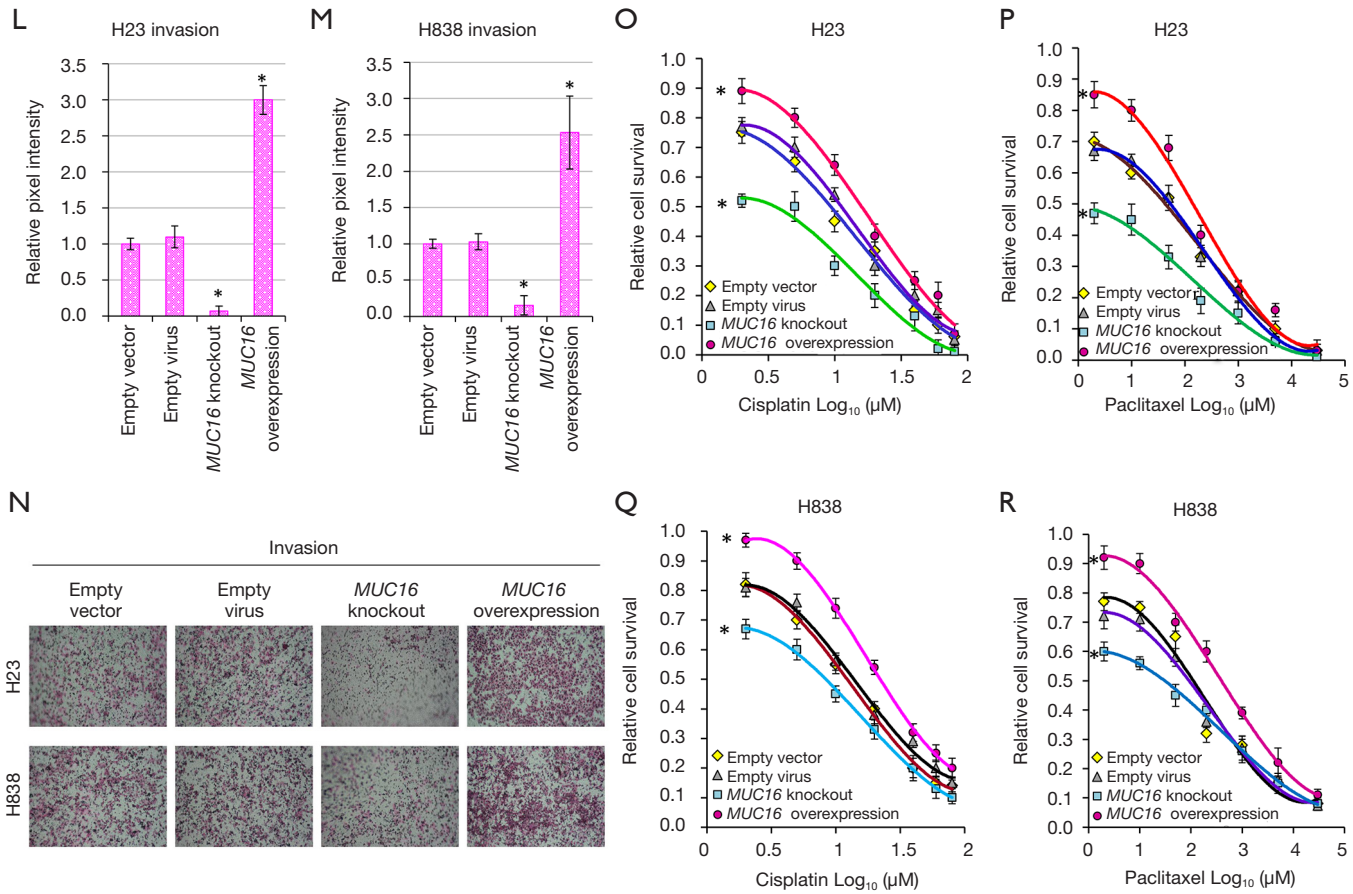


Figure 2 *MUC16* gene knockout and overexpression in human lung cancer cell lines and the influence on cell behaviors. To examine the function of *MUC16* in lung cancer, we performed *MUC16* knockout and overexpression in human lung cancer cell line H23 and H838. The up/down regulation could be detected at both mRNA (A,B) and protein level (C,D,E,F); there was no further cell selection after vector transfection, protein down-regulation was around 25% (H23) and 50% (H838), *MUC16* up-regulation was apparent in both cell lines (~200%). High *MUC16* level promoted lung cancer cell proliferation (G,H). *MUC16* knockout cells had clearly decreased growth rate compared to empty vector groups. *MUC16* overexpressed cells had significantly higher growth rate compared to empty virus infected cells (P<0.05). High *MUC16* level promoted lung cancer cell migration/invasion: *MUC16* knockout cells had decreased migratory capacity (P<0.05), while *MUC16* overexpressed cells had increased migratory capacity than empty virus infected cells (P<0.05) (I,J,K). *MUC16* knockout cells showed decreased invasion (P<0.05), while *MUC16* overexpressed cells revealed increased invasion capacity (P<0.05) (L,M,N). *MUC16* overexpression was associated with chemoresistance in lung cancer cells (O,P,Q,R). *MUC16* knockout cells were more sensitive to cisplatin and paclitaxel; no significant change was observed between empty vector and empty virus treated cells; *MUC16* overexpressed lung cancer cells were generally more resistant to the cytotoxic effects of cisplatin and paclitaxel. The cell line variations were also detected: in *MUC16* up-regulation, H838 had bigger cisplatin-resistance increase than H23; the paclitaxel-resistance increase also seemed higher for H838, only visible in low concentrations for H23; after *MUC16* down-regulation, H23 showed bigger drop in both cisplatin and paclitaxel resistance, while H838 had less resistance capacity decrease. *P<0.05.

2I,J,K). Compared to H23, the migratory capacity of H838 showed more increase and slightly less decrease. These results suggested that *MUC16* could contribute to the migration of lung cancer cells.

Similar results were found in invasion assay. *MUC16* knockout cells showed decreased invasion (P<0.05), while *MUC16* overexpressed cells revealed increased invasion capacity (P<0.05) (*Figure 2L,M,N*). Cell line variations could still be observed: H23 seemed to have bigger changes in invasion capacity for both *MUC16* up/down regulation.

These results reflected that *MUC16* might also boost the invasion capacity of lung cancer cells.

MUC16 overexpression is associated with chemoresistance in lung cancer cells

Overall, *MUC16* knockout cells were more sensitive to cisplatin and paclitaxel (*Figure 2O,P,Q,R*). In addition, no significant change was observed between empty vector and empty virus treated cells. On the other hand, *MUC16* overexpressed lung cancer cells were generally more resistant to the cytotoxic effects of cisplatin and paclitaxel. Taken together, these results indicated that *MUC16* could contribute to chemoresistance in lung cancer cells.

The cell line variations were also detected: after *MUC16* down-regulation, H23 showed bigger drop in both cisplatin and paclitaxel resistance (*Figure 2O,P*), while H838 had less resistance capacity decrease (*Figure 2Q,R*). In *MUC16* up-regulation, H838 had bigger cisplatin-resistance increase than H23, the paclitaxel-resistance increase was also higher for H838, only visible in low concentrations for H23.

Discussion

MUC16 has been studied in different cancers across populations. Our subject population has its signature characters: FLC history, indoor air pollution caused by coal use, and also the highest lung cancer incidence in the world among never smokers (3-6). All make it unique to study the complex interaction between genetic and environmental factors in lung cancer etiology. FLC showed strong association with early-onset ($P<0.01$) and later stage ($P<0.01$), which was consistent with previous findings (8,9,12). Indoor air pollution was associated with younger age ($P<0.01$), later stage ($P<0.05$) and AD histology type ($P<0.05$). Both factors were considered as crucial elements in lung cancer development (9-11). Interestingly, the double age peaks of FLC and sporadic group (*Figure 1A*) suggested multiple major contributors to lung cancer in our subject population, besides FLC and indoor air pollution.

We found that *MUC16* overexpression was associated with FLC ($P<0.05$), indoor air pollution ($P<0.01$), and later stage ($P<0.01$), furthermore, increased metastasis was observed in patients with up-regulated *MUC16* (18.1% *vs.* 10.3%). Similarly, many studies supported high *MUC16* was associated with increased metastasis and poor prognosis (17,22,24). Since *MUC16* functions as molecular barrier on epithelial cells, it would be reasonable to predict that compositions in polluted air could stimulate *MUC16* up-regulation as protective response. But in our study, no apparent correlation was found between high *MUC16* and smoking, possibly suggesting *MUC16* overexpression was a response to a wider spectrum of stimulants, and wasn't specific to cigarette ingredients. Furthermore, lung cancers in non-smokers were also different from those in smokers (9,10,31). Importantly, the mechanism underlining the association between FLC and elevated *MUC16* deserves further investigation. One study indicated that *MUC16* mutation was associated with tumor mutation load (32), and FLC patients could possibly carry larger tumor mutation load, since the susceptible elements in FLC subjects made them more vulnerable to mutation-inducing carcinogens. The mutation rate could also be varied for different genes in one individual, and evidence suggested that certain genes were more frequently mutated in FLC population (31). As a result, elevated *MUC16* might potentially be a unique feature to our subject population, like one molecular character of inherited lung cancer susceptibility in local residents.

To examine the function of *MUC16* in lung cancer, we carried gene knockout and overexpression in human lung cancer cell line H23 and H838. We found that high *MUC16* level promoted lung cancer cell proliferation, migration, invasion and also chemoresistance, additionally there were also variations among different cell lines. Our results were well supported by previous reports in different cancers (22-26). Some (22,23,26) found *MUC16* mediated *JAK2/STAT3/GR* signal pathway, and promoted cancer cell growth/migration through *TSPYL5*. Moreover, *MUC16* could induce resistance to chemotherapy drugs by up-regulating *TSPYL5*, which suppresses *p53* activity.

Beside its positive roles in cancer cells, *MUC16* also interferes with immune reaction. There were evidences that *MUC16* could suppress human innate immune responses by regulating NK cells and macrophages (16,30). *MUC16* can form aggregates with neutrophils, macrophages, and platelets, conferring protection to cancer cells during hematological dissemination (16). Intriguingly, we also found patients with more *MUC16* up-regulation seemed to have a lower number of white blood cells, especially neutrophils. Oppositely some subjects showed less *MUC16* could have much higher white blood cell count. It helped to explain that high *MUC16* meant poor prognosis. On the contrary, presence of *MUC16* neo-antigen-specific T cells in cancer patients suggested that *MUC16* could serve as a potential target for cancer immunotherapy and

radioimmunotherapy (16,18,19), which might possibly benefit our subject population.

Conclusions

MUC16 can play crucial roles in lung cancer pathogenesis, progression and chemoresistance. Interestingly, its association with FLC and indoor air pollution highlights the complexity of lung cancer etiology. Our findings provide useful information to study the intricate and dynamic interaction between environmental carcinogens and population genetic background.

Acknowledgments

Funding: This work was supported by National Natural Science Foundation of China (No: 81702274); Yunnan Applied Basic Research Projects-Union Foundation [No: 2017FE468 (-159), 2015FB069, 2017FE467 (-0187), 2017FE468 (-214), 2017FA039]; Internal Organization Research Projects of Yunnan Cancer Hospital (No: 2017NS198, 2017NS199); Yunnan Health Training Project of High Level Talent (D-2017012,D-201641); Doctor Research Foundation of Yunnan Cancer Hospital (No: BSKY201705).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The study was approved by the Ethical Committees of Yunnan Cancer Hospital (No. KY2019.57). All patients provided informed consent. The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. AMC. Cancer Facts and Figures. Atlanta: American Cancer Society; 2014.
2. Sugimura H, Yang P. Long-term survivorship in lung cancer: a review. Chest 2006;129:1088-97.
3. Barone-Adesi F, Chapman RS, Silverman DT, et al. Risk of lung cancer associated with domestic use of coal in Xuanwei, China: retrospective cohort study. BMJ 2012;345:e5414.
4. Chapman RS, Mumford JL, Harris DB, et al. The epidemiology of lung cancer in Xuan Wei, China: current progress, issues, and research strategies. Arch Environ Health 1988;43:180-5.
5. Mumford JL, Chapman RS, Harris DB, et al. Indoor air exposure to coal and wood combustion emissions associated with a high lung cancer rate in Xuan Wei, China. Environment International 1989;15:315-20.
6. Mumford JL, He XZ, Chapman RS, et al. Lung cancer and indoor air pollution in Xuan Wei, China. Science 1987;235:217-20.
7. Coté ML, Kardia SLR, Wenzlaff AS, et al. Risk of lung cancer among white and black relatives of individuals with early-onset lung cancer. JAMA 2005;293:3036-42.
8. Coté ML, Liu M, Bonassi S, et al. Increased risk of lung cancer in individuals with a family history of the disease: A pooled analysis from the International Lung Cancer Consortium. Eur J Cancer 2012;48:1957-68.
9. Lin KF, Wu HF, Huang WC, et al. Propensity score analysis of lung cancer risk in a population with high prevalence of non-smoking related lung cancer. BMC Pulm Med 2017;17:120.
10. Lin H, Huang YS, Yan HH, et al. A family history of cancer and lung cancer risk in never-smokers: A clinic-based case-control study. Lung Cancer 2015;89:94-8.
11. Karp I, Sylvestre MP, Abrahamowicz M, et al. Bridging the etiologic and prognostic outlooks in individualized assessment of absolute risk of an illness: Application in lung cancer. Eur J Epidemiol 2016;31:1091-9.
12. Musolf AM, Simpson CL, de Andrade M, et al. Familial Lung Cancer: A Brief History from the Earliest Work to the Most Recent Studies. Genes (Basel) 2017;8. doi: 10.3390/genes8010036.
13. Das S, Batra SK. Understanding the Unique Attributes of MUC16 (CA125): Potential Implications in Targeted Therapy. Cancer Res 2015;75:4669-74.
14. Hollingsworth MA, Swanson BJ. Mucins in cancer: protection and control of the cell surface. Nat Rev Cancer 2004;4:45-60.
15. Kufe DW. Mucins in cancer: function, prognosis and therapy. Nat Rev Cancer 2009;9:874-85.
16. Bhatia R, Gautam SK, Cannon A, et al. Cancer-associated mucins: role in immune modulation and metastasis. Cancer Metastasis Rev 2019;38:223-36.
17. Jonckheere N, van Seuningen I. Integrative analysis of the cancer genome atlas and cancer cell lines encyclopedia large-scale genomic databases: MUC4/MUC16/MUC20 signature is associated with poor survival in human

carcinomas. J Transl Med 2018;16:259.
18. Aithal A, Rauth S, Kshirsagar P, et al. MUC16 as a novel target for cancer therapy. Expert Opin Ther Targets 2018;22:675-86.
19. Rao TD, Fernández-Tejada A, Axelrod A, et al. Antibodies Against Specific MUC16 Glycosylation Sites Inhibit Ovarian Cancer Growth. ACS Chem Biol 2017;12:2085-96.
20. Kim N, Hong Y, Kwon D, et al. Somatic mutaome profile in human cancer tissues. Genomics Inform 2013;11:239-44.
21. Akita K, Tanaka M, Tanida S, et al. CA125/MUC16 interacts with Src family kinases, and over-expression of its C-terminal fragment in human epithelial cancer cells reduces cell-cell adhesion. Eur J Cell Biol 2013;92:257-63.
22. Das S, Rachagani S, Torres-Gonzalez MP, et al. Carboxyl-terminal domain of MUC16 imparts tumorigenic and metastatic functions through nuclear translocation of JAK2 to pancreatic cancer cells. Oncotarget 2015;6:5772-87.
23. Lakshmanan I, Ponnusamy MP, Das S, et al. MUC16 induced rapid G2/M transition via interactions with JAK2 for increased proliferation and anti-apoptosis in breast cancer cells. Oncogene 2012;31:805-17.
24. Thériault C, Pinard M, Comamala M, et al. MUC16 (CA125) regulates epithelial ovarian cancer cell growth, tumorigenesis and metastasis. Gynecol Oncol 2011;121:434-43.
25. Boivin M, Lane D, Piché A, et al. CA125 (MUC16) tumor antigen selectively modulates the sensitivity of ovarian cancer cells to genotoxic drug-induced apoptosis. Gynecol Oncol 2009;115:407-13.
26. Lakshmanan I, Salfity S, Seshacharyulu P, et al. MUC16 Regulates TSPYL5 for Lung Cancer Cell Growth and Chemoresistance by Suppressing p53. Clin Cancer Res 2017;23:3906-17.
27. Ran FA, Hsu PD, Wright J, et al. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 2013;8:2281-308.
28. Konermann S, Brigham MD, Trevino AE, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 2015;517:583-8.
29. Naito Y, Hino K, Bono H, et al. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics 2015;31:1120-3.
30. Felder M, Kapur A, Rakhmilevich AL, et al. MUC16 suppresses human and murine innate immune responses. Gynecol Oncol 2019;152:618-28.
31. Gaughan EM, Cryer SK, Yeap BY, et al. Family history of lung cancer in never smokers with non-small-cell lung cancer and its association with tumors harboring EGFR mutations. Lung Cancer 2013;79:193-7.
32. Li X, Pasche B, Zhang W, et al. Association of MUC16 mutation with tumor mutation load and outcomes in patients with gastric cancer. JAMA Oncol 2018;4:1691-8.

Cite this article as: Chen Y, Huang Y, Kanwal M, Li G, Yang J, Niu H, Li Z, Ding X. *MUC16* in non-small cell lung cancer patients affected by familial lung cancer and indoor air pollution: clinical characteristics and cell behaviors. Transl Lung Cancer Res 2019;8(4):476-488. doi: 10.21037/tlcr.2019.07.10

Supplementary

Subject population background

Our subject population were recruited from China's Yunnan Province, certain region here reported some of the highest lung cancer rates in the world, such as Xuanwei/Fuyuan (3-6). These areas have long been focus of lung cancer studies, including epidemiology, molecular or clinical research. Interestingly, the subject population has two characters: familial lung cancer (FLC) and indoor air pollution caused by coal combustion, because local residents use coal for cooking and heating for generations (3-6).

Present study was designed to investigate the clinical significance of *MUC16* in NSCLC patients affected by familial lung cancer (FLC) and indoor air pollution caused by coal use in Yunnan, China. Subjects were selected by the following criteria: (I) the case population was mainly composed of residents from Xuanwei/Fuyuan region of Yunnan Province, who primarily use coal for heating or cooking for more than 10 years. (II) The control subjects were patients from other areas in the same province, who reported no history of occupational or domestic coal use. In total, 185 cases and 92 controls were enrolled. (III) Subjects with familial lung cancer were defined as individuals with three or more first-degree relatives affected by lung cancer. There were 51 patients classified as having familial lung cancer. All the information was based on self-report and confirmed by personal medical records.

The subject population can be further divided into 4 subgroups: FLC+, indoor air pollution+; FLC+, indoor air pollution–; FLC–, indoor air pollution+; FLC–, indoor air pollution–. Both characters, one genetic and one environmental were analyzed in our study. Clinicopathologic data were shown in *Table 1* and *Table S1*.

Table S1 Clinical characteristics of 277 NSCLC patients divided into four subgroups

Variables	Total	Subgroup, n (%)			
		1	2	3	4
Familial lung cancer (FLC)		+	+	–	–
Indoor air pollution		+	–	+	–
Total number of patients	277	40	11	145	81
Gender					
Male	127	18 (45.0)	2 (18.2)	68 (46.9)	39 (48.1)
Female	150	22 (55.0)	9 (81.8)	77 (53.1)	42 (51.9)
Average age: 54 years (range, 30–82 years)					
<54 years	150	32 (80.0)	7 (63.6)	81 (55.9)	30 (37.0)
≥54 years	127	8 (20.0)	4 (36.4)	64 (44.1)	51 (63.0)
Histology type					
Adenocarcinoma	264	40 (100)	11 (100)	140 (96.6)	73 (90.1)
Squamous cell carcinoma	13	0 (0.0)	0 (0.0)	5 (3.4)	8 (9.9)
Metastasis					
Negative	236	35 (87.5)	7 (63.6)	128 (88.3)	66 (81.5)
Positive	41	5 (12.5)	4 (36.4)	17 (11.7)	15 (18.5)
Stage					
I–II	123	2 (5.0)	4 (36.4)	72 (49.7)	45 (55.6)
III–IV	154	38 (95.0)	7 (63.6)	73 (50.3)	36 (44.4)
Smoking history					
Yes (current or ex-smoker)	85	11 (27.5)	1 (9.1)	50 (34.5)	23 (28.4)
Never	192	29 (72.5)	10 (90.9)	95 (65.5)	58 (71.6)
MUC16 expression					
Up-regulated	160	29 (72.5)	8 (72.7)	91 (62.8)	32 (39.5)
Down-regulated or no-change	117	11 (27.5)	3 (27.3)	54 (37.2)	49 (60.5)

NSCLC, non-small cell lung cancer; +, positive; –, negative.

Table S2 Addgene plasmid for *MUC16* gene knockout and overexpression

Gene	Catalog #	Plasmid name
MUC16 knockout	Plasmid #62988	pSpCas9(BB)-2A-Puro(PX459)V2.0
MUC16 overexpression	Plasmid #61425	lenti dCAS-VP64_Blast
	Plasmid #61426	lenti MS2-P65-HSF1_Hygro
	Plasmid #61427	lenti sgRNA(MS2)_zeo backbone

Table S3 sgRNA sequence for *MUC16* gene knockout and overexpression

Gene	Seq NO.	sgRNA sequence
MUC16 knockout	sgRNA-1	ATGCAGAAACAAGTGCCGAA
	sgRNA-2	TCACTCCACGAGTGGAAAGG
MUC16 overexpression	sgRNA-1	GGTTTCTAAGACATCACACA
	sgRNA-2	AGGGAAGAATTGAGATCATC
	sgRNA-3	TGAGATTTGGTCTTCTCAAT

The sgRNA were designed based on GRCh37/hg19, using CRISPRdirect (http://crispr.dbcls.jp/).

Table S4 Specificity information for *MUC16* sgRNA sequence

Gene	Seq NO.	Sequence information		Number of site	
		GC (%)	Tm (℃)	hit_20mer	hit_12mer*
MUC16 knockout	sgRNA-1	45.00	70.78	1	2
	sgRNA-2	55.00	73.87	1	1
MUC16 overexpression	sgRNA-1	40.00	66.43	1	19
	sgRNA-2	40.00	66.41	1	40
	sgRNA-3	35.00	64.56	1	17

Specificity information was provided by CRISPRdirect, based on Human (Homo sapiens) genome, GRCh37/hg19. hit_20mer, number of site perfect match with the sgRNA sequence; hit_12mer, number of site match with 12 bases at 3' region of sgRNA sequence. *Because 200bp upstream the transcription start site is highly preferred for activation (28), sgRNA available in this short region can be specific in hit_20mer, but may not very low in hit_12mer.

Vectors construction for MUC16 gene knockout and overexpression

CRISPR-Cas9 vectors were constructed for *MUC16* gene knockout as described in (27). In order to effectively knockout *MUC16* gene, two sgRNA were combined to target the first exon of *MUC16* (PX459-*MUC16*-sgRNA-1 and PX459-*MUC16*-sgRNA-2). Lenti-CRISPR-dCas9 system was used for *MUC16* overexpression, three sgRNA were used simultaneously to increase activation efficiency. The vector construction and lentivirus packaging followed protocols in (28). PX459 and Lenti-CRISPR-dCas9 system were gift from Feng Zhang (*Table S2*, Addgene plasmid #62988; #61425, #61426, #61427). The sgRNA sequences were designed using CRISPRdirect (29) (http://crispr.dbcls.jp/) and listed in *Tables S3*,*S4*.

The sgRNA site on MUC16 genome

The *MUC16* genome sequence showed here includes 350 bp upstream the transcription start site and part of the first exon. 200bp upstream is preferred for activation (28). Upstream sequence is in lowercase and the 1st exon is in uppercase. *MUC16* knockout sgRNA is marked in yellow and *MUC16* overexpression sgRNA is marked in green.

>hg19_ensGene_ENST00000397910 range=chr19 strand=- repeatMasking=none
ccatgttggtcaggctgatctcaaactcccaacctcaggtgatcctccca
cctcacctcccaaagtgctgggattacaggcatgagccaccgtgcccagc
ctggttcctggtttctaagacatcacacacacacacacacacacacac
acactcacacactcagagagagagagagagagaggatcattaagacatga
tacactaagaaattctattctgcagacactgagaatccgttaaaaagttt
gaagggaagaattgagatcatcaggtgtttatttgaggaaattgtctgtg
gttgaactatcctttccttttctctccctgagatttggtcttctcaattag
AAGCGTTGCACAATTCCCCCAACCTCCATACATACGGCAGCTCTTCTAGA
CACAGGTTTTCCCAGGTCAAATGCGGGGACCCCAGCCATATCTCCCACCC

TGAGAAATTTTGGAGTTTCAGGGAGCTCAGAAGCTCTGCAGAGGCCACCC
TCTCTGAGGGGATTCTTCTTAGACCTCCATCCAGAGGCAAATGTTGACCT
GTCCATGCTGAAACCCTCAGGCCTTCCTGGGTCATCTTCTCCCACCCGCT
CCTTGATGACAGGGAGCAGGAGCACTAAAGCCACACCAGAAATGGATTCA
GGACTGACAGGAGCCACCTTGTCACCTAAGACATCTACAGGTGCAATCGT
GGTGACAGAACATACTCTGCCCTTTACTTCCCCAGATAAGACCTTGGCCA
GTCCTACATCTTCGGTTGTGGGAAGAACCACCCAGTCTTTGGGGGTGATG
TCCTCTGCTCTCCCTGAGTCAACCTCTAGAGGAATGACACACTCCGAGCA
AAGAACCAGCCCATCGCTGAGTCCCCAGGTCAATGGAACTCCCTCTAGGA
ACTACCCTGCTACAAGCATGGTTTCAGGATTGAGTTCCCCAAGGACCAGG
ACCAGTTCCACAGAAGGAAATTTTACCAAAGAAGCATCTACATACACACT
CACTGTAGAGACCACAAGTGGCCCAGTCACTGAGAAGTACACAGTCCCCA
CTGAGACCTCAACAACTGAAGGTGACAGCACAGAGACCCCCTGGGACACA
AGATATATTCCTGTAAAAATCACATCTCCAATGAAAACATTTGCAGATTC
AACTGCATCCAAGGAAAATGCCCCAGTGTCTATGACTCCAGCTGAGACCA
CAGTTACTGACTCACATACTCCAGGAAGGACAAACCCATCATTTGGGACA
CTTTATTCTTCCTTCCTTGACCTATCACCTAAAGGGACCCCAAATTCCAG
AGGTGAAACAAGCCTGGAACTGATTCTATCAACCACTGGATATCCCTTCT
CCTCTCCTGAACCTGGCTCTGCAGGACACAGCAGAATAAGTACCAGTGCG
CCTTTGTCATCATCTGCTTCAGTTCTCGATAATAAAATATCAGAGACCAG
CATATTCTCAGGCCAGAGTCTCACCTCCCCTCTGTCTCCTGGGGTGCCCG
AGGCCAGAGCCAGCACAATGCCCAACTCAGCTATCCCTTTTTCCATGACA
CTAAGCAATGCAGAAACAAGTGCCGAAAGGGTCAGAAGCACAATTTCCTC
TCTGGGGACTCCATCAATATCCACAAAGCAGACAGCAGAGACTATCCTTA
CCTTCCATGCCTTCGCTGAGACCATGGATATACCCAGCACCCACATAGCC
AAGACTTTGGCTTCAGAATGGTTGGGAAGTCCAGGTACCCTTGGTGGCAC
CAGCACTTCAGCGCTGACAACCACATCTCCATCTACCACTTTAGTCTCAG
AGGAGACCAACACCCATTCACTCCACGAGTGGAAAGGAAACAGAAGGAACT
TTGAATACATCTATGACTCCACTTGAGACCTCTGCTCCTGGAGAAGAGTC
CGAAATGACTGCCACCTTGGTCCCCACTCTAGGTTTTACAACTCTTGACA
GCAAGATCAGAAGTCCATCTCAGGTCTCTTCATCCCACCCAACAAGAGAG
CTCAGAACCACAGGCAGCACCTCTGGGAGGCAGAGTTCCAGCACAGCTGC
CCACGGGAGCTCTGACATCCTGAGGGCAACCACTTCCAGCACCTCAAAAG
CATCATCATGGACCAGTGAAAGCACAGCTCAGCAATTTAGTGAACCCCAG
CACACACAGTGGGTGGAGACAAGTCCTAGCATGAAAACAGAGAGACCCCC
AGCATCAACCAGTGTGGCAGCCCCTATCACCACTTCTGTTCCCTCAGTGG
TCTCTGGCTTCACCACCCTGAAGACCAGCTCCACAAAAGGGATTTGGCTT

Oligo annealing and cloning into backbone vectors

MUC16 knockout

(I) Digest 1ug of pX459 with BbsI for 30 min at 37 ℃
- 1 µg pX459
- 1 µL FastDigest BbsI (Fermentas)
- 2 µL 10× FastDigest Buffer
- X µL ddH_2O
- 20 µL in total

(II) Gel purify digested pX459 using Gel Extraction Kit (TIANGAN)

(III)Anneal each pair of oligos
1 μL oligo forward (100 mM)
1 μL oligo reverse (100 mM)
2 μL 5× annealing butter (TIANGAN)
6 μL ddH_2O
10 μL in total
Anneal in a thermocycler using the following parameters
95 ℃ 5 min and then ramp down to 25 ℃ at 5 ℃/min
(IV)Ligation reaction
X μL BbsI digested pX459 from step 2 (50 ng)
1 μL annealed oligo from step 3
1 μL 10× ligation Buffer
1 μL T4 Ligase (Fermentas)
X μL ddH_2O
10 μL in total
Incubate reaction at 22 ℃ for 40 min
(V) Transformation into Stbl3 bacteria

MUC16 overexpression

(I) Digest and 5 μg of lenti sgRNA zeo backbone with BsmBI for 60 min at 37 ℃
5 μg lenti sgRNA zeo backbone
3 μL FastDigest BsmBI (Fermentas)
6 μL 10× FastDigest Buffer
0.6 μL 100 mM DTT (freshly prepared)
X μL ddH_2O
60 μL in total
(II) Gel purify digested plasmid using Gel Extraction Kit (TIANGAN)
(III)Anneal each pair of oligos
1 μL oligo forward (100 mM)
1 μL oligo reverse (100 mM)
2 μL 5× annealing butter (TIANGAN)
6 μL ddH_2O
10 μL in total
Anneal in a thermocycler using the following parameters:
95 ℃ 5 min and then ramp down to 25 ℃ at 5 ℃/min
(IV)Ligation reaction
X μL digested sgRNA zeo backbone from step 2 (50 ng)
1 μL annealed oligo from step 3
1 μL 10× ligation Buffer
1 μL T4 Ligase (Fermentas)
X μL ddH_2O
10 μL in total
Incubate reaction at 22 ℃ for 60 min
(V) Transformation into Stbl3 bacteria

Plasmid transfection and lentivirus infection

MUC16-knockout vectors (PX459-*MUC16*-sgRNA-1; PX459-*MUC16*-sgRNA-2) were transfected into target cells using

Lipofectamine 2000 (Thermo Fisher, waltham, MA, USA) according to the manufacturer's instruction, empty vector was used as control. Lentivirus packing and infection were carried out as mentioned in (28) with empty virus as control.

Present work first studied the immediate effect of *MUC16* change on a population of cancer cells, and no clone selection was carried to remove *MUC16*-unchanged cells. Since drug selection would purify subpopulations featured with up/down-regulated *MUC16*, but other genes level may also change during subpopulation selection, when compared with the original cell population. Therefore, transfection and infection were performed freshly for each cell behavior experiment, *MUC16* levels were monitored by q-PCR, cell populations with more than 60% *MUC16* decrease and more than 3 times *MUC16* increase were immediately used for the behavior experiments. Furthermore, cancer cells show heterogeneity in patients' tumor as well as cultured cells, if *MUC16* change in a subpopulation could influence the behaviors of the whole cell population, it still provides meaningful information.

Cytotoxicity assay

We treated cells with increasing concentrations of cisplatin or paclitaxel for 72 h. The drug concentrations were listed in *Table S5*.

Table S5 Drug concentrations for cytotoxicity assay

Chemotherapy drug	Concentration
Cisplatin (μM)	2, 5, 10, 20, 40, 60, 80
Paclitaxel (nM)	2, 10, 50, 200, 1,000, 5,000, 30,000

Results part

Threshold setting standard based on MUC16-ANC

Two groups of subjects seemed drifting outside the main population (*Figure 1L,M*) (*Table S6*), threshold values were set to isolate those individuals for further study (based on *MUC16*-ANC). Group I: *MUC16* increase >2.8-fold (apparently elevated) and ANC <4.8×10^9/L (average-to-low). Group II: *MUC16* level ≤2.8-fold increase (included those with less *MUC16* increase, no-change and down-regulated) and ANC >5.5×10^9/L (higher-than-majority). The circles in *Figure 1L,M* covered major members to represent the group.

In FLC-ANC and FLC-*MUC16* (*Figure 1C,G*), FLC+ subjects were clearly divided by a gap (black arrow), suggesting potential subgroups. We used the upper limit (2.8-fold) of *MUC16* gap to separate group I and II as the first step (in *Figure 1G*, lg2.8=0.45), then we used the upper (5.5×10^9/L) and the lower (4.8×10^9/L) limit of ANC gap to separate group I and II further apart, so as to observe bigger difference.

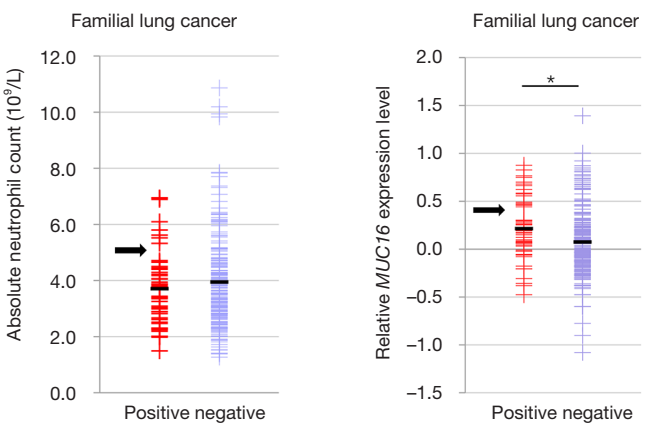


Figure S1 Threshold setting standard based on *MUC16*-ANC. ANC, absolute neutrophil count. The black arrow is pointing at the "gap" in FLC+ patients, and the gap separated FLC+ patients into two subgroups.

Table S6 Clinical characteristics of group I and group II which have special *MUC16*-white blood cell combination

Characters	Total	Group, n (%)		
		I	II	P values[a]
Total number of patients	277	38	33	
MUC16 level (fold)	0.125–24.6	>2.8	≤2.8	
ANC (10^9/L)	1.26–14.39	<4.8	>5.5	
ALC (10^9/L)	0.22–4.9	0.9–3.67	1.16–3.3	
ANC+ALC (10^9/L)	2.44–15.7	2.9–6.38	7.06–15.7	
Gender				0.230
Male	127	12 (31.6)	15 (45.5)	
Female	150	26 (68.4)	18 (54.5)	
Average age: 54 years (range, 30–82 years)				0.462
<54 years	150	24 (63.2)	18 (54.5)	
≥54 years	127	14 (36.8)	15 (45.5)	
Histology type				1.00
Adenocarcinoma	264	37 (97.4)	32 (97.0)	
Squamous cell carcinoma	13	1 (2.6)	1 (3.0)	
Metastasis				0.497
Negative	236	34 (89.5)	27 (81.8)	
Positive	41	4 (10.5)	6 (18.2)	
Stage				0.899
I–II	123	19 (50.0)	16 (48.5)	
III–IV	154	19 (50.0)	17 (51.5)	
Smoking history				0.150
Yes (current or ex-smoker)	85	7 (18.4)	11 (33.3)	
Never	192	31 (81.6)	22 (66.7)	
Indoor air pollution (solid fuel use)				0.042
Positive	185	33 (86.8)	22 (66.7)	
Negative	92	5 (13.2)	11 (33.3)	
Familial lung cancer				0.106
Positive	51	12 (31.6)	5 (15.2)	
Negative	226	26 (68.4)	28 (84.8)	

[a]P value calculated by chi-square test or Fisher's exact test, when there is at least one cell with expected count less than 5. ANC, absolute neutrophil count; ALC, absolute lymphocyte count.