Thoracic Cancer

Open Access

Thoracic Cancer ISSN 1759-7706

ORIGINAL ARTICLE

Lung cancer family history and exposure to occupational/domestic coal combustion contribute to variations in clinicopathologic features and gene fusion patterns in non-small cell lung cancer

Ying Chen[1†], Guangjian Li[1†], Yujie Lei[1], Kaiyun Yang[1], Huatao Niu[2], Jie Zhao[1], Rui He[1], Huanqi Ning[1], Qiubo Huang[1], Qinghua Zhou[3] & Yunchao Huang[1]

1 Department of Thoracic Surgery I, The Third Affiliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center, The International Cooperation Key Laboratory of Regional Tumor in High Altitude Area, Kunming, China

2 Department of Neurosurgery, The Third Affiliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, Yunnan Cancer Center, The International Cooperation Key Laboratory of Regional Tumor in High Altitude Area, Kunming, China

3 Lung Cancer Center, Institute, West China Hospital, Sichuan University, Chengdu, China

Keywords

Cancer susceptibility; coal combustion; lung cancer family history; non-small cell lung cancer; occupational exposure.

Correspondence

Yunchao Huang, Department of Thoracic Surgery, The Third Affiliated Hospital of Kunming Medical University/Yunnan Cancer Hospital, No.519, Kunzhou Road, Kunming 650106, Yunnan, China.
Tel: +86 137 0847 6546
Fax: +86 871 6818 5656
Email: huangych2001@aliyun.com

Qinghua Zhou, Lung Cancer Center/Institute, West China Hospital, Sichuan University, No.37,Guo Xue Xiang, Wuhuo District, Chengdu610041,Sichuan, China.
Tel: +86 189 8060 6202
Fax: +86 22 6036 3013
Email: zhouqh135@163.com

†Ying Chen and Guangjian Li contributed equally to this work.

doi: 10.1111/1759-7714.12987

Thoracic Cancer **10** (2019) 695–707

Abstract

Background: Both genetic and environmental factors contribute to the development of cancer and its mutant spectrum. Lung cancer has familial aggregation. Lung cancer caused by non-tobacco factors has unique pathological and molecular characteristics. The interaction between genetic lung cancer susceptibility and carcinogens from coal burning remains complex and understudied.

Methods: We selected 410 non-small cell lung cancer (NSCLC) patients with a family history of lung cancer (FLC) and exposure to coal combustion between 2014 and 2017. Clinicopathologic parameters were analyzed. Reverse transcription-PCR was performed to detect *ALK*, *ROS1*, *RET*, and *NTRK1* rearrangement.

Results: Among the 410 NSCLC patients, 192 had FLC and 204 (49.8%) were exposed to occupational or domestic coal combustion. FLC patients had the same characteristics regardless of gender and coal exposure: younger age, high female ratio, adenocarcinoma, increased metastasis, later stage at diagnosis, and higher frequency of gene fusion. Sixty-seven patients (16.3%) had gene rearrangement: 51 (12.4%) harbored *EML4-ALK* fusions and 16 *ROS1* fusions (3.9%). The highest gene fusion rate (35.1%, 33/94) occurred in patients with both FLC and high tobacco and coal exposure. *ALK* fusions and total gene rearrangement were closely associated with women, never smokers, younger age, FLC, and coal exposure.

Conclusion: FLC and exposure to coal combustion have an important impact on the clinicopathological characteristics and gene fusion mode of NSCLC, particularly in cases of higher levels of carcinogens, and genetic susceptibility has a greater impact. Our findings may help evaluate the effect of FLC and coal exposure on the pathogenesis of lung cancer.

Introduction

Lung cancer has been the most common cancer worldwide for decades.[1,2] Investigations conducted among different populations have found significant differences in the epidemiological, clinical, and molecular characteristics of lung cancer patients, especially non-smokers, and that natural genetic susceptibility and fossil fuel use are important factors in the etiology of lung cancer.[2–7] An epidemiological survey

proved that lung cancer has familial aggregation, after adjusting for tobacco smoking and other environmental factors.[8–11] Most studies have found an increased risk associated with a family history of lung cancer (FLC), which varies 1.3–3.5-fold.[8–14] Unlike other typical familial cancers, such as breast and colorectal cancer, genetic susceptibility to lung cancer has not been studied. Lung cancer has a range of well-known risk factors, such as smoking, mining, shipbuilding, industrial construction, and the occupational hazards of petroleum refining.[11] Globally, approximately 53% of lung cancer cases in women and 15% in men are not related to smoking.[5] Statistics suggest that approximately 3 billion people on earth use coal or biomass for cooking and heating purposes.[15,16] This practice poses a long-term risk of the development of cardiovascular and respiratory diseases, including lung cancer.[3,15–17] Evidence suggests that lung cancer caused by other non-tobacco factors has unique clinical and histological characteristics compared to lung cancer caused by smoking.[4–7]

Notably, Xuanwei, Yunnan Province, China, has long been the focus of large-scale epidemiological studies, reporting some of the highest rates of lung cancer in the world, especially among people who have never smoked.[7,17,18] This has partially been attributed to coal combustion.[3,16] The incidence of lung cancer in these rural areas also shows characteristics of family aggregation.[17] The shared environment can partly explain this phenomenon, but the influence of a genetic background of susceptibility cannot be ignored.

Patients from this region could represent a good model to study the intricate mechanisms of lung cancer etiology. Although much work has been done, the dynamic and complex interactions between environmental carcinogens and human genetic background remain a mystery.[4,19] Research has suggested that non-smokers and patients exposed to coal combustion exhibit unique driver mutation patterns and fusion genes, such as *ALK* and *ROS1*, which exhibit special signatures in NSCLC.[4,6,7,19] To better characterize the interaction between these two key players, this study investigated the clinical characteristics and genetic rearrangements of people with FLC and a history of exposure to coal combustion, one representing genetic causes and the other key environmental factors. Compared to fluorescence in situ hybridization (FISH) and immunohistochemistry (IHC), reverse transcription (RT)-PCR can detect fusion genes, including variant types. It is also a fast, sensitive, high-flux method.[20–22]

Methods

Patients and tissue samples

Patients with pathologically confirmed lung cancer in Ward One at the Department of Thoracic Surgery, Yunnan Cancer Hospital between February 2014 and June 2017 were included. To study the effect of FLC and exposure to occupational or household coal combustion, we focused on patients with both characteristics, based on the following criteria: (i) FLC was defined as individuals with three or more first-degree relatives affected by lung cancer (n = 192); and (ii) patients without FLC were enrolled as a control, defined as individuals with no reported cancer diagnosed among first-degree relatives for three generations. Eventually, 218 patients met the criteria. Patients' exposure to occupational or domestic coal combustion was also documented. Patients with occupational coal combustion exposure, mainly residents from Xuanwei and other rural areas, include farmers who use coal to heat agricultural facilities and workers, such as furnacemen, exposed to carcinogens related to burning coal. Residents use coal domestically to cook and heat their homes. Smoking history and occupational or domestic coal exposure was obtained based on self-report and confirmed by personal medical records.

Clinical and pathologic data were obtained from the hospital cooperated databank (https://www.linkdoc.com), which included age, gender, histologic type, family history, smoking status, occupational exposure, and critical environmental risk factors. Tumor node metastasis (TNM) stage was reviewed according to the 8th edition of the International Association for the Study of Lung Cancer (IASLC) staging system. The majority of patients enrolled had adenocarcinoma (AD) and squamous cell carcinoma (SCC); other histology types were not included as few patients met the inclusion criteria.

Lung cancer tissues were stored in RNAlater (Sigma, St. Louis, MO, USA) immediately after surgery. A slide was cut from each sample for hematoxylin and eosin staining. Those containing > 70% cancer cells and < 10% necrosis were enrolled. The ethical committees of Yunnan Cancer Hospital approved the study. All patients signed informed consent.

RNA extraction and reverse transcription-PCR

Total RNA was extracted using a TRIzol Reagent Kit (Invitrogen, Carlsbad, CA, USA), and then incubated with RNase-free DNase I (Qiagen, Hilden, Germany) to remove contaminating DNA. RT was performed using a reverse transcription kit (Promega, Madison, WI, USA) to generate complementary DNA as PCR templates. RT-PCR was performed using Sigma Taq (Sigma).

PCR products were first detected by electrophoresis and those with proper bands were sequenced by BGI Tech (Shenzhen, China; http://www.bgitechsolutions.com) for confirmation. The nucleotide sequences were verified using the BLAST program (National Center for Biotechnology

Information, Bethesda, MD, USA). All positive cases were confirmed by another independent PCR reaction.

ALK rearrangement

Primers were selected based on reference to identify different *EML4-ALK* variants (Table S1).[22–25] The thermal cycle conditions were as follows: *EML4-ALK* variant 1: 95°C, 5 minutes, 40 cycles of 95°C, 30 seconds, 55°C, 30 seconds, 72°C, 1 minute; and *EML4-ALK* variants 2–7: 95°C, 10 minutes, 40 cycles of 95°C, 30 seconds, 66°C, 30 seconds, 72°C, 2.5 minutes.

Other *ALK* rearrangements were also examined in the negative samples for *EML4-ALK*, including *TFG-ALK*, *KLC1-ALK*, and *KIF5B-ALK*. The primers and PCR conditions were designed based on previous reports (Table S1):[26–28] *KIF5B-ALK*: 95°C, 5 minutes, 40 cycles of 95°C, 30 seconds, 50°C, 30 seconds, 72°C, 2.5 minutes; *KLC1-ALK*: 95°C, 5 minutes, 40 cycles of 95°C, 40 seconds, 55°C, 40 seconds, 72°C, 3 minutes; and *TFG-ALK*: 95°C, 5 minutes, 40 cycles of 94°C, 40 seconds, 65°C, 40 seconds, 72°C, 1.5 minutes.

ROS1 rearrangement

Different *ROS1* fusions were examined, including *CD74-ROS1*, *TPM3-ROS1*, *EZR-ROS1*, *SLC34A2-ROS1*, *LRIG3-ROS1*, *GOPC-ROS1*, and *SDC4-ROS1*. The primers and PCR conditions were designed based on previous studies (Table S1).[29,30] The PCR conditions were as follows: 95°C, 5 minutes, 10 cycles of touchdown PCR (annealing from 63 to 58°C with a 0.5 decrease each cycle) and 30 cycles of 95°C, 40 seconds, 58°C, 40 seconds, 72°C, 1 minute, with a final extension of 72°C for 5 minutes.

RET rearrangement

The potential existence of *CCDC6-RET* and *KIF5B-RET* were also examined. The primers and PCR conditions were set according to a previous study (Table S1).[31] The program for *CCDC6-RET* was: 95°C, 4 minutes, 40 cycles of 95°C, 30 seconds, 60°C, 30 seconds, 72°C, 30 seconds. The program for *KIF5B-RET* was: 95°C, 4 minutes, 40 cycles of 95°C, 30 seconds, 62°C, 30 seconds, 72°C, 30 seconds.

NTRK1 rearrangement

The rarely reported *CD74-NTRK1* fusion was also examined. The primers and PCR conditions were set based on a previous report (Table S1).[32] The PCR conditions were as follows: 95°C, 4 minutes, 10 cycles of touchdown PCR (annealing from 62 to 57°C with a 0.5 decrease each cycle) and 30 cycles of 95°C, 40 seconds, 57°C, 40 seconds, 72°C, 1 minute.

Statistical analysis

Chi-square and Fischer's exact tests were used to analyze the association of clinicopathological parameters with FLC and occupational/domestic coal exposure. SPSS version 17.0 (SPSS, Chicago, IL, USA) was used. Statistical significance was set at $P < 0.05$ (two-sided P value).

Results

Clinicopathological features of the study population

In total, 410 lung cancer patients were included. The clinicopathological characteristics are shown in Tables 1 and 2. There were 277 men and 133 women at an average age of 59 (range 28–84) years; 192 patients (male: 114, female: 78) had FLC while 218 (male: 163, female: 55) reported no FLC. The histological subtypes included 326 (79.5%) AD and 84 (20.5%) SCC. A total of 204 (49.8%; male: 177, female: 27) patients were exposed to occupational or domestic coal combustion and 231 (56.3%, all male) were smokers, among whom 173 (42.2%) were both smokers and exposed to coal combustion. To better evaluate major carcinogen exposure in the sample population, both factors were considered together. The definition of high exposure was either being a smoker or exposed to coal combustion, while low exposure was defined as a non-smoker not exposed to coal combustion. A total of 262 (63.9%, male: 235, female: 27) subjects were classified in the high carcinogen exposure group.

Patients with a family history of lung cancer (FLC): Younger and with a higher female ratio

Significantly more women had FLC ($P < 0.01$). Because men dominated the smoking population, most women were never smokers. There were more men in the high exposure than in the low exposure group (Table 1, Fig 1a). Importantly, the distribution of patient age also reveals interesting variation in different subpopulations. The age structure curve of FLC patients increased compared to patients without FLC: there were significantly more individuals aged < 55 ($P < 0.01$) (Table 1, Fig 1b). A similar apparent age difference was observed in subgroups (Fig 1c). Men with FLC developed the disease much earlier than those without (median age: 55 vs. 64 years), while in women the age gap was smaller (median age: with FLC 53.5 vs. without FLC 57 years). On the other hand, if first

Table 1 Clinical characteristics of 410 NSCLC patients grouped by FLC and exposure level

Variables	Total	Total FLC Case (%)			High exposure group FLC Case (%)			Low exposure group FLC Case (%)		
		Positive	Negative	*P**	Positive	Negative	*P**	Positive	Negative	*P**
Total number of patients	410	192 (46.8)	218 (53.2)		94 (24.4)	168 (64.1)		98 (66.2)	50 (33.8)	
Gender				0.001			0			0
Male	277	114 (59.4)	163 (74.8)		75 (79.8)	160 (95.2)		39 (39.8)	3 (6.0)	
Female	133	78 (40.6)	55 (25.2)		19 (20.2)	8 (4.8)		59 (60.2)	47 (94.0)	
Average age: 59 years (range 28–84)				0			0			0.037
≤ 55 years	145	104 (54.2)	41 (18.8)		55 (58.5)	25 (14.9)		49 (50.0)	16 (32.0)	
> 55 years	265	88 (45.8)	177 (81.2)		39 (41.5)	143 (85.1)		49 (50.0)	34 (68.0)	
Histological type				0			0			0.550
Adenocarcinoma	326	180 (93.7)	146 (67.0)		84 (89.4)	96 (57.1)		96 (98.0)	50 (100)	
Squamous cell carcinoma	84	12 (6.3)	72 (33.0)		10 (10.6)	72 (42.9)		2 (2.0)	0 (0)	
T				0.138			0.54			0.25
T1	69	34 (17.7)	35 (16.1)		14 (14.9)	24 (14.3)		20 (20.4)	11 (22.0)	
T2	178	75 (39.1)	103 (47.2)		36 (38.3)	77 (45.8)		39 (39.8)	26 (52.0)	
T3	65	28 (14.6)	37 (17.0)		17 (18.1)	31 (18.5)		11 (11.2)	6 (12.0)	
T4	98	55 (28.6)	43 (19.7)		27 (28.7)	36 (21.4)		28 (28.6)	7 (14.0)	
N				0.033			0.21			0.18
N0	153	70 (36.5)	83 (38.1)		29 (30.9)	63 (37.5)		41 (41.8)	20 (40.0)	
N1	33	10 (5.2)	23 (10.6)		6 (6.4)	18 (10.7)		4 (4.1)	5 (10.0)	
N2	129	57 (29.7)	72 (33.0)		32 (34.0)	55 (32.7)		25 (25.5)	17 (34.0)	
N3	95	55 (28.6)	40 (18.3)		27 (28.7)	32 (19.0)		28 (28.6)	8 (16.0)	
M				0			0.007			0.004
M0	255	99 (54.4)	156 (71.6)		51 (54.3)	119 (70.8)		48 (49.0)	37 (74.0)	
M1	155	93 (45.6)	62 (28.4)		43 (45.7)	49 (29.2)		50 (51.0)	13 (26.0)	
Stage				0			0.037			0
I	98	45 (23.4)	53 (24.3)		16 (17.0)	42 (25.0)		29 (29.6)	11 (22.0)	
II	51	17 (8.8)	34 (15.6)		8 (8.5)	25 (14.9)		9 (9.2)	9 (18.0)	
III	106	37 (19.3)	69 (31.7)		27 (28.7)	52 (31.0)		10 (10.2)	17 (34.0)	
IV	155	93 (48.4)	62 (28.4)		43 (45.7)	49 (29.2)		50 (51.0)	13 (26.0)	

**P* value calculated by chi-square or Fisher's exact test, when there is at least one cell with an expected count < 5. High exposure is defined as being a smoker or exposed to coal use, while low exposure refers non-smokers and no coal exposure. FLC, family history of lung cancer; NSCLC, non-small cell lung cancer.

divided by overall exposure (tobacco and coal), subjects in the same exposure group were still affected by FLC: FLC patients were younger than their counterparts. The age gap was much larger in the high exposure (median age: with FLC 53 vs. without FLC 63 years) compared to the low exposure group (median age: with FLC 55.5 vs. without FLC 58.5 years).

FLC patients: Dominated by adenocarcinoma and a higher frequency of later stage diagnosis

Clinical characteristics were first divided according to gender and total exposure level, and then the effects of FLC were further compared between total and subgroup exposure level. AD was the major histological type in all subgroups (Table 1), but there were statistically more cases of SCC in subgroups negative for FLC, whether divided by gender or exposure (Fig 1d,i). There was a higher percentage of stage IV patients in the FLC subgroup ($P < 0.05$) within the same gender or exposure group (Fig 1e,j). FLC patients had more T4 and N3 disease, but only the N stage in total was statistically significant (Fig 1f,g,k,l). The greatest variation was in M stage, where FLC patients in all groups experienced significantly more distant metastasis ($P < 0.05$) (Fig 1h,m). Overall, patients with FLC had a higher risk of lymph node and other organ metastasis. The influence of FLC was consistent in the same gender or within the same exposure group.

ALK rearrangement: Associated with younger age, women, and never smokers

Of the total 410 patients, 51 (12.4%: 22 men 7.9%, 29 women 21.8%) had *ALK* rearrangement (Table 2). Patients with *ALK* fusion were younger ($P < 0.01$), with

Table 2 Clinical characteristics of gene rearrangements detected by reverse transcription-PCR in 410 NSCLC patients

Variables	Total	*ALK* rearrangements Case (%)			*ROS1* rearrangements Case (%)			Total rearrangements Case (%)		
		Positive	Negative	*P**	Positive	Negative	*P**	Positive	Negative	*P**
Total number of patients	410	51 (12.4)	359 (87.6)		16 (3.9)	394 (96.1)		67 (16.3)	343 (83.7)	
Gender				0			0			0
Male	277	22 (7.9)	255 (92.1)		3 (1.1)	274 (98.9)		25 (9.0)	252 (91.0)	
Female	133	29 (21.8)	104 (78.2)		13 (9.8)	120 (90.2)		42 (31.6)	91 (68.4)	
Average age: 59 years (range 28–84)				0.002			0.001			0
≤55 years	145	28 (19.3)	117 (80.7)		12 (8.3)	133 (91.7)		40 (27.6)	105 (72.4)	
> 55 years	265	23 (8.7)	242 (91.3)		4 (1.5)	261 (98.5)		27 (10.2)	238 (89.8)	
Histological type				0.006			0.05			0
Adenocarcinoma	326	48 (14.7)	278 (85.3)		16 (4.9)	310 (95.1)		64 (19.6)	262 (80.4)	
Squamous cell carcinoma	84	3 (3.6)	81 (96.4)		0 (0)	84 (100)		3 (3.6)	81 (96.4)	
Stage				0.013			0.87			0.014
I	98	15 (15.3)	83 (84.7)		5 (5.1)	93 (94.9)		20 (25.6)	78 (74.4)	
II	51	6 (11.8)	45 (88.2)		2 (3.9)	49 (96.1)		8 (15.7)	43 (84.3)	
III	106	4 (3.8)	102 (96.2)		3 (2.8)	103 (97.2)		7 (6.6)	99 (93.4)	
IV	155	26 (16.8)	129 (83.2)		6 (3.9)	149 (96.1)		32 (20.6)	123 (79.4)	
Smoking history				0			0.002			0
Yes (Current or ex-smoker)	231	17 (7.4)	214 (92.6)		3 (1.3)	228 (98.7)		20 (8.7)	211 (91.3)	
Never	179	34 (19.0)	145 (81.0)		13 (7.3)	166 (92.7)		47 (26.3)	132 (73.7)	
Occupational or domestic coal use				0.022			0.98			0.041
Yes (current or ex-user)	204	33 (16.2)	171 (83.8)		8 (3.9)	196 (96.1)		41 (20.1)	163 (79.9)	
Never	206	18 (8.7)	188 (91.3)		8 (3.9)	198 (96.1)		26 (12.6)	180 (87.4)	
Overall exposure level (tobacco and coal)				0.046			0.68			0.046
High (smoker or household coal-user)	262	39 (14.9)	223 (85.1)		11 (4.2)	251 (95.8)		50 (19.1)	212 (80.9)	
Low (never smoker & never coal-user)	148	12 (8.1)	136 (91.9)		5 (3.4)	143 (96.6)		17 (11.5)	131 (88.5)	
Family history of lung cancer				0.001			0.07			0
Present	192	35 (18.2)	157 (81.8)		11 (5.7)	181 (94.3)		46 (24.0)	146 (76.0)	
Absent	218	16 (7.3)	202 (92.7)		5 (2.3)	213 (97.7)		21 (9.6)	197 (90.4)	
Family history in high exposure group				0			0.06			0
Present	94	26 (27.7)	68 (72.3)		7 (7.4)	87 (92.6)		33 (35.1)	61 (64.9)	
Absent	168	13 (7.7)	155 (92.3)		4 (2.4)	164 (97.6)		17 (10.1)	151 (89.9)	
Family history in low exposure group				0.75			0.66			0.34
Present	98	9 (9.2)	89 (90.8)		4 (4.1)	94 (95.9)		13 (13.3)	85 (86.7)	
Absent	50	3 (6.0)	47 (94.0)		1 (2.0)	49 (98.0)		4 (8.0)	46 (92.0)	

**P* value calculated by chi-square or Fisher's exact test, when there is at least one cell with an expected count < 5. NSCLC, non-small cell lung cancer.

28 (28/145, 19.3%) aged < 55 and 23 (23/265, 8.7%) aged > 55. The pathological diagnoses included 48 cases with AD and three cases with SCC ($P < 0.01$). The positive cases mainly occurred in stages I (15/51, 29.4%) and IV (26/51, 51.0%). Importantly, 4/15 (26.7%) patients in stage I were FLC positive, while 24/26 (92.3%) in stage IV were FLC positive. *ALK* rearrangement was significantly associated with female gender ($P < 0.01$) and non-smoking status ($P < 0.01$). All 51 patients had *EML4-ALK* fusion: 17 were variant 1, 6 were variant 2, and 28 were variant 3a/3b (Table S2).

ALK rearrangement: Associated with FLC and coal exposure

In this series, *ALK* rearrangement was strongly associated with FLC ($P < 0.01$): 35 patients (35/192, 18.2%) were positive (Table 2). Because genetic and environmental factors

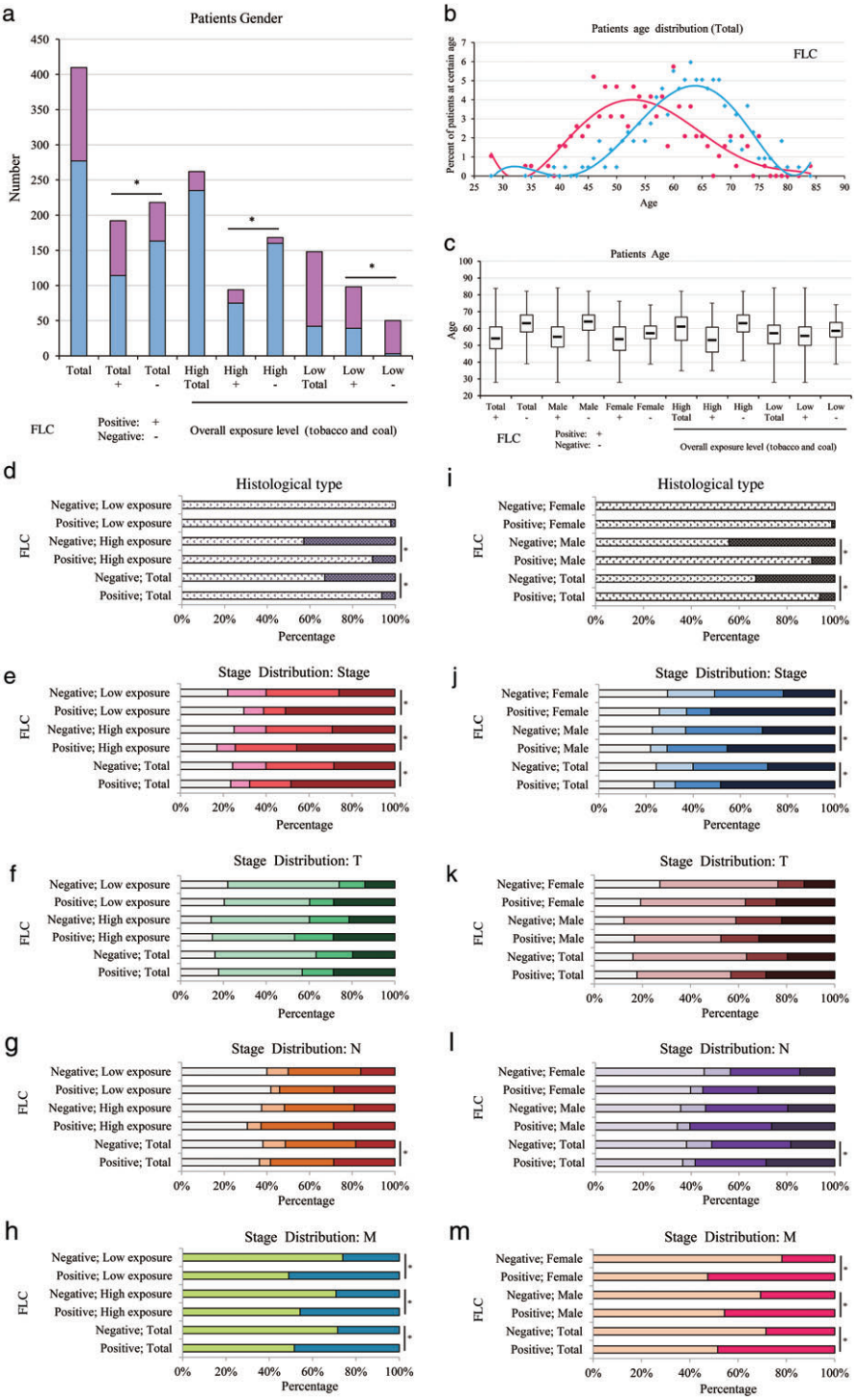

Figure 1 Patient age, gender, histology, and stage distribution. (**a**) Gender ratio in different subgroups. (■) Female, and (■) Male. (**b**) Patient age structure curves. Both fitted curves were 5th-degree polynomial; the function for FLC patients is: $y = -4E{-}07x^5 + 0.0001x^4 - 0.015x^3 + 0.8543x^2 - 23x + 234.51$ ($R^2 = 0.715$); and for patients without FLC is: $y = 8E{-}07x^5 - 0.0002x^4 + 0.0228x^3 - 1.1361x^2 + 27.063x - 247.71$ ($R^2 = 0.8616$) (●) Positive, and (●) Negative. (**c**) Patient age distribution in subgroups. Patient histology and tumor node metastasis stage divided by (**d–h**) exposure level (□) AD, and (▩) SCC; (□) Stage I, (■) Stage II, (■) Stage III, and (■) Stage IV; (□) T1, (■) T2, (■) T3, and (■) T4; (□) N0, (■) N1, (■) N2, and (■) N3; (■) M0, and (■) M1 and (**i–m**) gender (□) AD, and (▩) SCC; (□) Stage I, (■) Stage II, (■) Stage III, and (■) Stage IV; (□) T1, (■) T2, (■) T3, and (■) T4; (□) N0, (■) N1, (■) N2, and (■) N3; (□) M0, and (■) M1

both contribute to gene mutation, subjects were further compared within the same exposure group. In the high exposure group, a high association existed between *ALK* fusion and FLC ($P < 0.01$). However, no significant association was found between *ALK* fusion and FLC in the low exposure group. In addition, *ALK* rearrangement was significantly associated with occupational or domestic coal combustion ($P < 0.05$): 33 patients (33/204, 16.2%) were current or ex-coal users, while 18 (18/206, 8.7%) had never been exposed to coal combustion. When smoking and coal use were combined as overall exposure (tobacco and coal), the association still existed ($P < 0.05$): 39 (39/262, 14.9%) were either a coal user or a smoker, and 12 (12/148, 8.1%) were never-coal users and never-smokers.

ROS1 rearrangement: Associated with younger age, women, and non-smokers

Among the 410 patients, 16 (3.9%: 3 men, 3/277, 1.1%; 13 women, 13/133, 9.8%) harbored *ROS1* rearrangements (Table 2). A statistically significant association was found for female gender and non-smoking. Most subjects with *ROS1* fusion were also younger ($P < 0.01$): 12 (12/145, 8.3%) were aged < 55 years and only 4 (4/265, 1.5%) were aged > 55. All 16 cases were diagnosed as adenocarcinoma: 10 patients had *CD74-ROS1* (E6/E34) fusion and the other six had unknown type (Table S3); none had a co-existing *ALK* rearrangement. Other *ROS1* fusions were not identified. No statistically significant association was observed for stage, histology type, or exposure to occupational or domestic coal combustion. The association between *ROS1* translocation and FLC was also not significant ($P = 0.07$): 11 (11/16, 68.8%) *ROS1* fusion cases had FLC (Fig 2a,b).

Other gene rearrangements

Although potential *RET* and *NTRK1* rearrangement were analyzed in all samples, no positive case was identified. This may have been a result of sample size and selection standards. Detection of other gene rearrangements can be performed in a larger sample or in populations of a different composition.

Total gene fusion pattern: Associated with women, never smokers, younger age, FLC, and coal exposure

All detected gene rearrangements were combined for evaluation (Table 2). A total of 67 cases (67/410, 16.3%) were identified, including 64 AD and 3 SCC ($P < 0.01$). Most fusion cases occurred in stages I and IV (Fig 2). A statistically significant association was also found for female gender and never smokers ($P < 0.01$). The majority of gene fusions (40/67, 59.7%) occurred in patients aged < 55 years ($P < 0.01$). When evaluated by coal exposure status and overall exposure (tobacco and coal), the association was significant ($P < 0.05$). A total gene fusion event was statistically associated with FLC ($P < 0.01$): 46 patients (46/67, 68.7%) were positive for FLC, and the association was stronger in the high exposure ($P < 0.01$) than in the low exposure group ($P < 0.05$).

ALK fusion: Asian studies revealed cross-population/subpopulation similarities and heterogeneity

Asian studies on lung cancer patients that provided details of ethnicity, gender, and smoking status with *ALK* fusion are summarized in Table S4.[2,20–23,25,27,31,33–58] In total, 33 studies were included (China 14, Japan 12, Korea 7). Among 10 837 NSCLC patients, 614 (5.7%) *ALK* rearrangements were identified, with a major positive rate of 1.4–12%. On average, more *ALK* rearrangements occurred in Chinese patients (7.9%), followed by Korean (6.9%) and Japanese (3.2%) (Fig 3a,b). When considering gender ratio and smoking history, in all three nations *ALK* rearrangement in men occurred 20% less frequently than in women. Interestingly, when *ALK* fusion cases in the 33 studies were combined, never smokers (71.2%) made up the majority carrying *ALK* fusions. When divided by nation, the never smoker ratio harboring *ALK* fusions was 3 in the Chinese group and approximately 2 each in the Japanese and Korean groups (Fig 3c,d).

Analyses of the relationship between gender, smoking ratio, and *ALK* positive rate yielded quite interesting findings. In all three countries, increases in the number of *ALK* positive patients were associated with increases in the male: female ratio (Fig 3e). In terms of study distribution, China and Korea were similar, while Japan had less overlapping work with the other two. The trend line was based on 31 studies, excluding one[27] that did not include a gender ratio and one[55] that only included stage IIIB–IV NSCLC patients. On the other hand, in studies with a lower *ALK* positive rate, more never smokers tended to have *ALK* fusion, but the frequency of smokers became higher with increasing *ALK* positive rates; the distribution was similar for all three countries (Fig 3f). In summary, an increase in the frequency of *ALK* fusion in the subject population was accompanied by rising male ratios in the positive pool. The increase appeared to be related to smoking males, suggesting a potential link between smoking and *ALK* fusion in men.

When gender, smoking history, and *ALK* positive rates were integrated into a plot for evaluation, the majority of studies gathered around the center, but a few drifted away

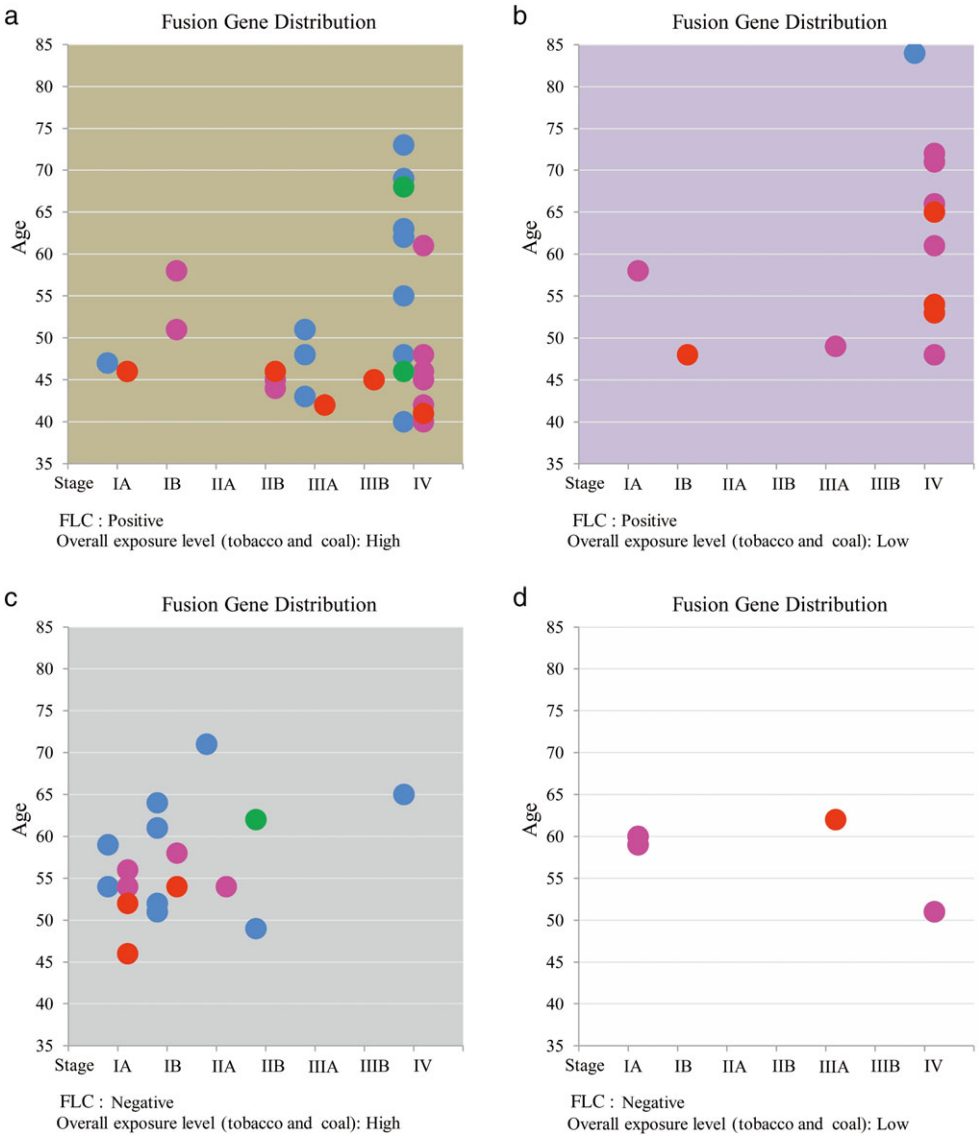

Figure 2 Total gene fusion distribution based on family history of lung cancer (FLC) and overall exposure level. FLC positive and (**a**) high and (**b**) low exposure. FLC negative and (**c**) high and (**d**) low exposure. (●) *ALK* male, (●) *ALK* female, (●) *ROS1* male, and (●) *ROS1* female

from the main group (Fig 3g). Most studies fell into the category with a higher frequency of women and never smokers, exhibited a relatively lower positive rate, and focused on stage IIIB–IV NSCLC (with the exception of one study with a much higher positive rate, represented by the large white bubble in Fig 3g, which focused on stage IIIB–IV NSCLC).[55] Studies located in the right upper quadrant included more men and never smokers but showed higher positive rates. Most Chinese, Japanese, and Korean studies focused on the same geographic regions, but a few of the Japanese studies contained little co-existing work from other ethnic groups, possibly reflecting population variation.

Subgroups by genetic, occupational, or environmental factors could partly recreate heterogeneous distribution of *ALK* fusion from previous reports

Our overall results were consistent with previous work on *ALK* rearrangement, especially those from China (Fig 3e–g). When our study population was further divided by FLC, smoking status, coal exposure, and overall exposure, the diversity of subgroups gradually took shape.

In total, 12 (6 pairs) subgroups were analyzed: 8 corresponded with the category featuring women and never smokers, similar to the majority of the studies; and 4 with

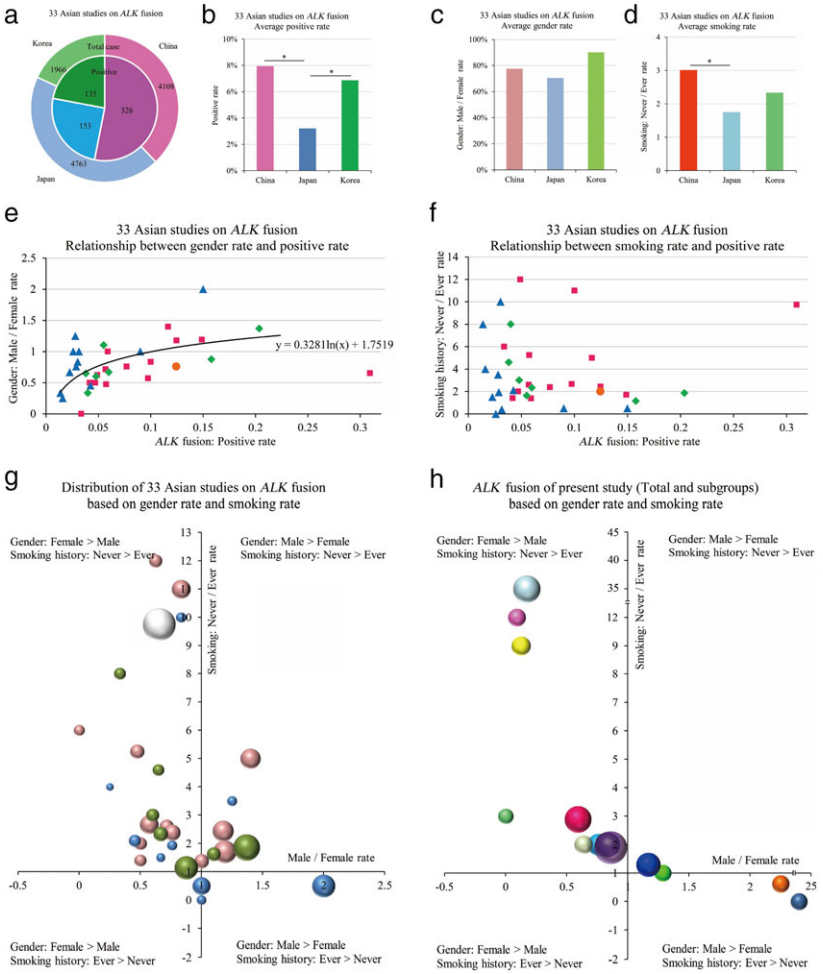

Figure 3 Comparison of 33 Asian studies on *ALK* fusion based on gender and smoking history. (**a**) *ALK* fusions over three nations. Average (**b**) positive rate of *ALK* fusion, (**c**) gender ratio, and (**d**) smoking rate. Relationships between *ALK*+ and (**e**) gender and (**f**) smoking rates. (■) China, (▲) Japan, (♦) Korea, and (●) Present study. (**g**) Distribution of 33 Asian studies on *ALK* fusion based on gender and smoking rates. (●) China, (●) Japan, and (●) Korea. (**h**) *ALK* fusion of present study (total and subgroups) based on gender and smoking rates. (*$P < 0.05$) Note: the value of the numerator based on the condition that the denominator is zero was directly used in figures. The black bubble representing the coal exposure subgroup is right behind the transparent purple bubble (high exposure and family history of lung cancer [FLC]+). (●) Present study-Total, (●) FLC+, (●) FLC−, (●) Smoker+, (●) Smoker−, (●) Coal exposure+, (●) Coal exposure−, (●) High exposure group, (●) Low exposure group, (●) High exposure + FLC+, (●) High exposure + FLC−, (●) Low exposure + FLC+, (●) Low exposure + FLC−. High exposure: either being a smoker or exposed to coal combustion; low exposure: not exposed to coal combustion and also a non-smoker.

a higher male ratio. Moreover, compared to the 33 Asian studies, 6 subgroups were located near the center, while the other 6 drifted some distance away. Notably, 4 pairs were separated into different quadrants; in only 2 pairs were both subgroups located in the same category (coal user +/−; low exposure, FLC +/−) (Fig 3h). Interestingly, dividing the population into subgroups featured by genetic, occupational, or environmental factors could at least partly recreate the diversified distribution picture of *ALK* fusion from previous reports.

ALK fusion occurred in 18.2% (35/192) of patients with FLC, with a higher ratio of women and non-smokers, and 7.3% (16/218) in patients without FLC. The *ALK*+ rate in never smokers was 19.0% versus 7.4% in smokers. In *ALK* + cases, all smokers were men and 85.3% of non-smokers were women; this subgroup exhibited the greatest difference in results. *ALK*+ occurred more frequently in subjects exposed to coal combustion (16.2% vs. 8.7%) but *ALK*+ cases in both exposed and non-exposed groups were higher in women and never smokers. In the high exposure group 14.9% (39/262) were *ALK*+ compared to 8.1% (12/148) in the low exposure group; women and never smokers dominated these categories. Patients with both FLC and high exposure exhibited the highest rate of *ALK* rearrangement (26/94, 27.7%) in all subgroups, almost matching the results of a study on advanced NSCLC (30.9%).

Interestingly, *ALK* fusion in patients without FLC but high exposure was characterized by more men, smokers, and a moderate positive rate (13/168, 7.7%), contrasting with the results of the majority of the 33 reports included in this analysis;[2,20,45,48,54] however, this subgroup was similar as the results of one of the Japanese studies.[46] Finally, regardless of FLC, *ALK*+ patients in the low exposure group were mainly women and non-smokers, but the positive rate was slightly higher in FLC patients (9.2% vs. 6.0%). Overall, different population compositions contribute to variations in *ALK* rearrangement.

Discussion

Both genetic and environmental factors contribute to the development of cancer and its mutant spectrum.[2,6,7,13,59,60] There is abundant evidence of genetic predisposition to lung cancer.[8,11,13,14,61] Our study found that regardless of gender and exposure level, FLC patients exhibited some common characteristics: younger age, a high female ratio, AD dominant, increased metastasis, later stage at diagnosis, and higher frequency of gene fusion. Many of these are indicative of increased aggressiveness. Our population had unique characteristics but our results were consistent with previous studies. Findings across previous studies showed both similarity and heterogeneity, depending on the subject source. Some studies reported that certain ethnic groups are affected more by inherited lung cancer susceptibility;[13,14] one indicated a lower association between never smokers and FLC;[13] another suggested that FLC is common in never smokers with NSCLC;[9] in regard to gender, some studies found that female relatives have a higher risk compared to male relatives[10,61] and similarly, 34% of Spanish women with lung cancer had FLC;[12] and a study that enrolled mainly Caucasian patients found a link between FLC and *EGFR*, rather than *ALK* or *KRAS* mutation.[9] These results indicate that susceptibility to lung cancer may be inherited in a complex pattern across populations, and each population or subpopulation has unique characteristics.

Importantly, FLC tend to have greater impact on lung cancer when subjects were exposed to a high level of carcinogens; studies also report that the lung cancer risk is further amplified by smoking.[14] This was reflected in our results (Fig 1c): the median age of FLC patients was 10 years younger (53 vs. 63 years) in the high exposure group. On the other hand, if all factors were equal and the study population was divided by pollution level, certain somatic mutation may occur less frequently in an unpolluted environment. Our results showed a smaller age gap in the low exposure group (median age: with FLC 55.5 vs. without FLC 58.5 years). Similarly, regarding *ALK* fusion and total gene rearrangement, the association with FLC was stronger in the high than the low exposure group (Table 2). Our findings indicate that even within the same exposure group, patients are still affected by FLC, but a susceptible genetic background makes them more vulnerable in a polluted environment.

The *ALK* fusion gene is recognized as an important oncogenic driver gene in NSCLC (main type *EML4-ALK*).[19,24,25,50] Incidence of *ALK* fusion in NSCLC is approximately 0.99–12%, with no significant differences between Asian and Western populations.[19,20,49] *ROS1* rearrangement occurs in a small subset (0.5%–2%) of NSCLC patients and is associated with slight or never smokers and adenocarcinoma histology. *CD74-ROS1* is the major type (~40%), followed by *EZR-ROS1* and unknown type (both ~15%).[62] Our results regarding *ALK* and *ROS1* fusion were similar to previous reports with the exception of elevated *ROS1* fusion frequency (3.9%).[29,30,49,62] A meta-analysis including 1178 *ALK* rearranged cases from 20 541 NSCLC patients indicated that age, gender, smoking status, histology, tumor stage, and ethnicity may be a source of between-study heterogeneity.[19] According to the reference studies from three nations, there is obvious across population/subpopulation similarity and heterogeneity. Interestingly, similar to the reference reports, our result of an *ALK* fusion increase with a rising male ratio in the positive pool was associated with smoking (Fig 3e,f), suggesting a potential relationship between smoking and *ALK* fusion in men. However, as many previously published reports from Asian countries found that *ALK* rearrangement is associated with female gender and non-smokers, this hypothesis requires further investigation.[22,36,37,50,54]

One parameter only represents one piece of the puzzle, and for many investigators it can be difficult to collect complete patient data, such as: disease history, personal genetic makeup, and every occupational or environmental risk factor. Therefore, it can be difficult to determine what kinds of missing data may cause differences between study results. Although it is a well-known assumption that variation between studies is mainly caused by different population compositions,[2,19] the division of our study sample into subgroups by genetic, occupational, or environmental factors enabled a picture of diversity in *ALK* fusion to take shape. This not only provided support for the assumption, but also suggested that different kinds of parameters should be carefully considered when conducing cancer research.

When all identified gene rearrangement events were combined for consideration (Fig 2), inherited susceptibility and exposure to environmental carcinogens appeared to be significant factors causing gene rearrangement. The combination pushed the gene fusion rate to 35.1% (33/94) in the subgroup with both characteristics. The high frequency of gene fusion events and the association with women, never

smokers, and younger age may be unique characteristics of the subject population. Other research on lung cancer conducted in this region also found special mutation signatures, such as higher *K-RAS* mutation rates (15%–29%), compared to other Asian populations (2–7%), including smokers and never smokers.[59,63–65] Coal combustion releases a cocktail of carcinogens including polycyclic aromatic hydrocarbons and fine particulate matter ($PM_{2.5}$), many of which are defined by the International Agency for Research on Cancer and the World Health Organization as class I carcinogens.[3,16,66] Certain carcinogens may have a specific effect on the human genome. For example, *p53* gene mutations in tumors from female never smokers in Xuanwei exhibited characteristics induced by polycyclic aromatic hydrocarbons, different from those in lung cancer attributed to smoking.[6] All of these findings support the hypothesis that patients exposed to coal combustion could display a unique mutational spectrum in lung cancer tumors.[5,7,63,64]

Our study found that FLC and exposure to coal combustion influence clinicopathologic features and gene fusion patterns in NSCLC. Our results reveal the potential uniqueness of the subject population, particularly that inherited susceptibility exerts a greater impact on the age of onset and frequency of gene fusion when subjects are exposed to high levels of carcinogens. Finally, our study showed that dividing a population by genetic, occupational, or environmental factors could partly recreate the diversified distribution picture of gene fusion presented over different reports. Our findings may help to evaluate the effect of FLC and coal exposure on the pathogenesis of lung cancer and highlight the significance of integrating various parameters into clinical and theoretical research.

Acknowledgments

This study was supported by National Natural Science Foundation of China (No. 81702274); the Key Project of International Cooperation of Science and Technology Innovation between Governments, the National Key Research and Development Plan of China (No. 2016YEE0103400); Yunnan Applied Basic Research Projects-Union Foundation (No. 2017FE468 [-159], 2015FB069, 2017FE467 [-0187], 2017FE468 [-214]); Internal Organization Research Projects of Yunnan Cancer Hospital (No. 2017NS198); and the Doctor Research Foundation of Yunnan Cancer Hospital (No. BSKY201705).

Disclosure

No authors report any conflict of interest.

References

1 Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; **61**: 69–90.
2 Seo J-S, Ju YS, Lee W-C *et al.* The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res* 2012; **22**: 2109–19.
3 Hosgood HD, Boffetta P, Greenland S *et al.* In-home coal and wood use and lung cancer risk: A pooled analysis of the international lung cancer consortium. *Environ Health Perspect* 2010; **118**: 1743–7.
4 Lee YJ, Kim J-H, Kim SK *et al.* Lung cancer in never smokers: Change of a mindset in the molecular era. *Lung Cancer* 2011; **72**: 9–15.
5 Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers – A different disease. *Nat Rev Cancer* 2007; **7**: 778–90.
6 DeMarini DM, Landi S, Tian D *et al.* Lung tumor KRAS and TP53 mutations in nonsmokers reflect exposure to PAH-rich coal combustion emissions. *Cancer Res* 2001; **61**: 6679–81.
7 Hosgood HD, Pao W, Rothman N *et al.* Driver mutations among never smoking female lung cancer tissues in China identify unique EGFR and KRAS mutation pattern associated with household coal burning. *Respir Med* 2013; **107**: 1755–62.
8 Frank C, Sundquist J, Yu H, Hemminki A, Hemminki K. Concordant and discordant familial cancer: Familial risks, proportions and population impact. *Int J Cancer* 2017; **140**: 1510–6.
9 Gaughan EM, Cryer SK, Yeap BY, Jackman DM, Costa DB. 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.
10 Lin H, Huang Y-S, Yan H-H *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 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**: 36. https://doi.org/10.3390/genes8010036.
12 Isla D, Felip E, Viñolas N *et al.* Lung cancer in women with a family history of cancer: The Spanish female-specific database WORLD07. *Anticancer Res* 2016; **36**: 6647–53.
13 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.
14 Coté ML, Kardia SLR, Wenzlaff AS, Ruckdeschel JC, Schwartz AG. Risk of lung cancer among white and black relatives of individuals with early-onset lung cancer. *JAMA* 2005; **293**: 3036–42.
15 Ezzati M, Kammen DM. The health impacts of exposure to indoor air pollution from solid fuels in developing countries: Knowledge, gaps, and data needs. *Environ Health Perspect* 2002; **110**: 1057–68.

16 Zhang JJ, Smith KR. Household air pollution from coal and biomass fuels in China: Measurements, health impacts, and interventions. *Environ Health Perspect* 2007; **115**: 848–55.
17 Mumford JL, He XZ, Chapman RS *et al.* Lung cancer and indoor air pollution in Xuan Wei, China. *Science* 1987; **235**: 217–20.
18 Wang TW, Vermeulen RCH, Hu W *et al.* Gene-expression profiling of buccal epithelium among non-smoking women exposed to household air pollution from smoky coal. *Carcinogenesis* 2015; **36**: 1494–501.
19 Fan L, Feng Y, Wan H, Shi G, Niu W. Clinicopathological and demographical characteristics of non-small cell lung cancer patients with ALK rearrangements: A systematic review and meta-analysis. *PLoS One* 2014; **9**: e100866.
20 Soda M, Isobe K, Inoue A *et al.* A prospective PCR-based screening for the EML4-ALK oncogene in non-small cell lung cancer. *Clin Cancer Res* 2012; **18**: 5682–9.
21 Shan L, Lian F, Guo L, Yang X, Ying J, Lin D. Combination of conventional immunohistochemistry and qRT-PCR to detect ALK rearrangement. *Diagn Pathol* 2014; **9**: 3.
22 Wu Y-C, Chang I-C, Wang C-L *et al.* Comparison of IHC, FISH and RT-PCR methods for detection of ALK rearrangements in 312 non-small cell lung cancer patients in Taiwan. *PLoS One* 2013; **8**: e70839.
23 Takeuchi K, Choi YL, Soda M *et al.* Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts. *Clin Cancer Res* 2008; **14**: 6618–24.
24 Sanders HR, Li H-R, Bruey J-M *et al.* Exon scanning by reverse transcriptase-polymerase chain reaction for detection of known and novel EML4-ALK fusion variants in non-small cell lung cancer. *Cancer Genet* 2011; **204**: 45–52.
25 Koivunen JP, Mermel C, Zejnullahu K *et al.* EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. *Clin Cancer Res* 2008; **14**: 4275–83.
26 Togashi Y, Soda M, Sakata S *et al.* KLC1-ALK: A novel fusion in lung cancer identified using a formalin-fixed paraffin-embedded tissue only. *PLoS One* 2012; **7**: e31323.
27 Takeuchi K, Choi YL, Togashi Y *et al.* KIF5B-ALK, a novel fusion oncokinase identified by an immunohistochemistry-based diagnostic system for ALK-positive lung cancer. *Clin Cancer Res* 2009; **15**: 3143–9.
28 Wong DW-S, Leung EL-H, Wong SK-M *et al.* A novel KIF5B-ALK variant in nonsmall cell lung cancer. *Cancer* 2011; **117**: 2709–18.
29 Davies KD, Le AT, Theodoro MF *et al.* Identifying and targeting ROS1 gene fusions in non-small cell lung cancer. *Clin Cancer Res* 2012; **18**: 4570–9.
30 Fu S, Liang Y, Lin Y-B *et al.* The frequency and clinical implication of ROS1 and RET rearrangements in resected stage IIIA–N2 non-small cell lung cancer patients. *PLoS One.* 2015; **10**: e0124354.
31 Wang R, Hu H, Pan Y *et al.* RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. *J Clin Oncol* 2012; **30**: 4352–9.
32 Vaishnavi A, Capelletti M, Le AT *et al.* Oncogenic and drug-sensitive NTRK1 rearrangements in lung cancer. *Nat Med* 2013; **19**: 1469–72.
33 Shinmura K, Kageyama S, Tao H *et al.* EML4-ALK fusion transcripts, but no NPM-, TPM3-, CLTC-, ATIC-, or TFG-ALK fusion transcripts, in non-small cell lung carcinomas. *Lung Cancer* 2008; **61**: 163–9.
34 Inamura K, Takeuchi K, Togashi Y *et al.* EML4-ALK fusion is linked to histological characteristics in a subset of lung cancers. *J Thorac Oncol* 2008; **3**: 13–7.
35 Inamura K, Takeuchi K, Togashi Y *et al.* EML4-ALK lung cancers are characterized by rare other mutations, a TTF-1 cell lineage, an acinar histology, and young onset. *Mod Pathol* 2009; **22**: 508–15.
36 Wong DW-S, Leung EL-H, So KK-T *et al.* The EML4-ALK fusion gene is involved in various histologic types of lung cancers from nonsmokers with wild-type EGFR and KRAS. *Cancer* 2009; **115**: 1723–33.
37 Takahashi T, Sonobe M, Kobayashi M *et al.* Clinicopathologic features of non-small-cell lung cancer with EML4-ALK fusion gene. *Ann Surg Oncol* 2010; **17**: 889–97.
38 Jokoji R, Yamasaki T, Minami S *et al.* Combination of morphological feature analysis and immunohistochemistry is useful for screening of EML4-ALK-positive lung adenocarcinoma. *J Clin Pathol* 2010; **63**: 1066–70.
39 Zhang X, Zhang S, Yang X *et al.* Fusion of EML4 and ALK is associated with development of lung adenocarcinomas lacking EGFR and KRAS mutations and is correlated with ALK expression. *Mol Cancer* 2010; **9**: 188.
40 Koh Y, Kim D-W, Kim TM *et al.* Clinicopathologic characteristics and outcomes of patients with anaplastic lymphoma kinase-positive advanced pulmonary adenocarcinoma: Suggestion for an effective screening strategy for these tumors. *J Thorac Oncol* 2011; **6**: 905–12.
41 Lee J-O, Kim TM, Lee S-H *et al.* Anaplastic lymphoma kinase translocation: A predictive biomarker of pemetrexed in patients with non-small cell lung cancer. *J Thorac Oncol* 2011; **6**: 1474–80.
42 Jin G, Jeon H-S, Lee EB *et al.* EML4-ALK fusion gene in Korean non-small cell lung cancer. *J Korean Med Sci* 2012; **27**: 228–30.
43 Paik JH, Choi C-M, Kim H *et al.* Clinicopathologic implication of ALK rearrangement in surgically resected lung cancer: A proposal of diagnostic algorithm for ALK-rearranged adenocarcinoma. *Lung Cancer* 2012; **76**: 403–9.
44 Chen T-D, Chang I-C, Liu H-P *et al.* Correlation of anaplastic lymphoma kinase overexpression and the EML4-ALK fusion gene in non-small cell lung cancer by immunohistochemical study. *Chang Gung Med J* 2012; **35**: 309–17.
45 Wang Z, Zhang X, Bai H *et al.* EML4-ALK rearrangement and its clinical significance in Chinese patients with advanced non-small cell lung cancer. *Oncology* 2012; **83**: 248–56.

46 Sakai K, Okamoto I, Takezawa K *et al.* A novel mass spectrometry-based assay for diagnosis of EML4-ALK-positive non-small cell lung cancer. *J Thorac Oncol* 2012; **7**: 913–8.
47 Takeda M, Okamoto I, Sakai K, Kawakami H, Nishio K, Nakagawa K. Clinical outcome for EML4-ALK-positive patients with advanced non-small-cell lung cancer treated with first-line platinum-based chemotherapy. *Ann Oncol* 2012; **23**: 2931–6.
48 Kobayashi M, Sonobe M, Takahashi T, Yoshizawa A, Kikuchi R, Date H. Detection of ALK fusion in lung cancer using fluorescence in situ hybridization. *Asian Cardiovasc Thorac Ann* 2012; **20**: 426–31.
49 Takeuchi K, Soda M, Togashi Y *et al.* RET, ROS1 and ALK fusions in lung cancer. *Nat Med* 2012; **18**: 378–81.
50 Li Y, Li Y, Yang T *et al.* Clinical significance of EML4-ALK fusion gene and association with EGFR and KRAS gene mutations in 208 Chinese patients with non-small cell lung cancer. *PLoS One* 2013; **8**: e52093.
51 Li Y, Pan Y, Wang R *et al.* ALK-rearranged lung cancer in Chinese: A comprehensive assessment of clinicopathology, IHC, FISH and RT-PCR. *PLoS One* 2013; **8**: e69016.
52 To K-F, Tong JHM, Yeung KSF *et al.* Detection of ALK rearrangement by immunohistochemistry in lung adenocarcinoma and the identification of a novel EML4-ALK variant. *J Thorac Oncol* 2013; **8**: 883–91.
53 Xia N, An J, Jiang Q-Q, Li M, Tan J, Hu C-P. Analysis of EGFR, EML4-ALK, KRAS, and c-MET mutations in Chinese lung adenocarcinoma patients. *Exp Lung Res* 2013; **39**: 328–35.
54 Zhou JX, Yang H, Deng Q *et al.* Oncogenic driver mutations in patients with non-small-cell lung cancer at various clinical stages. *Ann Oncol* 2013; **24**: 1319–25.
55 Han X-H, Zhang N-N, Ma L *et al.* Immunohistochemistry reliably detects ALK rearrangements in patients with advanced non-small-cell lung cancer. *Virchows Arch* 2013; **463**: 583–91.
56 Lee HY, Ahn HK, Jeong JY *et al.* Favorable clinical outcomes of pemetrexed treatment in anaplastic lymphoma kinase positive non-small-cell lung cancer. *Lung Cancer* 2013; **79**: 40–5.
57 Sakai Y, Nakai T, Ohbayashi C *et al.* Immunohistochemical profiling of ALK fusion gene-positive adenocarcinomas of the lung. *Int J Surg Pathol* 2013; **21**: 476–82.
58 Wang J, Cai Y, Dong Y *et al.* Clinical characteristics and outcomes of patients with primary lung adenocarcinoma harboring ALK rearrangements detected by FISH, IHC, and RT-PCR. *PLoS One.* 2014; **9**: e101551.
59 Kawaguchi T, Ando M, Kubo A *et al.* Long exposure of environmental tobacco smoke associated with activating EGFR mutations in never-smokers with non-small cell lung cancer. *Clin Cancer Res* 2011; **17**: 39–45.
60 An S-J, Chen Z-H, Su J *et al.* Identification of enriched driver gene alterations in subgroups of non-small cell lung cancer patients based on histology and smoking status. *PLoS One.* 2012; **7**: e40109.
61 Lin K-F, Wu H-F, Huang W-C, Tang P-L, Wu M-T, Wu F-Z. 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.
62 Rossi G, Jocollé G, Conti A *et al.* Detection of ROS1 rearrangement in non-small cell lung cancer: Current and future perspectives. *Lung Cancer* 2017; **8**: 45–55.
63 Sun Y, Ren Y, Fang Z *et al.* Lung adenocarcinoma from east Asian never-smokers is a disease largely defined by targetable oncogenic mutant kinases. *J Clin Oncol* 2010; **28**: 4616–20.
64 Wu C-C, Hsu H-Y, Liu H-P *et al.* Reversed mutation rates of KRAS and EGFR genes in adenocarcinoma of the lung in Taiwan and their implications. *Cancer* 2008; **113**: 3199–208.
65 Koivunen JP, Kim J, Lee J *et al.* Mutations in the LKB1 tumour suppressor are frequently detected in tumours from Caucasian but not Asian lung cancer patients. *Br J Cancer* 2008; **99**: 245–52.
66 Gordon SB, Bruce NG, Grigg J *et al.* Respiratory risks from household air pollution in low and middle income countries. *Lancet Respir Med* 2014; **2**: 823–60.

Supporting Information

Additional Supporting Informationmay be found in the online version of this article at the publisher's website:

Table S1. Primer sets for detection of gene rearrangements.
Table S2. Clinicopathological characteristics of 51 non-small cell lung cancer (NSCLC) patients with *ALK* rearrangements.
Table S3. Clinicopathological characteristics of 16 non-small cell lung cancer (NSCLC) patients with *ROS1* rearrangement.
Table S4. Investigation of 33 *ALK* rearrangement studies.

