



# Presentation of *EGFR* mutations in 162 family probands with multiple primary lung cancer

Chunxiang Li<sup>1#</sup>, Yalong Wang<sup>2#</sup>, Kai Su<sup>1</sup>, Yu Liu<sup>1</sup>, Liyu Wang<sup>1</sup>, Bo Zheng<sup>3</sup>, Na Yan<sup>4</sup>, Dawei Yuan<sup>4</sup>, Yanxiang Zhang<sup>4</sup>, Liyan Xue<sup>3</sup>, Shugeng Gao<sup>1</sup>, Jie He<sup>1</sup>

<sup>1</sup>Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; <sup>2</sup>Department of Anesthesiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; <sup>3</sup>Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; <sup>4</sup>Geneis (Beijing) Co., Ltd., Beijing, China

**Contributions:** (I) Conception and design: C Li, Y Wang; (II) Administrative support: C Li, L Xue, S Gao; (III) Provision of study materials or patients: C Li, Y Wang; (IV) Collection and assembly of data: K Su, Y Liu, L Wang; (V) Data analysis and interpretation: B Zheng, N Yan, D Yuan; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

**Correspondence to:** Professor Jie He. Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No. 17 Panjiayuan Nanli, Beijing 100021, China. Email: prof.jiehe@gmail.com.

**Background:** The number of multiple primary lung cancer (MPLC) patients has rapidly increased in recent years. However, information regarding the etiology of MPLC and responsiveness to epidermal growth factor receptor (*EGFR*)-tyrosine kinase inhibitors (TKIs) is rare. The present study aims to describe the mutation signatures of *EGFR* in MPLC.

**Methods:** Polymerase chain reaction (PCR) and Sanger sequencing were used to screen *EGFR* mutations in 162 family probands comprising 366 tumor lesions and 162 paired noncancerous adjacent tissues (NATs). Sequencing data from 3,243 sporadic lung adenocarcinoma (LUAD) samples were analyzed as a control.

**Results:** Candidate germline mutations were observed in exons 19 (3, 1.85%), 20 (8, 4.94%) and 21 (5, 3.10%), with a total frequency of 9.88% in NATs (16/162). There were 63 probands harboring somatic mutations (63/162, 38.89%), 9 patients harbored the consistency mutations among lesions, and 8 patients carried 2 or more mutations. The overall rate of *EGFR* somatic mutations was lower in the MPLC probands, but those of exon 19 p.747-752del, 20 p.V769indelsVASV and 20 p.D770indelsDSVD were significantly higher in MPLC probands than in patients with sporadic LUAD.

**Conclusions:** There exists unique *EGFR* mutation signatures in a large cohort of MPLC probands, which might provide objective evidence of the etiology and effectiveness of clinical TKI treatment of high-risk MPLC patients.

**Keywords:** Multiple primary lung cancer (MPLC); lung adenocarcinoma (LUAD); probands; candidate germline mutation; somatic mutation

Submitted Sep 02, 2020. Accepted for publication Feb 26, 2021.

doi: 10.21037/tlcr-20-1001

View this article at: <http://dx.doi.org/10.21037/tlcr-20-1001>

## Introduction

Multiple primary lung cancer (MPLC) refers to the synchronous or metachronous occurrence of two or

more primary lung cancers in the same patient (1). With advances in chest computed tomography (CT) imaging in recent years, the number of MPLC patients has rapidly increased (2-4). MPLC is relatively common among

female nonsmokers and has an excellent prognosis (3). The incidence of first-degree relatives of MPLC patients with a family history of this disease is significantly higher (5). Our previous study also revealed that the incidence of malignant tumors in first-degree relatives of MPLC patients was as high as 35.8% (102/285) (6). A recent study also suggested that MPLC has unique genetic characteristics and should therefore be treated as an independent disease (4). All the above findings suggested a potentially different etiology for MPLC; however, etiological studies of MPLC are still rare.

Epidermal growth factor receptor (*EGFR*) germline mutations, including the mutations p.T790M (7-9) and p.R776H (10,11) in exon 20 and p.V843I in exon 21 (12,13), are associated with genetic susceptibility to lung cancer. The *EGFR*-T790M germline mutation was first identified in a European family with a history of MPLC and was associated with drug resistance (14). Several other studies reported the existence of the *EGFR*-T790M germline mutation in familial MPLC cases (14,15). Our previous study also revealed that the *EGFR*-R776H germline mutation existed in a family with a high incidence of MPLC (11). These studies have facilitated the understanding of how germline mutations in *EGFR* may regulate the occurrence of MPLC. The few described cases to date have been insufficient to uncover the genetic relationship between *EGFR* mutation and MPLC due to the low frequency of MPLC patients with mutant *EGFR*.

Lung cancers with specific somatic *EGFR* mutations have a favorable clinical response to *EGFR*-tyrosine kinase inhibitors (TKIs) (16-18). The most frequent *EGFR* mutations are small insertions and deletions (indels) around exon 19 and the p.L858R mutation in exon 21, which are oncogenic and hypersensitive to *EGFR*-TKIs. Other less prevalent but clinically relevant *EGFR* mutations that predict a resistant response to *EGFR*-TKIs include p.G719A/S/C in exon 18; p.dels in exon 19; p.S768I, p.T790M, and p.indels in exon 20; and p.L861Q in exon 21 (19).

To identify the signatures of candidate germline and tumor-derived mutations in family members and to provide evidence of etiology and clinical TKI treatment for high-risk MPLC patients, 162 probands [162 noncancerous adjacent tissues (NATs) and 366 tumor lesions] were screened in the current study. Additionally, somatic mutations were described in the other 3,243 sporadic lung adenocarcinoma (LUAD) patient cohorts. We present the following article in accordance with the STROBE reporting checklist (available at <http://dx.doi.org/10.21037/tlcr-20-1001>).

## Methods

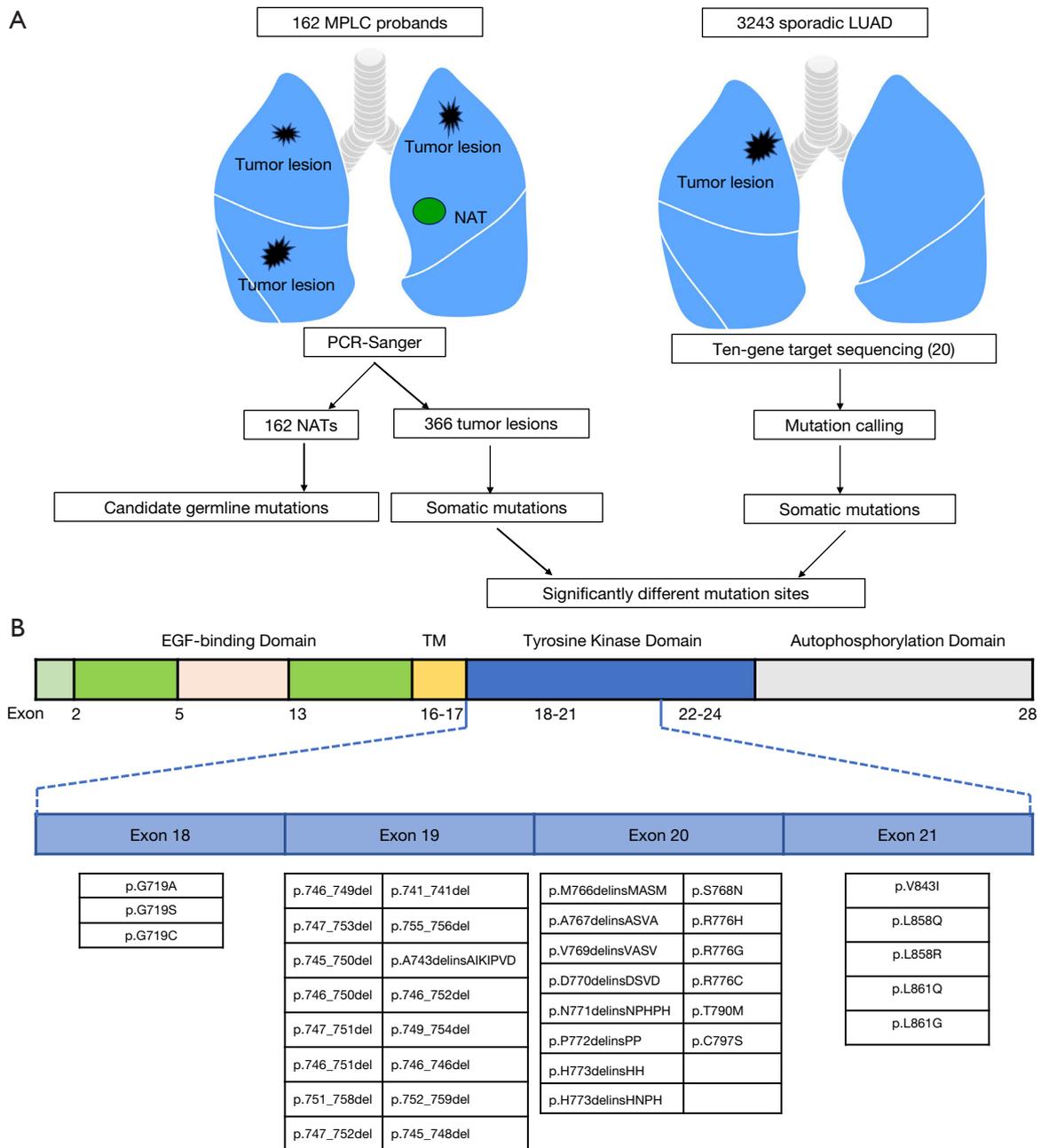
### *Study design and patient selection*

In all, 366 tumor lesions and 162 NATs from a cohort of 162 family probands were separately genotyped for candidate germline mutations and somatic mutations (Figure 1A). All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013). Ethical approval was obtained from the medical ethics committee of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (NCC2018AA-015). And all the patients provided written informed consent. As a control cohort for somatic mutations in the current study, samples from 3243 LUAD patients were previously subjected to ten-gene target sequencing (20). All *EGFR* mutant sites were described in detail in both cohorts (Figure 1B).

Selection criteria included a diagnosis of MPLC and sufficient archived tissue samples available for PCR sequencing. Each proband represents one family with at least one affected first-degree relative with MPLC, and a total of 162 probands were analyzed in our study. The study population was limited to surgically treated family probands with pathological diagnosis of MPLC at our hospital between January 2011 and January 2017. The inclusion criteria for patients were as follows: (I) pathologically confirmed MPLC (according to the 8th American Joint Committee on Cancer (AJCC) edition; typical CT and pathological images of MPLC are shown in Figure 2) and at least one first-degree relative with pathologically confirmed primary lung cancer; and (II) available complete medical records. The source for detecting *EGFR* mutations was formalin-fixed, paraffin-embedded (FFPE) normal tissues and tumor tissues that were collected after resection. Pathological diagnosis and staging (according to the 8th AJCC edition) were performed by two pathologists and two thoracic surgeons, respectively, who were blinded to the patient data. Detailed characteristics of the 3243 patients with sporadic LUAD were described in a previous study by the current authors (20).

### *FFPE tissue preparation and DNA extraction*

Lung tumor and adjacent normal tissue samples were formalin fixed and paraffin embedded after surgical resection to produce FFPE samples. Genomic DNA (gDNA) was extracted from the FFPE samples using the GeneRead DNA FFPE Kit (Qiagen, Hilden, Germany)



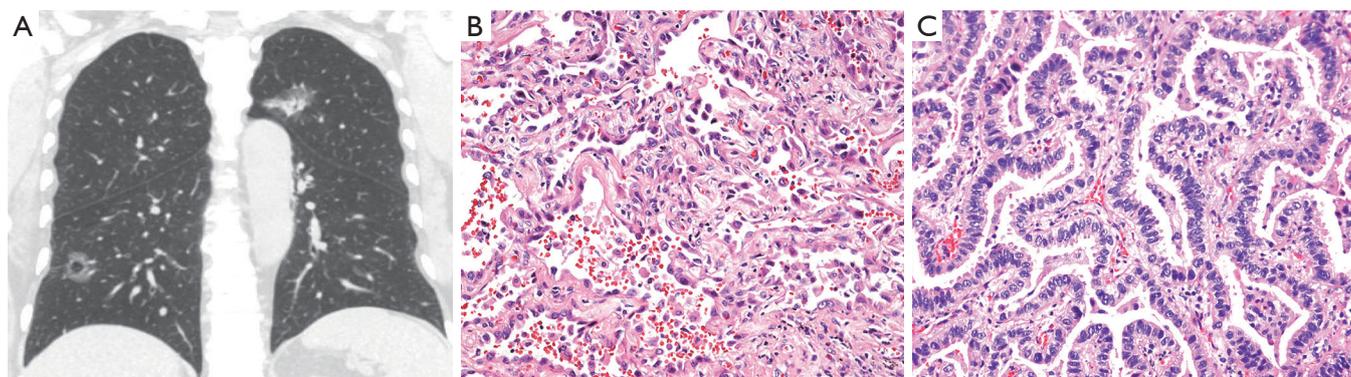
**Figure 1** Study design and *EGFR* mutation sites. (A) Mutation detection flow chart for the two cohorts. (B) *EGFR* mutation sites (20).

and quantified with Qubit (Life Technologies, Carlsbad, California, USA) in accordance with the manufacturer's protocol.

#### *EGFR* mutation detection in 162 MPLC patients

All mutations were amplified from the gDNA of 162

probands using standard PCR. PCR was conducted using the following gene-specific sense and antisense primers: exon 18: F (5'-GAGGTGACCCTTGTCTCTGTGT-3') and R (5'-CCCAAACACTCAGTGAACAAA-3'), exon 19: F (5'-CAATATCAGCCTTAGGTGCGG-3') and R (5'-GATGTGGAGATGAGCAGGGTC-3'), exon 20: F (5'-GTCCATGTGCCCTCCTTCT-3') and



**Figure 2** Representative CT and histopathological images for an MPLC patient. (A) The CT image shows two lung lesions in the left and right lungs. (B) The histopathological findings indicated adenocarcinoma of the right lung (200 $\times$ , hematoxylin-eosin staining). (C) The histopathological findings indicated adenocarcinoma of the left lung (400 $\times$ , hematoxylin-eosin staining).

R (5'-TTCCCTGATTACCTTTGCGA-3'), and exon 21: F (5'-CGTTCGCCAGCCATAAGTCCT-3') and R (5'-GCATCCTCCCCTGCATGTGTT-3'). PCR amplification and agarose gel detection were repeated 3 times, and two researchers independently interpreted the mutation data and gave reports.

#### ***EGFR somatic mutation analysis in 3,243 patients with sporadic LUAD***

Sequencing data for 3,243 patients with sporadic LUAD from Geneis (Beijing) Co., Ltd. were read using Finch TV software (Geospiza Inc., Seattle, USA), and genotyping was performed. Variant calling was performed on the Lung Cancer Ten-Genes Panel (Geneis Co., Ltd.) from NextSeq CN500 sequencing with Burrows-Wheeler Aligner (BWA) and FreeBayes software. The common clinical databases used in this study included PharmGKB, the Human Gene Mutation Database (HGMD), Clinvar, Catalogue of Somatic Mutations in Cancer (COSMIC), SNPedia, 1000 Genomes and dbSNP (20). A blinded approach was followed using frequency thresholds of  $\geq 1\%$  and  $\geq 5\%$  to define a mutation for the tumor tissue samples analyzed.

#### ***Statistical analysis***

SPSS version 13.0 (IBM Corporation, Chicago, IL, USA) was used. Intergroup differences were compared using the chi-square test or Fisher's exact test. Mann-Whitney U test and Kruskal-Wallis test were used for analysis of continuous variables with a non-normal distribution respectively. A value of  $P < 0.05$  was considered significant.

## **Results**

### ***Patients and samples***

A total of 162 MPLC probands with 366 tumor lesions and 162 corresponding nontumor normal tissue samples were screened for *EGFR* mutations. Samples were obtained from paraffin-embedded tissues deposited in a tumor bank. The clinicopathological characteristics of the 162 probands are summarized in *Table 1*. We chose multiple lesion samples and corresponding nontumor normal tissue samples for PCR sequencing.

Detailed characteristics of the 3,243 patients with sporadic LUAD were described in a previous study by the current authors (20).

### ***Distribution of candidate germline EGFR mutations in 162 MPLC patients***

We proposed that *EGFR* mutations are common variants that predispose individuals to lung cancer and that they are potentially enriched in probands representing lung cancer families. gDNA from 162 MPLC family probands was analyzed. PCR amplification and DNA sequencing of *EGFR* exon 18 p.G719X, 19 p.dels, 20 p.S768I, 20 p.R776H/G/C, 20 p.T790M, 20 p.C797S, 20 p.indels (p.M766indelsMASM, p.A767indelsASVA, p.V769indelsVASV, p.D770indelsDSVD, p.N771indelsNPHPH, p.P772indelsPP, p.H773indelsHH, and p.H773indelsHNPH), 21 p.V843I, 21 p.L858R and 21 p.L861Q/G were performed (*Figure 3A,B,C*). The mutations described above were detected in 162 nontumor tissue samples adjacent to cancerous tissue with the following distribution: exon 19, p.746\_750del (2, 1.23%)

**Table 1** Clinical characteristics of 162 family probands

Patient characteristics	Results
Age (y), mean [range]	59 [37–82]
Sex, No. (%)	
Male	58 (35.8%)
Female	104 (64.2%)
Smoking history, No. (%)	
Never smokers	118 (72.8%)
Heavy smokers (pack-year >20)	44 (27.2%)
Number of lung cancer patients in first relatives, No. (%)	
1	136 (84.0%)
2	21 (13.0%)
3	5 (3.1%)
Number of tumor lesions, No. (%)	
2	129 (79.6%)
3	26 (16.0%)
4	5 (3.1%)
5	2 (1.2%)
Histology of the sequenced tumor lesions, No. (%)	
Adenocarcinoma <i>in situ</i>	3 (1.9%)
Microinvasive adenocarcinoma	5 (3.1%)
Adenocarcinoma	154 (95.1%)
Pathological stage of the sequenced tumors, No. (%)	
0	3 (1.9%)
IA1	39 (24.1%)
IA2	38 (23.5%)
IA3	6 (3.7%)
IB	62 (38.3%)
IIB	5 (3.1%)
IIIA	8 (4.9%)
IIIB	1 (0.6%)
EGFR-TKI treatment, No. (%)	
Yes	15 (9.3%)
No	147 (90.7%)

and p.746\_751del (1, 0.62%); exon 20, p.V769indelsVASV (2, 1.23%), p.S768I (3, 1.85%), p.S768N (1, 0.62%), p.T790M (1, 0.62%), and p.Q791\* (1, 0.62%); and exon 21, p.V843I (1, 0.62%) and p.L858R (4, 2.74%) (Figure 3B,C, Table 2 and Table S1). As shown in Table 2, the total frequency of mutations in nontumor tissues was 9.88%.

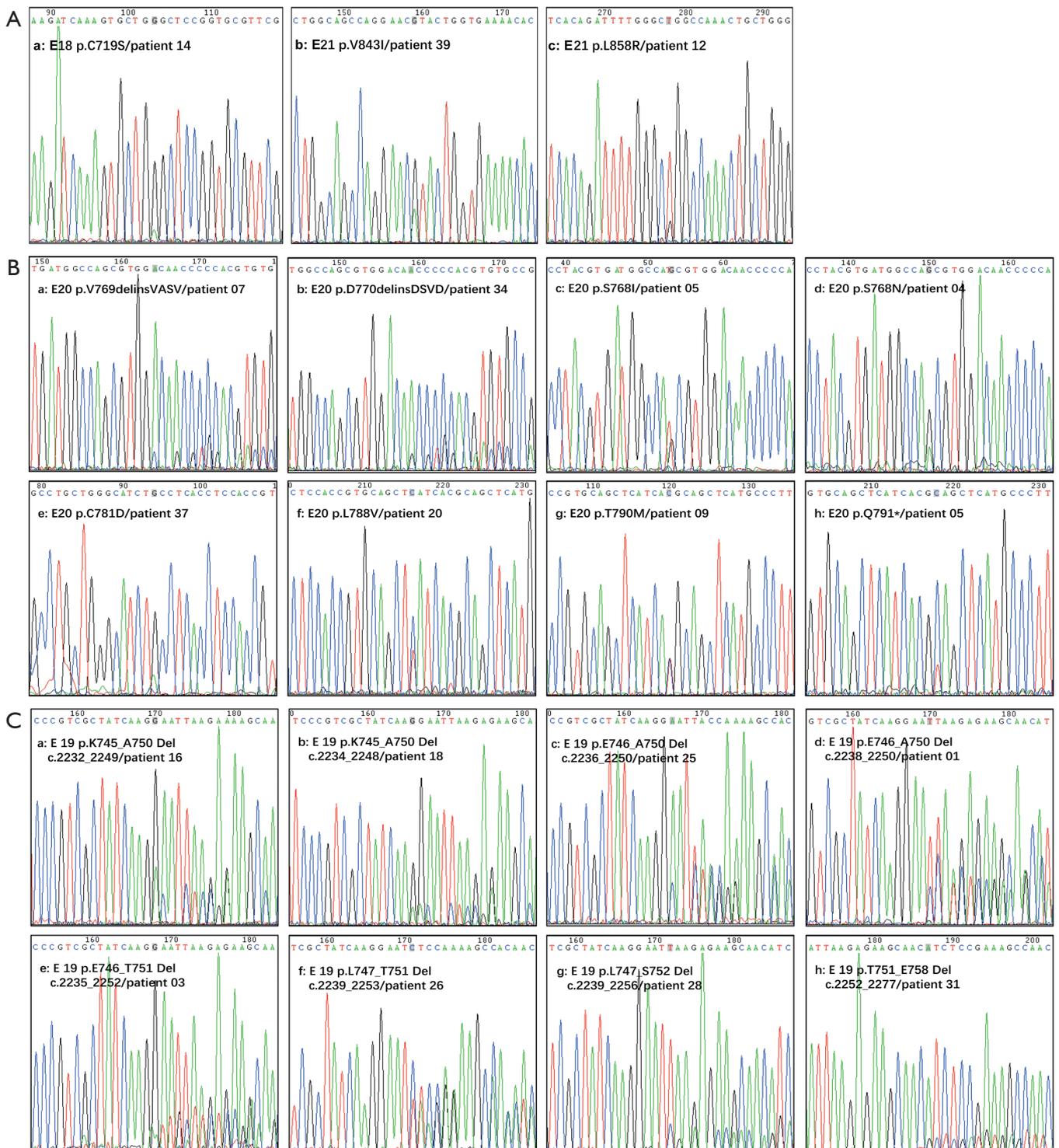
#### ***Distribution of somatic EGFR mutations in 162 MPLC patients***

PCR sequencing of 366 tumor lesion samples for somatic *EGFR* driver mutations was also carried out. The results showed 74 different heterozygous mutations, including exon 18 p.G719S (1, 0.62%), 19 p.dels (21, 12.35%), 20 p.insX (8, 4.94%), 20 p.S768I (3, 1.85%), 20 p.S768N (1, 0.62%), 20 p.C781Y (1, 0.62%), 20 p.L788V (1, 0.62%), 20 p.T790M (3, 1.85%), 20 p.Q791\* (1, 0.62%), 21 p.V843I (2, 1.23%) and 21 p.L858R (32, 19.75%) (Figure 3A,B,C and Table 2). Of the 74 mutations, multiple mutations were observed in eight patients, and repeated counting was avoided. A total of 63 MPLC patients among initial 162 (38.89%) were identified as having somatic mutations (duplications removed) (Table S1). We performed a stratified analysis to assess sex, smoking status, number of lung cancer patients in first relatives and pathological stage. Results showed smoking status impacted on somatic mutant rate ( $P < 0.05$ , Table 3 and Table S2).

The detailed information for the patients with coexisting mutations was as follows: patient 20 (exon 19 p.745\_750del and 20 p.L788V), patient 24 (exon 19 p.746\_750del and 21 p.L858R), patient 1 (exon 19 p.746\_750del and 20 p.S768I), patient 3 (exon 19 p.746\_751del, 20 p.T790M and 21 p.L858R), patient 7 (exon 20 p.V769indelsVASV and 21 p.L858R), patient 4 (exon 20 p.V769indelsVASV, 20 p.S768N and 21 p.L858R), patient 5 (exon 20 p.S768I and 20 p.Q791\*) and patient 9 (exon 20 p.T790M and 21 p.L858R) (Table S1). We also analyzed the mutation presentation in the 366 tumor lesion samples. The results showed low consistency between lesions individually, and only 9 cases showed consistency (Tables S1,S2).

#### ***Comparison of EGFR somatic mutations in 162 MPLC probands and in 3243 LUAD patients***

In total, 3243 sporadic LUAD tissue samples were previously subjected to targeted sequencing using the lung cancer ten-gene panel (Geneis Co., Ltd.). We reanalyzed the



**Figure 3** Sequencing peak maps of EGFR mutations in exons 18, 19, 20 and 21. (A-a) exon 18 p.C719S (c.2155G>A); (A-b) exon 21 p.V843I (c.2527 G>A); (A-c) exon 21 p.L858R (c.2573 T>G). (B-a/b) exon 19 p.K745\_A750 del; (B-c/d) exon 19 p.E746\_A750 del; (B-e) exon 19 p.E746\_T751del; (B-f) exon 19 p.L747\_T751del; (B-g) exon 19 p.L747\_S752del; (B-h) exon 19 p.T751\_E758del. (C-a) exon 20 p.V769indelsVASV; (C-b) exon 20 p.D770DSVD; (C-c) exon 20 p.S768I (c.2303G>T); (C-d) exon 20 p.S768N (c.2303G>A); (C-e) exon 20 p.C781Y (c.2339G>A); (C-f) exon 20 p.L788V (c.2364C>T); (C-g) exon 20 p.T790M (2369 C>T); (C-h) exon 20 p.Q791\* (c.2371 C>T).

**Table 2** The frequencies of EGFR mutations in 162 MPLC probands

Exon	Type	CGMF in 162 MPLC probands (%)	SMF in 162 MPLC probands (%)	SMF in 3,243 lung patients (%)
Total		16 (9.88)	63 (38.89), 73 (45.06)*	1,814 (55.94)*
18	p.G719A	0	0	42 (1.30)
	p.G719S	0	1 (0.62)	10 (0.31)
	p.G719C	0	0	26 (0.80)
19del	p.746_749del	0	0	43 (1.33)
	p.747_753del	0	0	78 (2.14)
	p.745_750del	0	8 (4.94)	296 (9.13)
	p.746_750del	2 (1.23)	6 (3.70)	154 (4.75)
	p.747_751del	0	2 (1.23)	63 (1.94)
	p.746_751del	1 (0.62)	1 (0.62)	14 (0.43)
	p.751_758del	0	1 (0.62)	2 (0.06)
	p.747_752del	0	3 (1.85)*	11 (0.34)
	p.741_741del	0	0	0 (0)
	p.755_756del	0	0	0 (0)
	p.A743delinsAIKIPVD	0	0	0 (0)
	p.746_752del	0	0	2 (0.06)
	p.749_754del	0	0	2 (0.06)
	p.746_746del	0	0	1 (0.03)
	p.752_759del	0	0	2 (0.06)
	p.745_748del	0	0	0 (0)
20indels	p.M766delinsMASM	0	0	0 (0)
	p.A767delinsASVA	0	0	0 (0)
	p.V769delinsVASV	2 (1.23)	4 (2.47)***	3 (0.09)
	p.D770delinsDSVD	0	4 (2.47)***	0 (0)
	p.N771delinsNPHPH	0	0	1 (0.03)
	p.P772delinsPP	0	0	0 (0)
	p.H773delinsHH	0	0	0 (0)
	p.H773delinsHNPH	0	0	0 (0)
20	p.S768I	3 (1.85)	3 (1.85)	36 (1.11)
	p.S768N	1 (0.62)	1 (0.62)	0 (0)
	p.R776H	0	0	4 (0.12)
	p.R776G	0	0	0 (0)
	p.R776C	0	0	8 (0.25)
	p.L788V	0	1 (0.62)	0 (0)
	p.C781Y	0	1 (0.62)	0 (0)

**Table 2** (continued)

**Table 2** (continued)

Exon	Type	CGMF in 162 MPLC probands (%)	SMF in 162 MPLC probands (%)	SMF in 3,243 lung patients (%)
	p.T790M	1 (0.62)	3 (1.85)	68 (2.10)
	Q791*	1 (0.62)	1 (0.62)	1 (0.03)
	p.C797S	0	0	0 (0)
21	p.V843I	1 (0.62)	2 (1.23)**	0 (0)
	p.L858Q	0	0	0 (0)
	p.L858R	4 (2.47)*	32 (19.75)*	911 (28.09)
	p.L861Q	0	0	36 (1.11)
	p.L861G	0	0	0 (0)

\*, significantly different for SMFs in 162 MPLC probands compared to those in 3,243 LUAD patients, which were calculated using the Fisher's exact test, P<0.05; \*\*, significantly different for SMFs in 162 MPLC probands compared to those in 3,243 LUAD patients, which were calculated using the Fisher's exact test, P<0.01; \*\*\*, significantly different for SMFs in 162 MPLC probands compared to those in 3,243 LUAD patients, which were calculated using the Fisher's exact test, P<0.001; \*, duplication included. CGMF, candidate germline mutation frequencies; SMF, somatic mutation frequencies; Q791\*, nonsense mutation at the 791 site.

**Table 3** Information regarding EGFR mutations in 162 MPLC patients

Characteristic	CGM (16/162)	P value	SM (73/162)	P value
Sex, No. (%)		0.585 <sup>a</sup>		0.101 <sup>a</sup>
Male	7/58 (12.07)		21/58 (36.21)	
Female	9/104 (8.65)		52/104 (50.00)	
Smoking history, n (%), No. (%)		0.769 <sup>a</sup>		0.020 <sup>a*</sup>
Never smokers	11/118 (9.32)		60/118 (50.85)	
Heavy smokers (pack-year >20)	5/44 (11.36)		13/44 (29.55)	
Number of lung cancer patients in first relatives, No. (%)		<0.001 <sup>b***</sup>		0.184 <sup>b</sup>
1	8/136 (5.88)		57/136 (41.91)	
2	6/21 (28.57)		13/21 (61.90)	
3	2/5 (40.00)		3/5 (60.00)	
Pathological stage of the sequenced tumors, No. (%)		0.029 <sup>b*</sup>		0.474 <sup>b</sup>
0	0/3 (0.00)		0/3 (0.00)	
IA1	3/39 (7.69)		15/39 (38.46)	
IA2	6/38 (15.79)		19/38 (50.00)	
IA3	0/6 (0.00)		3/6 (50.00)	
IB	5/62 (8.06)		29/62 (46.77)	
IIB	0/5 (0.00)		3/5 (60.00)	
IIIA	0/8 (0.00)		2/8 (25.00)	
IIIB	2/1 (100.00) <sup>c</sup>		2/1 (100.00) <sup>c</sup>	

<sup>a</sup>, Mann-Whitney U test; <sup>b</sup>, Kruskal-Wallis test; <sup>c</sup>, T mutations/1 patient. \*, significantly different between groups, P<0.05; \*\*\*, significantly different between groups, P<0.001. CGM, candidate germline mutation; SM, somatic mutation.

data and identified the *EGFR* somatic mutation frequencies (threshold of  $\geq 1\%$ ) of exon 18 p.G719X (78, 2.41%), 19 p.dels (668, 20.60%), 20 p.indels (4, 0.12%), 20 p.S768I (36, 1.11%), 20 p.R776H/G/C (12, 0.37%), 20 p.T790M (68, 2.10%), 21 p.L858R (911, 28.06%) and 21 p.L861Q (36, 1.11%). These results are summarized in *Table 2*. First, as the most common mutations, the frequencies of exon 19 p.dels and 21 p.L858R mutation were indeed the highest in the two cohorts. However, there were significant differences in several mutations: exon 19 p.747-752del, 20 p.V769indelsVASV, 20 p.D770indelsDSVD, 21 p.V843I and 21 p.L858R. Specifically, the chi-square test showed the exon 19 p.747\_752del, 20 p.V769indelsVASV, 20 p.D770indelsDSVD and 21 p.V843I mutations occurred at significantly higher frequencies in the MPLC population (1.85%, 2.47%, 2.47% and 1.23%, respectively) than in the LUAD population (0.34%, 0.09%, and 0.00%, respectively; P value  $< 0.05$ ) (*Figure S1*).

## Discussion

Lung cancer is a major cause of death in China and other countries. Genetic factors play a major role in the development of lung cancer, especially familial lung cancer (21). Although there have been many studies addressing the molecular characteristics of MPLC in recent years, the number of cases involved in these studies has not been higher than 100 (4,22,23). Moreover, only several case reports have documented germline mutations in MPLC (9-13). Our study, for the first time, reported candidate germline mutations and tumor somatic mutations in a large-scale cohort of MPLC probands. In all, 162 patients with MPLCs and a first-degree relative with lung cancer were identified, which is a very impressive cohort size that can only be achieved at a high-volume thoracic surgical center in a region with a high incidence of *EGFR* mutant lung cancer.

Germline mutations in *EGFR* are rare but may contribute to oncogenesis (7-13,24-26). A larger study on the p.T790M germline mutation in 31,414 patients reported that 48 patients (0.15%) carried the germline p.T790M mutation (27). This mutation is one of the best studied germline mutations, which has been previously estimated to be involved in 1% to 4% of *EGFR*-positive lung cancers and implicated in hereditary lung cancer (28). The exon 20 p.R776H germline mutation is of special concern in MPLC patients according to our previous results (11); interestingly, the p.R776H/G/C mutation was

not detected in any of the family probands analyzed in this study. Here, we systematically investigated the signature of *EGFR* mutations in a population of MPLC families. The most important result was that several candidate germline mutations with frequencies above 1% were detected. These candidate germline mutations are as follows: exon 19 p.746\_750del (2, 1.23%), 20 p.V769indelsVASV (2, 1.23%), 20 p.S768I (3, 1.85%), and 21 p.L858R (4, 2.74%). Exon 20 exhibited the highest frequency of mutations in NATs (8/16, 50%). Exon 20 encodes the third most predominant kinase domain, and the mutation class encompasses in-frame insertions and indels following or within the regulatory C-helix amino acids (29,30). These oncogene variants engage the phosphoinositide 3-kinase/protein kinase B and extracellular-signal-regulated kinase/mitogen-activated protein kinase downstream pathways with *EGFR* dependency (30). Although the *EGFR* p.S768I mutation is considered to be highly rare, we detected it in a total of 1.85% (3/162) of the probands with MPLC. It is worth emphasizing that all p.S768I mutations identified were candidate germline mutations. All of the above mentioned candidate germline mutations might be responsible for the etiology of MPLC.

Predictive biomarker testing for *EGFR* point mutations and indels is the standard of care in NSCLC patients for the selection of an optimal first-line therapy (31,32). However, for MPLC patients, an *EGFR* mutation spectrum to guide clinical medication is lacking. In our study, 162 probands with 366 tumor lesions were screened for *EGFR* somatic mutations. This study is the first to report exon 18 p.G719S (1, 0.62%); exon 19 p.dels (21, 12.35%); exon 20 p.ins (8, 4.94%), p.S768I (3, 1.85%), p.S768N (1, 0.62%), p.C781Y (1, 0.62%), p.L788V (1, 0.62%), p.T790M (3, 1.85%) and p.Q791\* (1, 0.62%); and exon 21 p.V843I (2, 1.23%) and p.L858R (32, 19.75%) mutations in family numbers with MPLC. Exon 19 deletions and exon 21 p.L858R were most common in the MPLC family population. Genetic heterogeneity is a significant molecular genetic feature of MPLC. High-throughput sequencing analysis has previously shown high heterogeneity among multiple lesions in the same MPLC patient (4,33,34). Heterogeneity was also detected in our population, and only 9 patients in the 162 MPLC probands showed consistency between tumor lesions.

In China and other Asian countries, female never-smokers with an *EGFR* mutation account for a large percentage (40-50%) of LUAD cases (35-38). More than 60% of females and nonsmokers were in our cohort,

and the total *EGFR* mutation rate was 45.06% (73/163) (duplicates included) (Table 2). Smoking status and gender are important factors for *EGFR* mutation patterns, and we performed a stratified analysis to assess sex, smoking status, number of lung cancer patients in first relatives and pathological stage. We found no difference in smoking status between the two cohorts, and the proportion of female patients in the MPLC cohort was significantly higher (Table S3). However, whether gender differences are the fundamental cause of higher *EGFR* mutations remains to be further verified. In addition, it is interesting that only some specific site mutation frequencies (Figure S1) rather than all sites in the MPLC population are higher (Table 2), suggesting that they may have more value in clinical applications.

In the current study, we found that the frequencies of exon 19 p.747-752del, 20 p.V769indelsVASV, 20 p.D770indelsDSVD and 21 p.V843I were significantly higher in the MPLC probands (Figure S1). We performed functional prediction with SIFT and PROVEAN, and the results showed that the mutated sites were probably damaging/deleterious (Table S4). Nine nucleotide insertion mutations (p.V769indelsVASV and p.D770indelsDSVD) in exon 20 of *EGFR* were detected in eight patients with MPLC (Figure 3B). The frequencies of exon 20 p.indels were 4.94% in our family population, 0.09% in the 3243 patients with sporadic LUAD and 0.31% (88/28,768) in the COSMIC database. This implies that these mutations might be related to the occurrence of familial MPLC. The effect of EGFR-TKIs has been examined in cultured cell lines transfected with exon 20 insertion mutations. The 50% inhibitory concentration (IC<sub>50</sub>) of osimertinib was lower than that of other EGFR-TKIs in all cell lines tested (39). Clinical trials are currently underway to examine the effects of several EGFR-TKIs on NSCLC with confirmed exon 20 insertion mutations (UMIN000031929, ClinicalTrials.gov Identifier: NCT03066206, and ClinicalTrials.gov Identifier: NCT03414814). We hope that these results will establish evidence that osimertinib and other EGFR-TKIs are effective for NSCLC positive for exon 20 insertion mutations.

There are several limitations to our study. First, the distribution of *EGFR* mutations in the 162 MPLC cases was identified by PCR-based Sanger sequencing; however, the 3243 sporadic adenocarcinoma samples had been characterized by next-generation sequencing (NGS) with a ten-gene panel. The PCR-Sanger is highly effective and accurate for sequencing single targeted genes and

a single reaction can easily yield DNA sequence data between 500 and 1,000 base pairs in length. One major issue of this technology can not suitable for detecting low-frequency mutations for low sensitivity (40). *EGFR* has a high-frequency mutation in Chinese, and our sample size is relatively large (162 probands). NGS has advantages in high sensitivity and multiplexed testing. However, the quality of the results is dependent on sample type, method, sequencing platform and bioinformatic pipeline and finally interpretation by a team of molecular and genetic experts (40). The most important results in the study is that the rate of *EGFR* mutations in exon 19 p.747-752del, 20 p.V769indelsVASV and 20 p.D770indelsDSVD were significantly higher in MPLC probands detected with sanger than those in sporadic LUAD patients detected with NGS. For the low sensitivity of sanger sequencing compared with NGS, this result gives more stronger evidence that mutations in exon 19 p.747-752del, 20 p.V769indelsVASV and 20 p.D770indelsDSVD may play significant roles in family MPLC. Second, candidate germline mutations were identified in normal tissue adjacent to cancerous tissue. A recent study emphasized that blood is superior to these adjacent normal tissues (41), but NATs were used for identification of candidate germline mutations in our study due to the lack of peripheral blood samples. Third, the frequency of *EGFR* mutation in NSCLC has been described to differ across ethnic groups and the occurrence was shown to be evidently higher in East-Asians compared with that in Europeans (42,43). The two cohorts involved in the study are all from East Asian populations, and this conclusion should be carefully interpreted when applied to other ethnicities.

## Conclusions

This study first showed the unique *EGFR* mutation signatures in MPLC probands. We also first screened high-frequency *EGFR* candidate germline mutations, such as exon 19 p.del, 20 p.ins, 20 p.S768X and 21 p.L861R. Compared with the LUAD patients, MPLC probands had a high frequency of pathogenic somatic mutations, including exon 19 p.747-752del and 20 p.V769indelsVASV and p.D770indelsDSVD. These findings might provide evidence that relates to the etiology of these mutations and shed light on the clinical treatment of high-risk patients.

## Acknowledgments

*Funding:* This work was supported by the National Key

R&D Program of China (2018YFC1312100), the National Natural Science Foundation of China (81972196), The CAMS Innovation Fund for Medical Sciences (CIFMS) (2017-I2M-1-005, 2019-I2M-2-002), The Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2018PT32033), The Innovation team development project of Ministry of Education (IRT\_17R10).

## Footnote

*Reporting Checklist:* The authors have completed the STROBE reporting checklist. Available at <http://dx.doi.org/10.21037/tlcr-20-1001>

*Data Sharing Statement:* Available at <http://dx.doi.org/10.21037/tlcr-20-1001>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tlcr-20-1001>). The authors have no conflicts of interest to declare.

*Ethical Statement:* 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. All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013). Ethical approval was obtained from the medical ethics committee of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (NCC2018AA-015). And all the patients provided written informed consent.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

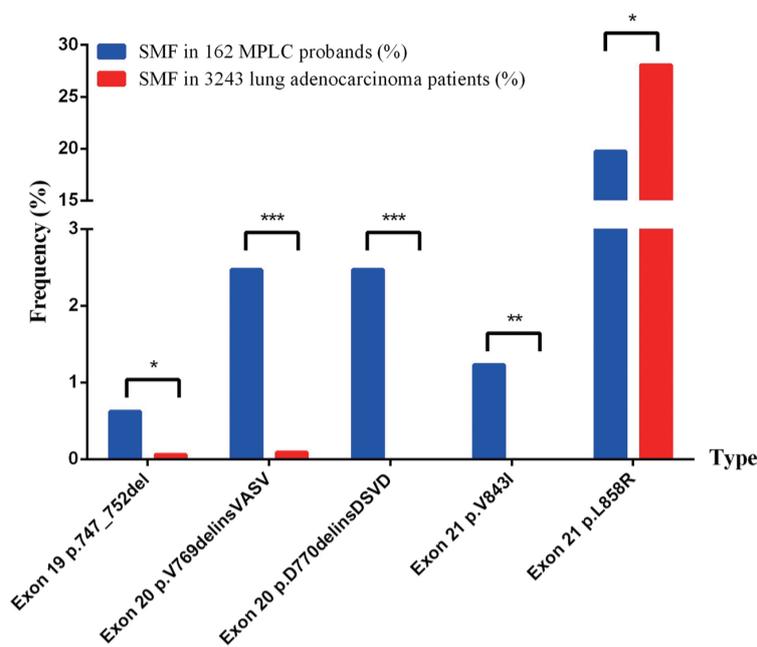
## References

- Martini N, Melamed MR. Multiple primary lung cancers. *J Thorac Cardiovasc Surg* 1975;70:606-12.
- Jung EJ, Lee JH, Jeon K, et al. Treatment outcomes for patients with synchronous multiple primary non-small cell lung cancer. *Lung Cancer* 2011;73:237-42.
- Detterbeck FC, Bolejack V, Arenberg DA, et al. The IASLC Lung Cancer Staging Project: Background Data and Proposals for the Classification of Lung Cancer with Separate Tumor Nodules in the Forthcoming Eighth Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* 2016;11:681-92.
- Park E, Ahn S, Kim H, et al. Targeted Sequencing Analysis of Pulmonary Adenocarcinoma with Multiple Synchronous Ground-Glass/Lepidic Nodules. *J Thorac Oncol* 2018;13:1776-83.
- Li X, Hemminki K. Familial multiple primary lung cancers: a population-based analysis from Sweden. *Lung Cancer* 2005;47:301-7.
- Zhang Z, Gao S, Mao Y, et al. Surgical Outcomes of Synchronous Multiple Primary Non-Small Cell Lung Cancers. *Sci Rep* 2016;6:23252.
- Prudkin L, Tang X, Wistuba, II. Germ-line and somatic presentations of the EGFR T790M mutation in lung cancer. *J Thorac Oncol* 2009;4:139-41.
- Tibaldi C, Giovannetti E, Vasile E, et al. Inherited germline T790M mutation and somatic epidermal growth factor receptor mutations in non-small cell lung cancer patients. *J Thorac Oncol* 2011;6:395-6.
- Oxnard GR, Miller VA, Robson ME, et al. Screening for germline EGFR T790M mutations through lung cancer genotyping. *J Thorac Oncol* 2012;7:1049-52.
- van Noesel J, van der Ven WH, van Os TA, et al. Activating germline R776H mutation in the epidermal growth factor receptor associated with lung cancer with squamous differentiation. *J Clin Oncol* 2013;31:e161-4.
- Su K, Gao S, Ying J, et al. Sequencing a super multiple synchronous lung cancer reveals a novel variant in driver gene ARID1B. *J Thorac Cardiovasc Surg* 2018;155:e185-91.
- Ikeda K, Nomori H, Mori T, et al. Novel germline mutation: EGFR V843I in patient with multiple lung adenocarcinomas and family members with lung cancer. *Ann Thorac Surg* 2008;85:1430-2.
- Ohtsuka K, Ohnishi H, Kurai D, et al. Familial lung adenocarcinoma caused by the EGFR V843I germ-line mutation. *J Clin Oncol* 2011;29:e191-2.
- Bell DW, Gore I, Okimoto RA, et al. Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. *Nat Genet* 2005;37:1315-6.
- Lou Y, Pecot CV, Tran HT, et al. Germline Mutation

- of T790M and Dual/Multiple EGFR Mutations in Patients With Lung Adenocarcinoma. *Clin Lung Cancer* 2016;17:e5-11.
16. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129-39.
  17. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497-500.
  18. Pao W, Miller V, Zakowski M, et al. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci U S A* 2004;101:13306-11.
  19. Costa DB. Kinase inhibitor-responsive genotypes in EGFR mutated lung adenocarcinomas: moving past common point mutations or indels into uncommon kinase domain duplications and rearrangements. *Transl Lung Cancer Res* 2016;5:331-7.
  20. Meng H, Guo X, Sun D, et al. Genomic Profiling of Driver Gene Mutations in Chinese Patients With Non-Small Cell Lung Cancer. *Front Genet* 2019;10:1008.
  21. Tomoshige K, Matsumoto K, Tsuchiya T, et al. Germline mutations causing familial lung cancer. *J Hum Genet* 2015;60:597-603.
  22. Chung JH, Choe G, Jheon S, et al. Epidermal growth factor receptor mutation and pathologic-radiologic correlation between multiple lung nodules with ground-glass opacity differentiates multicentric origin from intrapulmonary spread. *J Thorac Oncol* 2009;4:1490-5.
  23. Asmar R, Sonett JR, Singh G, et al. Use of Oncogenic Driver Mutations in Staging of Multiple Primary Lung Carcinomas: A Single-Center Experience. *J Thorac Oncol* 2017;12:1524-35.
  24. Thomas A, Xi L, Carter CA, et al. Concurrent molecular alterations in tumors with germ line epidermal growth factor receptor T790M mutations. *Clin Lung Cancer* 2013;14:452-6.
  25. Gazdar A, Robinson L, Oliver D, et al. Hereditary lung cancer syndrome targets never smokers with germline EGFR gene T790M mutations. *J Thorac Oncol* 2014;9:456-63.
  26. Girard N, Lou E, Azzoli CG, et al. Analysis of genetic variants in never-smokers with lung cancer facilitated by an Internet-based blood collection protocol: a preliminary report. *Clin Cancer Res* 2010;16:755-63.
  27. Hu Y, Alden RS, Odegaard JI, et al. Discrimination of Germline EGFR T790M Mutations in Plasma Cell-Free DNA Allows Study of Prevalence Across 31,414 Cancer Patients. *Clin Cancer Res* 2017;23:7351-9.
  28. 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.
  29. Yasuda H, Kobayashi S, Costa DB. EGFR exon 20 insertion mutations in non-small-cell lung cancer: preclinical data and clinical implications. *Lancet Oncol* 2012;13:e23-31.
  30. Yasuda H, Park E, Yun CH, et al. Structural, biochemical, and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer. *Sci Transl Med* 2013;5:216ra177.
  31. Lindeman NI, Cagle PT, Beasley MB, et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *J Thorac Oncol* 2013;8:823-59.
  32. Leighl NB, Rekhtman N, Biermann WA, et al. Molecular testing for selection of patients with lung cancer for epidermal growth factor receptor and anaplastic lymphoma kinase tyrosine kinase inhibitors: American Society of Clinical Oncology endorsement of the College of American Pathologists/International Association for the study of lung cancer/association for molecular pathology guideline. *J Clin Oncol* 2014;32:3673-9.
  33. Wu C, Zhao C, Yang Y, et al. High Discrepancy of Driver Mutations in Patients with NSCLC and Synchronous Multiple Lung Ground-Glass Nodules. *J Thorac Oncol* 2015;10:778-83.
  34. Liu Y, Zhang J, Li L, et al. Genomic heterogeneity of multiple synchronous lung cancer. *Nat Commun* 2016;7:13200.
  35. Chen J, Yang H, Teo ASM, et al. Genomic landscape of lung adenocarcinoma in East Asians. *Nat Genet* 2020;52:177-86.
  36. Nahar R, Zhai W, Zhang T, et al. Elucidating the genomic architecture of Asian EGFR-mutant lung adenocarcinoma through multi-region exome sequencing. *Nat Commun* 2018;9:216.
  37. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers--a different disease. *Nat Rev Cancer* 2007;7:778-90.
  38. Wang C, Yin R, Dai J, et al. Whole-genome sequencing reveals genomic signatures associated with the inflammatory microenvironments in Chinese NSCLC

- patients. *Nat Commun* 2018;9:2054.
39. Hirano T, Yasuda H, Tani T, et al. In vitro modeling to determine mutation specificity of EGFR tyrosine kinase inhibitors against clinically relevant EGFR mutants in non-small-cell lung cancer. *Oncotarget* 2015;6:38789-803.
  40. Sridhar K, Singh A, Butzmann A, et al. Molecular genetic testing methodologies in hematopoietic diseases: current and future methods. *Int J Lab Hematol* 2019;41 Suppl 1:102-16.
  41. Yizhak K, Aguet F, Kim J, et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* 2019;364:eaaw0726.
  42. Smits AJ, Kummer JA, Hinrichs JW, et al. EGFR and KRAS mutations in lung carcinomas in the Dutch population: increased EGFR mutation frequency in malignant pleural effusion of lung adenocarcinoma. *Cell Oncol (Dordr)* 2012;35:189-96.
  43. Soh J, Toyooka S, Matsuo K, et al. Ethnicity affects EGFR and KRAS gene alterations of lung adenocarcinoma. *Oncol Lett* 2015;10:1775-82.

**Cite this article as:** Li C, Wang Y, Su K, Liu Y, Wang L, Zheng B, Yan N, Yuan D, Zhang Y, Xue L, Gao S, He J. Presentation of *EGFR* mutations in 162 family probands with multiple primary lung cancer. *Transl Lung Cancer Res* 2021;10(4):1734-1746. doi: 10.21037/tlcr-20-1001



**Figure S1** Comparisons of somatic mutation frequencies. Significant differences between the 162 MPLC probands and 3,243 LUAD patients are shown (Mann-Whitney U test). One star represents a P value <0.05, two stars represent a P value <0.01, and three stars represent a P value <0.001.

**Table S1** The frequencies of *EGFR* mutations in 366 MPLC tumor lesions

Exon	Type	SMF in 366 MPLC lesions (%)
Total		93 (25.41) (102-9)
18	p.G719A	0
	p.G719S	1 (0.27)
	p.G719C	0
19del	p.746_749del	0
	p.747_753del	0
	p.745_750del	10 (2.73)
	p.746_750del	9 (2.46)
	p.747_751del	2 (0.55)
	p.746_751del	2 (0.55)
	p.751_758del	1 (0.27)
	p.747_752del	4 (1.09)
	p.741_741del	0
	p.755_756del	0
	p.A743delinsAIKIPVD	0
	p.746_752del	0
	p.749_754del	0
	p.746_746del	0
	p.752_759del	0
	p.745_748del	0
	p.M766delinsMASM	0
20	p.A767delinsASVA	0
	p.V769delinsVASV	7 (1.91)
	p.D770delinsDSVD	5 (1.37)
	p.N771delinsNPHPH	0
	p.P772delinsPP	0
	p.H773delinsHH	0
	p.H773delinsHNPH	0
	p.S768I	7 (1.91)
	p.S768N	2 (0.55)
	p.R776H	0
	p.R776G	0
	p.R776C	0
	p.L788V	1 (0.27)
	p.C781Y	1 (0.27)
	p.T790M	4 (1.09)
Q791*	1 (0.27)	
p.C797S	0	
p.V843I	4 (1.09)	
21	p.L858Q	0
	p.L858R	42 (11.48)
	p.L861Q	0
	p.L861G	0

SMF, somatic mutation frequencies.

**Table S2** Information regarding smoking status and sex in 162 MPLC and 3,243 LUAD patients

	162 MPLC		3,243 sporadic LUAD		Chi-square test
	Smoker	Nonsmoker	Smoker	Nonsmoker	
Smoking status	44/162 (27.16%)	118/162 (72.84%)	1118/3243 (34.47%)	2125/3243 (65.53%)	P=0.055
<i>EGFR</i> SMF*	13/44 (29.55%)	61/118 (51.69%)	453/1118 (40.52%)	1361/2125 (64.05%)	
<i>EGFR</i> SMF	11/44 (25.00%)	52/118 (44.07%)			
	Male	Female	Male	Female	
Sex	58/162 (35.80%)	104/162 (64.20%)	1687/3243 (52.02%)	1556 (47.98%)	P<0.001
<i>EGFR</i> SMF*	21/58 (36.21%)	53/104 (50.96%)	689/1687 (40.84%)	1095/1556 (72.30%)	
<i>EGFR</i> SMF	16/58 (27.59%)	47/104 (45.19%)			

\*, duplication included. MPLC, multiple primary lung cancer; LUAD, lung adenocarcinoma; SMF, somatic mutation frequencies.

**Table S3** Information regarding *EGFR* mutations in 162 MPLC patients

Exon	Type	NT (%)	Sample	Case No.	Sex	NFR	Smoke	Stage
Germine	Total	16 (9.88)						
19	p.746_750del	1	NAT	01	F	1	N	IB
	p.746_750del	1	NAT	02	F	1	N	IA1
19	p.746_751del	1	NAT	03	F	1	N	IB
20	p.S768N	1	NAT	04	F	1	N	IA1
	p.S768I	1	NAT	05	M	2	Y	IA2
	p.S768I	1	NAT	06	F	2	N	IA2
	p.S768I	1	NAT	01	F	1	N	IB
20 ins	p.V769_D770 delins ASV	1	NAT	07	M	2	N	IA1
	p.V769_D770 delins ASV	1	NAT	08	M	2	Y	IA2
20	p.T790M	1	NAT	09	M	3	Y	IIIB
20	p.Q791*	1	NAT	05	M	2	Y	IA2
21	p.V843I(G>A)	1	NAT	10	F	2	N	IB
	L858R	1	NAT	11	F	1	N	IA2
	L858R	1	NAT	12	M	1	N	IB
	L858R	1	NAT	09	M	3	Y	IIIB
	L858R	1	NAT	13	F	1	N	IA2
Somatic	Total	73 (45.06)						
18	C719S	1	Tumor lesion	14	F	1	N	IB
	Total (19)	20 (12.35)						
19del	p.745_750del	1	Tumor lesion	15	F	1	N	IB
	p.745_750del	1	Tumor lesion 1		F	1	N	IA2
			Tumor lesion 2	16				
	p.745_750del	1	Tumor lesion	17	F	1	N	IA2
	p.745_750del	1	Tumor lesion	18	F	1	N	IB
	p.745_750del	1	Tumor lesion	19	M	1	Y	IA1
	p.745_750del	1	Tumor lesion	20	F	1	N	IA2
	p.745_750del	1	Tumor lesion 1		F	2	N	IIB
			Tumor lesion 2	21				
	p.745_750del	1	Tumor lesion	22	F	1	N	IB
	p.746_750del	1	Tumor lesion	23	F	1	N	IIIA
	p.746_750del	1	Tumor lesion 1	2	F	1	N	IA1
			Tumor lesion 2					
	p.746_750del	1	Tumor lesion	24	F	1	N	IB
	p.746_750del	1	Tumor lesion	25	F	1	N	IB
	p.746_750del	1	Tumor lesion 1		F	1	N	IB
			Tumor lesion 2	1				
	p.746_751del	1	Tumor lesion 1		M	1	N	IB
			Tumor lesion 2	3				
	p.747_751del	1	Tumor lesion	26	M	1	Y	IB
	p.747_751del	1	Tumor lesion	27	F	1	N	IIB
	p.747_752del	1	Tumor lesion	28	F	1	N	IA1
	p.747_752del	1	Tumor lesion	29	F	1	N	IB
	p.747_752del	1	Tumor lesion 1		F	1	N	IB
			Tumor lesion 2	30				
	p.751_758del	1	Tumor lesion	31	F	1	N	IA1
	Total (20)	18 (11.11)?						
20ins	p.V769delinsVASV	1	Tumor lesion	32	F	1	N	IB
	p.V769delinsVASV	1	Tumor lesion 1		M	2	N	IA1
			Tumor lesion 2	7				
	p.V769delinsVASV	1	Tumor lesion 1		M	2	Y	IA2
			Tumor lesion 2	8				
	p.V769delinsVASV	1	Tumor lesion 1		F	1	N	IA1
			Tumor lesion 2	4				
	p.D770delinsDSVD	1	Tumor lesion	33	F	1	N	IA1
	p.D770delinsDSVD	1	Tumor lesion 1		F	3	N	IB
			Tumor lesion 2	34				
	p.D770delinsDSVD	1	Tumor lesion	35	F	1	N	IA2
	p.D770delinsDSVD	1	Tumor lesion	36	F	1	N	IB
20	p.S768N	1	Tumor lesion 1		F	1	N	IA1
			Tumor lesion 2	4				
	p.S768I	1	Tumor lesion 1		M	2	Y	IA3
			Tumor lesion 2	5				
	p.S768I	1	Tumor lesion 1		F	2	N	IA2
			Tumor lesion 2	6				
	p.S768I	1	Tumor lesion 1		F	1	N	IB
			Tumor lesion 2	1				
	p.C781Y	1	Tumor lesion	37	M	1	N	IIB
	p.L788V	1	Tumor lesion	20	F	1	N	IA2
	p.T790M	1	Tumor lesion	38	F	1	Y	IA1
	p.T790M	1	Tumor lesion 1		M	3	Y	IIIB
			Tumor lesion 2	9				
	p.T790M	1	Tumor lesion	3	M	1	N	IB
	p.Q791*	1	Tumor lesion	5	M	2	Y	IA3
	Total (21)	34 (20.99)						
21	p.V843I(G>A)	1	Tumor lesion 1		F	2	N	IB
			Tumor lesion 2					
			Tumor lesion 3	10				
	p.V843I(G>A)	1	Tumor lesion	39	M	1	N	IIIA
	p.L858R	1	Tumor lesion	40	M	2	Y	IB
	p.L858R	1	Tumor lesion	41	F	1	N	IA2
	p.L858R	1	Tumor lesion 1		F	1	N	IA2
			Tumor lesion 2	11				
	p.L858R	1	Tumor lesion	42	F	1	N	IB
	p.L858R	1	Tumor lesion	43	F	1	N	IA2
	p.L858R	1	Tumor lesion	44	F	1	N	IA2
	p.L858R	1	Tumor lesion 1		M	1	Y	IB
			Tumor lesion 2	45				
	p.L858R	1	Tumor lesion	46	M	1	N	IB
	p.L858R	1	Tumor lesion 1		M	1	N	IB
			Tumor lesion 2	12				
	p.L858R	1	Tumor lesion	47	F	1	N	IA2
	p.L858R	1	Tumor lesion	48	F	1	N	IB
	p.L858R	1	Tumor lesion	49	F	1	N	IA2
	p.L858R	1	Tumor lesion	50	F	1	N	IA2
	p.L858R	1	Tumor lesion	51	F	1	N	IB
	p.L858R	1	Tumor lesion 1		F	1	N	IA2
			Tumor lesion 2	13				
	p.L858R	1	Tumor lesion	52	F	1	N	IA2
	p.L858R	1	Tumor lesion	53	F	1	N	IA1
	p.L858R	1	Tumor lesion	54	M	1	Y	IA1
	p.L858R	1	Tumor lesion	55	F	1	N	IA1
	p.L858R	1	Tumor lesion	56	F	2	N	IB
	p.L858R	1	Tumor lesion	57	F	1	N	IA2
	p.L858R	1	Tumor lesion	58	M	2	N	IB
	p.L858R	1	Tumor lesion 1		F	1	N	IA2
			Tumor lesion 2	59				
	p.L858R	1	Tumor lesion	60	F	1	Y	IA2
	p.L858R	1	Tumor lesion	61	F	1	N	IB
	p.L858R	1	Tumor lesion	3	M	1	N	IB
	p.L858R	1	Tumor lesion	4	F	2	N	IA1
	p.L858R	1	Tumor lesion 1		M	3	Y	IIIB
			Tumor lesion 2	9				
	p.L858R	1	Tumor lesion	62	M	1	Y	IA3
	p.L858R	1	Tumor lesion	63	F	1	N	IA1
	p.L858R	1	Tumor lesion	24	F	2	N	IB
	p.L858R	1	Tumor lesion	7	M	2	N	IA1

**Bold:** patients (patient numbers are marked in bold) who have more than one mutation at the same time. NT, number of types; NFR, number of first-order relative lung cancer patients of MPLC patients.

**Table S4** Functional prediction of EGFR SNVs by PolyPhen-2 and PROVEAN

Variations	PolyPhen-2		PROVEAN <sup>a</sup>	
	Score	Prediction	Score	Prediction
p. G719A	1.000	Probably damaging	-5.056	Deleterious
p. G719S	1.000	Probably damaging	-5.056	Deleterious
p. G719C	1.000	Probably damaging	-7.584	Deleterious
p. S768I	1.000	Probably damaging	-4.836	Deleterious
p. S768N	1.000	Probably damaging	-2.000	Neutral
p. R776H	1.000	Probably damaging	-4.393	Deleterious
p. R776G	1.000	Probably damaging	-6.148	Deleterious
p. R776C	1.000	Probably damaging	-7.028	Deleterious
p. L788V	1.000	Probably damaging	-2.645	Deleterious
p. T790M	1.000	Probably damaging	-5.292	Deleterious
p. C797S	0.989	Probably damaging	-8.007	Deleterious
p. C781Y	1.000	Probably damaging	-10.360	Deleterious
p. V843I	0.999	Probably damaging	-0.887	Neutral
p. L858Q	1.000	Probably damaging	-5.145	Deleterious
p. L858R	1.000	Probably damaging	-5.202	Deleterious
p. L861Q	0.999	Probably damaging	-5.289	Deleterious
p. L861G	1.000	Probably damaging	-7.073	Deleterious

<sup>a</sup> PROVEAN prediction cutoff = -2.5.