Original Article

Mutational characteristics of Chinese Han and ethnic minorities in southwestern China: a propensity score matched analysis

Yongchun Zhou[1,2#], Xin Liu[1,2#], Yinjin Guo[1,2], Zhuoying Li[1,2], Shunhui Zhu[3], Quan Li[1,2], Xiaoxiong Wang[1,2^], Ruijiao Yang[1,2], Zhenghai Shen[1,2], Hongsheng Li[1,2], Yaxi Du[1,2], Luyao Ma[1,2], Shaocong Shen[1,2], Yunyi Lan[1,2^], Yunchao Huang[4]

[1]Molecular Diagnostic Center, the Third Affiliated Hospital of Kunming Medical University, Kunming, China; [2]International Joint Laboratory on High Altitude Regional Cancer, Kunming, China; [3]Tumor Molecular and Cell Biology Center, Honghe Prefecture Third People's Hospital, Gejiu, China; [4]Yunnan Key Laboratory of Lung Cancer Research, Kunming, China

Contributions: (I) Conception and design: Y Huang; (II) Administrative support: Y Zhou, X Liu, Y Guo; (III) Provision of study materials or patients: Z Li, S Zhu, Q Li, X Wang, R Yang, Z Shen; (IV) Collection and assembly of data: H Li, Y Du, L Ma, S Shen, Y Lan; (V) Data analysis and interpretation: Y Zhou, X Liu, Y Guo; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Dr. Yunchao Huang. Yunnan Key Laboratory of Lung Cancer Research, Kunming, China. Email: hycynszlyy1982@sina.com.

Background: In China, there has never been a comprehensive analysis of lung cancer-associated genetic mutations focused on ethnic minorities in the southwestern region. Our study aimed to provide valuable information on lung cancer-associated genetic alterations for cancer diagnosis and treatment, especially in ethnic minorities.

Methods: Retrospective data acquisition was conducted spanning 3 years (2016.01–2019.06) among all patients who were diagnosed with lung cancer at the Third Affiliated Hospital of Kunming Medical University Hospital. A total of 5,167 patients including 373 ethnic minorities were included in this study. Propensity score matching (PSM) was used to eliminate the bias between Han and other ethnic minorities, including gender, age, smoking history, metastasis status, clinical stage, histological type, sample type, region, and Xuanwei origin. All tests were two-tailed, and significance was defined as P less than 0.05.

Results: In terms of the prevalence of *EGFR*, *EGFR* L858R, *EGFR* T790M, *ROS1*, *RET*, *MET*, *BRAF*, and *ERBB2* mutations, there was no significant difference among ethnic groups in Yunnan Province (P>0.05). A higher proportion of *EGFR* 19 deletion was observed in Hui patients with lung cancer compared with patients of other ethnicities in Yunnan (P=0.048). The prevalence of *KRAS* mutations was higher in Hani (17.65%, 3/17) and Han patients (11.44%, 80/699) than that in other Yunnan ethnicities (6.04%, 9/149; P=0.07). In Hui patients, *ALK* fusion was correlated with a history of non-smoking and male gender. In Bai patients, *BRAF* mutation was also correlated with a history of non-smoking. In all ethnic groups, *EGFR* mutation was more frequent in women.

Conclusions: This study is the first in-depth large case-control study on genetic mutation profiles among multi-ethnic patients in southwestern China, especially focused on ethnic minorities in this area. Our study may facilitate the understanding of the etiology of this malignant disease and consequently help to reduce the incidence of lung cancer in Yunnan ethnic minority areas.

Keywords: Multi-ethnic patients; propensity score matching (PSM); Yunnan; lung cancer; *EGFR*

Submitted Jan 27, 2022. Accepted for publication Apr 08, 2022.
doi: 10.21037/atm-22-987
View this article at: https://dx.doi.org/10.21037/atm-22-987

^ ORCID: Xiaoxiong Wang, 0000-0001-7745-3110; Yunyi Lan, 0000-0002-6453-9270.

Introduction

Lung cancer is the most common cancer and a leading cause of cancer-related death worldwide (1). In China, it was estimated that the total number of newly diagnosed cases of lung cancer in 2015 was about 787,000, which is the equivalent of more than 2,100 cases diagnosed each day (2). High prevalence of major driver gene mutations and fusions in *EGFR*, *ALK*, *BRAF*, *ROS1*, and *KRAS* has been found in lung cancer patients, and certain mutations are associated with drug sensitivity or resistance, which have been previously observed in China (3,4). In particular, point mutations *EGFR* L858R and exon 19 deletion comprised nearly 90% of all *EGFR* mutations in non-small cell lung cancer (NSCLC) (5). The mutation profiles of NSCLC patients are vital to guide targeted therapy and monitor tumor recurrence, thereby improving the survival rate.

Apart from the known factors (gender, age, and smoking history, among others), racial and ethnic variations may also be important in lung cancer incidence and mortality (6,7). China is a multi-ethnic country with more than 1.3 billion people and 56 ethnic groups. Yunnan Province is located in the Yunnan-Guizhou Plateau in southwestern China. It is the largest habitation of ethnic minorities, including the Yi, Bai, Hani, Zhuang, Dai, and Hui, among others. Most of the ethnic groups have their own social settlements, unique customs, and lifestyles. Such a special population composition enables systematic research on lung cancer to be performed in Yunnan Province, taking into account clinical, lifestyle, and demographic (Xuanwei region) factors, which offers great practical significance to the effective prevention and early diagnosis of lung cancer for ethnic minorities.

The genetic profiles of lung cancer patients in ethnicities is still unclear, so we investigate the influence of ethnicities on the genetic profiles based on a cohort of 5,167 lung cancer patients from the Third Affiliated Hospital of Kunming Medical University Hospital in the period from 2016 to 2019, after performing propensity score matching (PSM). This is the first in-depth large case-control study on lung cancer genetic mutation profiles among multi-ethnic patients in southwestern China, especially focused on ethnic minorities in this area. We present the following article in accordance with the STROBE reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-22-987/rc).

Methods

Study population

Retrospective data acquisition was conducted spanning 3 years (2016.01–2019.06) among all patients who were diagnosed with lung cancer at the Third Affiliated Hospital of Kunming Medical University Hospital. The data were collected from hospital medical records which comprised clinical medical history, radiology reports, and pathology reports. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional ethics board of Third Affiliated Hospital of Kunming Medical University (No. KY202015) and informed consent was taken from all the patients.

In this study, 860 patients [376 tested by amplification refractory mutation system (ARMS), 484 tested by next-generation sequencing (NGS)] had 8 major gene alterations including *EGFR*, *ALK*, *KRAS*, *ROS1*, *RET*, *MET*, *BRAF*, and *ERBB2* alterations (Table S1). Of these, 618 patients had a single *EGFR* gene detection term. A total of 342 patients had a panel of *EGFR*, *ALK*, and *ROS1* detection terms, while 40 patients had a panel of *EGFR*, *ALK*, *ROS1*, and *MET* gene detection terms.

Statistical analysis

Patients were divided into groups based on the ethnicities they reported. As an observational study, significant bias might be introduced by the large Han population and other clinical factors. A 1:4 (other ethnic groups: Han) PSM method was performed with a 0.05 standard deviation caliper width on the R package "*MatchIt*" (8). Matching variables included gender, age, smoking history, metastasis status, clinical stage, histological type, sample type, region, and Xuanwei origin. The *EGFR* mutation spectrum in Xuanwei area is different from other area in Yunnan.

In unadjusted analyses, multivariate logistic regression was used to explore associations between clinical factors (smoking status and gender) and genetic alterations in each individual group. All analyses were conducted by Statistical Analysis Systems (SAS) for Windows version 9.4. All tests were two-tailed, and significance was defined as P less than 0.05.

Results

Clinicopathological characteristics of patients with lung cancer in southwestern China

During the study period, there were a total of 5167 lung cancer patients with clinical data that were recorded from January 2016 to June 2019 in the Third Affiliated Hospital of Kunming Medical University Hospital (Table S2). As the number of Han participants was significantly higher than that of ethnic minorities, PSM was used to eliminate the baseline variations and reduce the impact of patient number and other known factors on the results. There were four major ethnic groups (175 Yi, 100 Bai, 50 Hui, and 48 Hani) included in this study. Other ethnic groups were excluded as the number of patients was not sufficient to conduct an effective comparative analysis.

After a 1:4 (ethnic minorities: Han) PSM, the total analytic cohort included 1865 patients. Of these, 1492 were Han, 175 were Yi, 100 were Bai, 50 were Hui, and 48 were Hani (*Table 1*). Han patients were selected to match with other ethnic groups. The clinical, demographic, health, and lifestyle characteristics including gender, age, smoking history, metastasis status, clinical stage, histological type, sample type, and Xuanwei origin were well balanced among ethnic groups. In addition, a marked difference was observed in the regional distribution of lung cancer patients in varied ethnic groups (*Table 1*). The majority of the Yi patients were from central Yunnan and south Yunnan. The majority of the Bai patients were from northwest Yunnan. Most of the Hani patients were from south Yunnan. In addition, the majority of the Hui patients were from northeast Yunnan and central Yunnan. The NGS gene panel testing detected genetic alterations in 25.95% (484/1,865) of the patients (Table S1). The ARMS or super-ARMS detected genetic alterations in the rest of the patients.

Mutational status of driver genes in patients with lung cancer in southwestern China

In the PSM adjusted cohort, the *EGFR* mutation rate in Han patients was 44.5% (662/1,488; *Figure 1A*), with the major mutations being *EGFR* 19 deletion, *EGFR* L858R, and *EGFR* T790M. Han patients with lung cancer had a significantly lower prevalence of *EGFR* 19 deletion than other ethnic groups including the Yi, Bai, Hui, and Hani (P=0.074; *Figure 1B,1C*). We also found that Hui patients had a higher prevalence of *EGFR* 19 deletion than other ethnic groups (P=0.048; *Figure 1D*). In addition, the incidence of *EGFR* (P=0.569), *EGFR* L858R (P=0.612), and *EGFR* T790M (P=0.325) had no significant association among Yunnan major ethnic groups (*Figure 1A,1E,1F*).

The prevalence of *ALK* fusion was higher in Hui patients than that in non-Hui patients (*Figure 2A*). Of the Hui patients, 13% (5/38) were *ALK* positive, while 6.08% (7/115) of Yi patients, 7.58% (5/66) of Bai patients, 6.45% (2/31) of Hani patients, and 6.91% (69/999) of Han patients were *ALK* positive (*Figure 2A,2B*). In comparison with other Yunnan ethnicities, the incidence of *KRAS* mutation was higher in Hani and Han patients (*Figure 2C*). Of the Hani patients, 17.65% (3/17) were *KRAS* positive, while 11.44% (80/699) of Han patients and 6.04% (9/149) of other ethnic groups were *KRAS* positive (*Figure 2C,2D*). The incidence of *ROS1*, *RET*, *MET*, *BRAF*, and *ERBB2* alterations were similar among Yunnan ethnic groups (*Figure 3*).

The correlation between clinical factors (smoking status and gender) and genetic alterations within each ethnic group

Figure 3 illustrates the correlation between clinical factors (current smoking status and gender) and genetic alterations within each ethnic group in this study. Of all the patients included in this study, 3,664 (88.48%) of the patients were Han and 1,503 (11.52%) were from ethnic minorities, including 175 Yi, 100 Bai, 50 Hui, 48 Hani, 32 Zhuang, 21 Naxi, 72 patients from other ethnic minorities, and 21 patients from unknown ethnic groups (Table S2). In Hui patients, non-smoking and male were correlated with *ALK* fusion (*Figure 4A*). In this cohort, non-smoking was correlated with *BRAF* mutation (*Figure 4A*). In Bai patients, non-smoking was prone to *BRAF* mutation (*Figure 4B*). In Han and non-Han groups, *EGFR* and *EGFR* L858R mutations were correlated with female patients.

Relationship between mediastinal sites and targeted mutation statuses in patients with lung cancer in southwestern China

The associations between mediastinal sites and major driver mutations were analyzed in our cohort (study cohort before PSM, Figures S1-S3). The incidence of *EGFR* mutation was higher in patients with bone metastasis than that in patients without metastasis (49.2% *vs.* 43.94%; P= 0.072; Figure S1A). *BRAF* mutation was significantly associated with bone metastasis (P=0.01; Figure S1F). Although not statistically significant, *BRAF* mutation was also higher in

Table 1 Characteristics of lung cancer patients from ethnic groups in southwestern China after PSM

Characteristics	Han (n=1,492)	Yi (n=175)	Bai (n=100)	Hui (n=50)	Hani (n=48)	P value
Gender, n [%]						0.898
Female	628 [42]	73 [42]	45 [45]	23 [46]	18 [38]	
Male	864 [58]	102 [58]	55 [55]	27 [54]	30 [62]	
Age (years)						0.7
Mean ± SD	57.2±10.4	56.7±10.3	57.9±11.0	55.4±9.75	57.0±11.6	
Median (min, max)	57.0 (17.0, 92.0)	56.0 (26.0, 81.0)	59.5 (28.0, 80.0)	54.0 (28.0, 82.0)	57.5 (31.0, 82.0)	
Smoking, n [%]						0.965
Non-smoker	946 [63]	112 [64]	62 [62]	34 [68]	31 [65]	
Smoker	546 [37]	63 [36]	38 [38]	16 [32]	17 [35]	
Brain metastasis, n [%]						0.617
No	1,384 [93]	159 [91]	90 [90]	48 [96]	44 [92]	
Yes	108 [7]	16 [9]	10 [10]	2 [4]	4 [8]	
Bone metastasis, n [%]						0.902
No	1,191 [80]	137 [78]	82 [82]	38 [76]	39 [81]	
Yes	301 [20]	38 [22]	18 [18]	12 [24]	9 [19]	
Visceral metastasis, n [%]						0.393
No	1,386 [93]	166 [95]	90 [90]	44 [88]	44 [92]	
Yes	106 [7]	9 [5]	10 [10]	6 [12]	4 [8]	
Clinical stage, n [%]						0.479
I	306 [21]	37 [21]	17 [17]	14 [28]	8 [17]	
II	47 [3]	8 [5]	6 [6]	0 [0]	1 [2]	
III	225 [15]	29 [17]	16 [16]	3 [6]	6 [12]	
IV	914 [61]	101 [58]	61 [61]	33 [66]	33 [69]	
Histology, n [%]						0.917
LUAD	1,328 [89]	153 [87]	86 [86]	45 [90]	46 [96]	
LUSC	127 [9]	15 [9]	11 [11]	5 [10]	1 [2]	
NSCLC nos.	17 [1]	3 [2]	1 [1]	0 [0]	0 [0]	
ADSC	9 [1]	1 [1]	1 [1]	0 [0]	1 [2]	
LCLC	6 [0]	2 [1]	0 [0]	0 [0]	0 [0]	
SCLC	5 [0]	1 [1]	1 [1]	0 [0]	0 [0]	
Sample type, n [%]						0.788
Plasma	511 [34]	57 [33]	37 [37]	16 [32]	13 [27]	
Tissue	981 [66]	118 [67]	63 [63]	34 [68]	35 [73]	

Table 1 (*continued*)

Table 1 (*continued*)

Characteristics	Han (n=1,492)	Yi (n=175)	Bai (n=100)	Hui (n=50)	Hani (n=48)	P value
Xuanwei origin, n [%]						0.437
No	1,450 [97]	168 [96]	99 [99]	48 [96]	48 [100]	
Yes	42 [3]	7 [4]	1 [1]	2 [4]	0 [0]	
Region distribution, n [%]						<0.001
Northeast Yunnan	465 [31]	26 [15]	3 [3]	19 [38]	0 [0]	
South Yunnan	224 [15]	57 [33]	4 [4]	6 [12]	41 [85]	
Northwest Yunnan	150 [10]	16 [9]	75 [75]	5 [10]	0 [0]	
Central Yunnan	492 [33]	59 [34]	17 [17]	14 [28]	7 [15]	
Yunnan unknown	5 [0]	0 [0]	0 [0]	0 [0]	0 [0]	
Non-Yunnan	156 [10]	17 [10]	1 [1]	6 [12]	0 [0]	

PSM, propensity score matching; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; NSCLC nos., non-small cell lung cancer not otherwise specified; ADSC, adipose-derived stem cell; LCLC, large-cell lung cancer; SCLC, small-cell lung cancer.

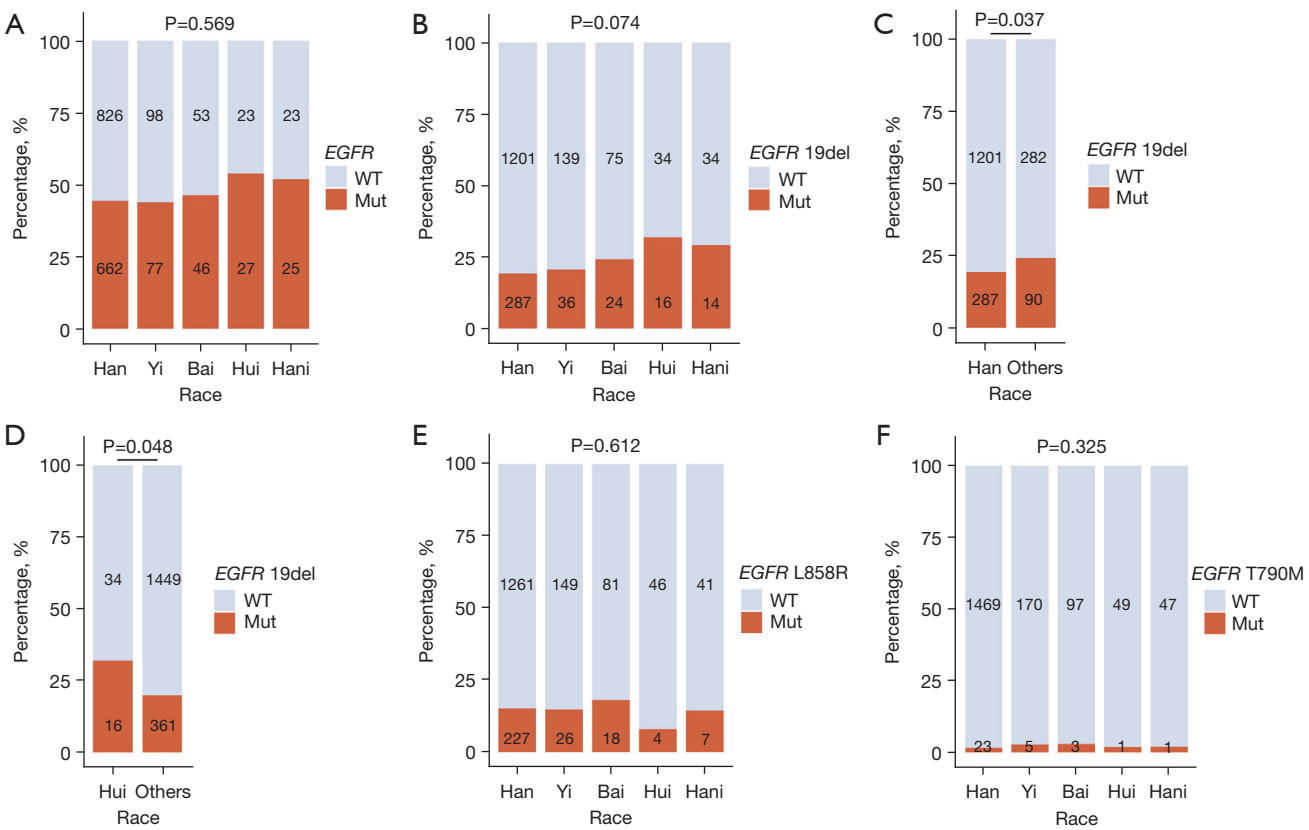

Figure 1 Comparison of the prevalence of *EGFR* mutation among major ethnic groups. (A) The incidence of *EGFR* mutation in Han, Yi, Bai, Hui, and Hani groups; (B) the incidence of *EGFR* 19 deletion in Han, Yi, Bai, Hui, and Hani groups; (C) the incidence of *EGFR* 19 deletion in Han and non-Han groups; (D) the incidence of *EGFR* 19 deletion in Hui and others groups; (E) the point mutation frequency of *EGFR* L858R in Han, Yi, Bai, Hui, and Hani groups; (F) the point mutation frequency of *EGFR* T790M in Han, Yi, Bai, Hui, and Hani groups. WT, wild type; Mut, mutation.

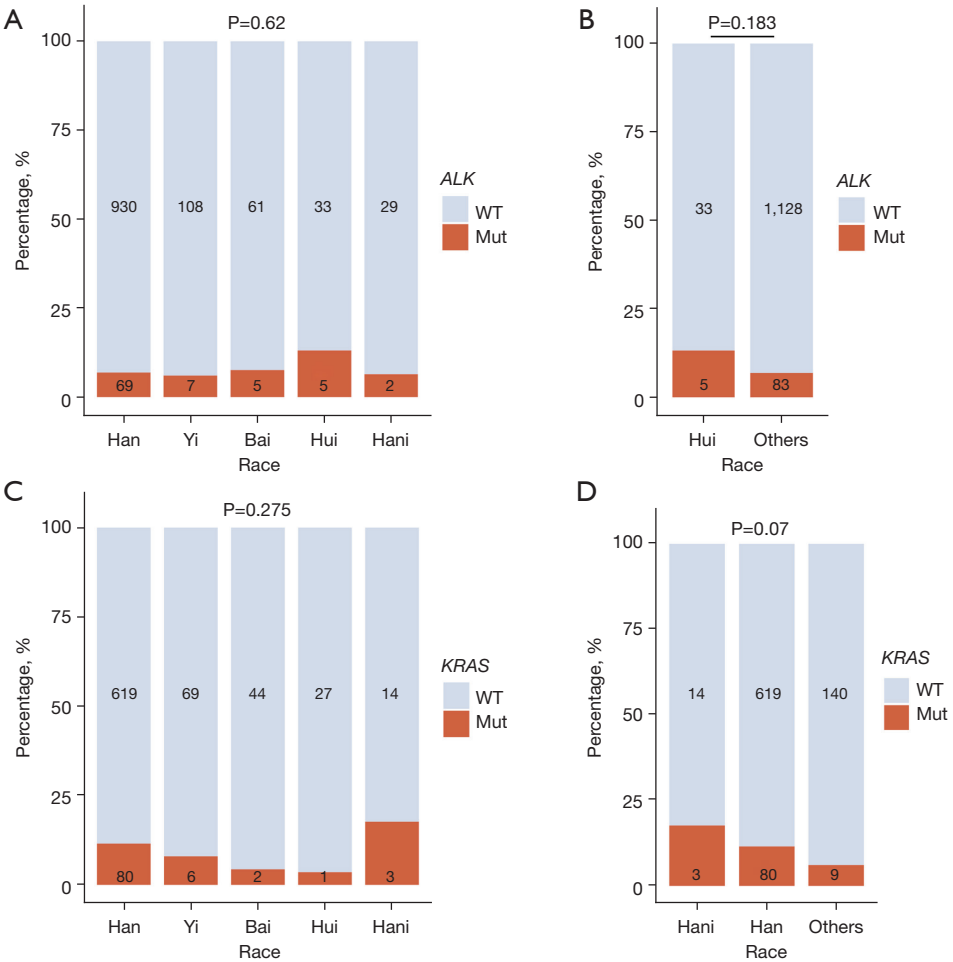

Figure 2 Comparison of the prevalence of *ALK* fusion and *KRAS* mutation among major ethnic groups. (A) The incidence of *ALK* fusion in Han, Yi, Bai, Hui, and Hani groups; (B) the prevalence of *ALK* fusion in Hui and others groups; (C) the incidence of *KRAS* mutation in Han, Yi, Bai, Hui, and Hani groups; (D) the incidence of *KRAS* mutation in Hani, Han, and other ethnic groups. WT, wild type; Mut, mutation.

patients with brain metastasis (P=0.082; Figure S2F) and visceral metastasis (P=0.07; Figure S3F) than that in patients without metastasis. In terms of the incidence of *ROS1*, *RET*, *MET*, and *ERBB2* mutations, there was no significant difference between patients with metastasis and patients without metastasis (Figures S1-S3).

Discussion

Lung cancer-associated genetic mutations are widespread in China. In this study, we presented an in-depth large case-control study on the lung cancer mutation profiles of major drug-targetable genes among multi-ethnic patients in southwestern China, especially focused on ethnic minority patients in this area. It has long been recognized that different ethnicities exhibit distinct patterns of ontogenetic profiles. In our study, there were differences in genetics, race, culture, diet, living habits, residence, and other aspects among different ethnic groups. We also observed that most of the ethnic groups had their unique social settlements. Hani patients were mainly settled in south Yunnan. The majority of Hui patients were settled in northeast and central Yunnan, including Qujing City. Qujing, located in southwest Yunnan, is an area with an extremely high incidence of lung cancer, especially in Xuanwei County (9).

EGFR is one of the most prevalent genetic alterations among lung cancer patients, even though other genetic aberrations also exist. In our previous study, we found that the *EGFR* mutation rates were 34.9% and 42.3% among patients with NSCLC and adenocarcinoma,

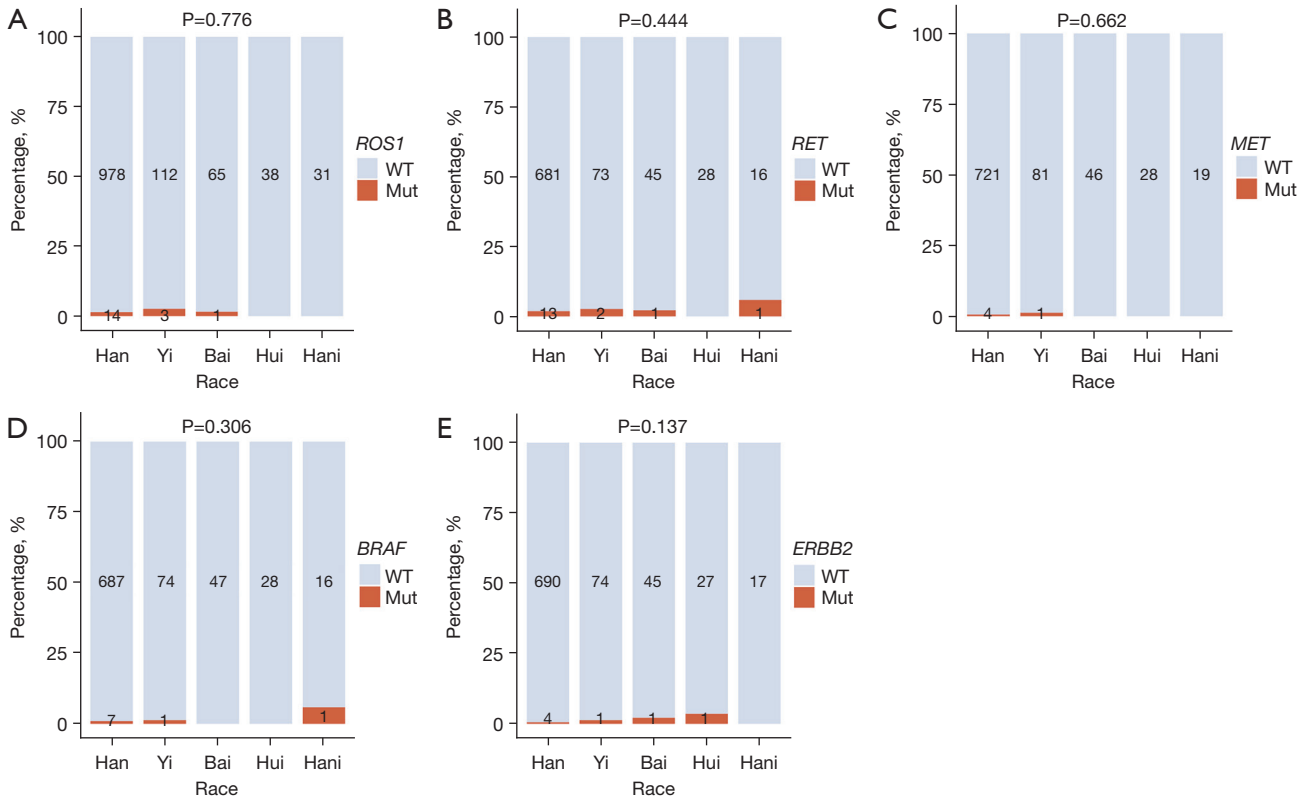

Figure 3 Comparison of the prevalence of *ROS1*, *RET*, *MET*, *BRAF*, and *ERBB2* alterations among major ethnic groups. (A-E) The prevalence of *ROS1*, *RET*, *MET*, *BRAF*, and *ERBB2* alterations in the Han, Yi, Bai, Hui, and Hani groups. WT, wild type; Mut, mutation.

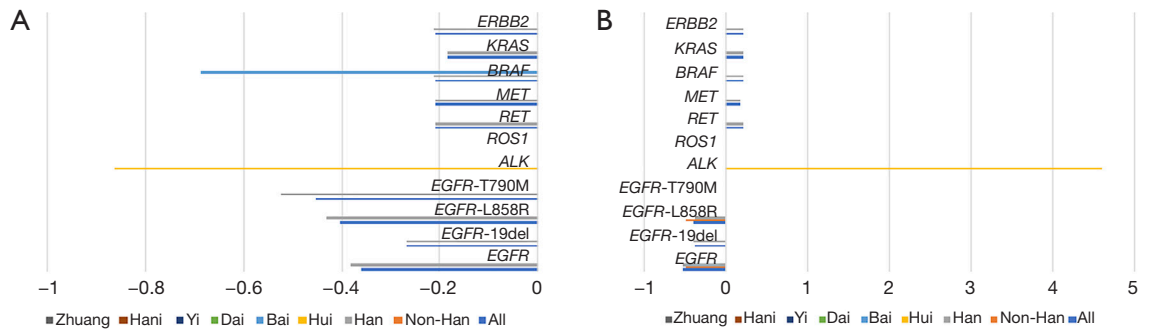

Figure 4 Chart depicting the interrelationships between genetic mutations, smoking status (A), and gender (B) in each study ethnic group.

respectively (10). Patients with NSCLC in Xuanwei (one of the highest mortality and incidence areas in China) displayed a unique profile of driver gene mutations, especially a higher prevalence of *EGFR* compound mutations and dominant *KRAS* G12C subtype compared to the general population in Yunnan (11). Female, never-smokers, and adenocarcinoma were correlated with a higher rate of NSCLC patients (12). These results were consistent with our present study and previous studies in East Asian populations (13-15). However, the prevalence of *EGFR* mutation was much higher than that in Caucasian lung adenocarcinoma patients (16,17). Like people from East Asia, the majority of the ethnic groups (Han, Yi, Bai, Hui, and Hani) from southwestern China are of Mongoloid race, which is one possible explanation.

In this study, we used PSM to balance the baseline

variations to further interpret the correlation between major drug-targetable genes and multiple ethnic groups. There were 8 major oncogenic genes addressed in this study including *EGFR*, *ALK*, *KRAS*, *ROS1*, *RET*, *MET*, *BRAF*, and *ERBB2* mutations. We found that the frequencies of major gene mutations were quite similar among multiple ethnic groups, except the Hui group. The prevalence of *EGFR* 19 deletion was relatively high in Hui patients and relatively low in Han patients. The mutation rate of *EGFR* was similar among ethnic groups in our cohort. However, entirely opposite results were found in Ningxia Hui Autonomous region. Yuan *et al.* indicated that, in contrast to patients from Yunnan areas, the mutation rate of *EGFR* in Han patients was lower than that in Hui patients from Ningxia ($P<0.05$) (18). A wide variety of known risk factors correlated with genetic mutations were included in the Ningxia study. Failure to exclude key risk factors may result in misspecified models that incorrectly attribute risk to members of the ethnic groups. Our study comprehensively examined differences in gene profiles by ethnicity, and also excluded a wide variety of known risk factors. Surprisingly, for the first time, we also found that Hui ethnicity may be associated with an increased likelihood of harboring rare oncogenic mutations in NSCLC. Hui patients from Yunnan had a higher incidence of *ALK* fusion and *KRAS* mutations than other ethnic groups.

In addition, many epidemiological investigations have shown that races might be associated with the frequency of metastasis in lung cancer (19). It was reported that blacks had a significantly increased frequency of mediastinal lymph node involvement and metastasis than whites (20). Another study reported that Asian and Pacific Islander patients presented the highest prevalence of bone metastasis (24.6%), followed by white (20.7%) and black patients (19.9%) (21). However, our study showed that there is not necessarily a link between metastasis and major ethnic groups in southwestern China, although the small sample size of ethnic groups might limit this interpretation.

Despite the significance of our study, there were several limitations that might impede the interpretation of our results. Firstly, our data might not represent the true prevalence of lung cancer mutations in Yunnan Province. This was a retrospective analysis and only included a single center, which might lead to selection bias. Although Yunnan is a multi-ethnic region, the Han accounted for 90.7% of our included patients (before PSM). The samples from other ethnic patients were lacking, which may not reflect the actual genetic background of Yunnan residents. A multicenter or nationwide well designed prospective study would be helpful to further confirm our results. Secondly, the detection of lung cancer-associated genetic mutations was performed by different methods (ARMS and NGS) and in various screening centers. The reliability and consistence of the analysis results might also be impacted by these factors.

In conclusion, we provide results from a large retrospective study on mutation profiles in major ethnic groups in southwestern China. We believe this work represents the largest and broadest study of genetic mutation profiles in the southwestern region, and can serve as a reference for future research. With the increasing incidence of lung cancer and the development of targeted kinase inhibitors, oncogenic genetic profile investigations of lung cancer in Yunnan, a region with a high concentration of ethnic minorities, may facilitate our understanding of the etiology of this malignant disease and consequently help to reduce the incidence of lung cancer in Yunnan ethnic minority areas.

Acknowledgments

The authors thank Dr. Xi Li (Burning Rock Biotech, Guangzhou, China), Dr. Ting Hou (Burning Rock Biotech, Guangzhou, China), and Dr. Wenbo Dong (Burning Rock Biotech, Guangzhou, China) for discussing the results.
Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 81860513).

Footnote

Reporting Checklist: The authors have completed the STROBE reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-22-987/rc

Data Sharing Statement: Available at https://atm.amegroups.com/article/view/10.21037/atm-22-987/dss

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-22-987/coif). 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. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional ethics board of Third Affiliated Hospital of Kunming

Medical University (No. KY202015) and informed consent was taken from all the patients.

References

1. Siegel R, Ma J, Zou Z, et al. Cancer statistics, 2014. CA Cancer J Clin 2014;64:9-29. Erratum in: CA Cancer J Clin 2014;64:364.
2. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115-32.
3. Feng H, Wang X, Zhang Z, et al. Identification of Genetic Mutations in Human Lung Cancer by Targeted Sequencing. Cancer Inform 2015;14:83-93.
4. El-Deiry WS, Goldberg RM, Lenz HJ, et al. The current state of molecular testing in the treatment of patients with solid tumors, 2019. CA Cancer J Clin 2019;69:305-43.
5. da Cunha Santos G, Shepherd FA, Tsao MS. EGFR mutations and lung cancer. Annu Rev Pathol 2011;6:49-69.
6. Mitsudomi T, Yatabe Y. Mutations of the epidermal growth factor receptor gene and related genes as determinants of epidermal growth factor receptor tyrosine kinase inhibitors sensitivity in lung cancer. Cancer Sci 2007;98:1817-24.
7. Zhang Y, Sun Y, Pan Y, et al. Frequency of driver mutations in lung adenocarcinoma from female never-smokers varies with histologic subtypes and age at diagnosis. Clin Cancer Res 2012;18:1947-53.
8. Ho D, Imai K, King G, et al. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. J Stat Softw 2011;42:1-28.
9. Chen G, Sun X, Ren H, et al. The mortality patterns of lung cancer between 1990 and 2013 in Xuanwei, China. Lung Cancer 2015;90:155-60.
10. Ma Y, Li Q, Du Y, et al. Oncogenic Genetic Alterations in Non-Small-Cell Lung Cancer (NSCLC) in Southwestern China. Cancer Manag Res 2020;12:10861-74.
11. Zhou Y, Ge F, Du Y, et al. Unique Profile of Driver Gene Mutations in Patients With Non-Small-Cell Lung Cancer in Qujing City, Yunnan Province, Southwest China. Front Oncol 2021;11:644895.
12. Powell HA, Iyen-Omofoman B, Hubbard RB, et al. The association between smoking quantity and lung cancer in men and women. Chest 2013;143:123-9.
13. Liam CK, Wahid MI, Rajadurai P, et al. Epidermal growth factor receptor mutations in lung adenocarcinoma in Malaysian patients. J Thorac Oncol 2013;8:766-72.
14. Shi Y, Au JS, Thongprasert S, et al. A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). J Thorac Oncol 2014;9:154-62.
15. Shi Y, Li J, Zhang S, et al. Molecular Epidemiology of EGFR Mutations in Asian Patients with Advanced Non-Small-Cell Lung Cancer of Adenocarcinoma Histology - Mainland China Subset Analysis of the PIONEER study. PLoS One 2015;10:e0143515.
16. Rosell R, Moran T, Queralt C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med 2009;361:958-67.
17. Marchetti A, Martella C, Felicioni L, et al. EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. J Clin Oncol 2005;23:857-65.
18. Yuan LZ, Zhou W, Meng FZ. An Analysis of EGFR Mutation and Clinicopathologic Characteristics in Hui and Han Patients with Non-small-cell Lung Cancer. Ningxia Medical Journal 2017;39:484-7.
19. Gadgeel SM, Kalemkerian GP. Racial differences in lung cancer. Cancer Metastasis Rev 2003;22:39-46.
20. Akerley WL 3rd, Moritz TE, Ryan LS, et al. Racial comparison of outcomes of male Department of Veterans Affairs patients with lung and colon cancer. Arch Intern Med 1993;153:1681-8.
21. Xu G, Cui P, Zhang C, et al. Racial disparities in bone metastasis patterns and targeted screening and treatment strategies in newly diagnosed lung cancer patients. Ethn Health 2022;27:329-42.

(English Language Editor: C. Betlazar-Maseh)

Cite this article as: Zhou Y, Liu X, Guo Y, Li Z, Zhu S, Li Q, Wang X, Yang R, Shen Z, Li H, Du Y, Ma L, Shen S, Lan Y, Huang Y. Mutational characteristics of Chinese Han and ethnic minorities in southwestern China: a propensity score matched analysis. Ann Transl Med 2022;10(14):782. doi: 10.21037/atm-22-987

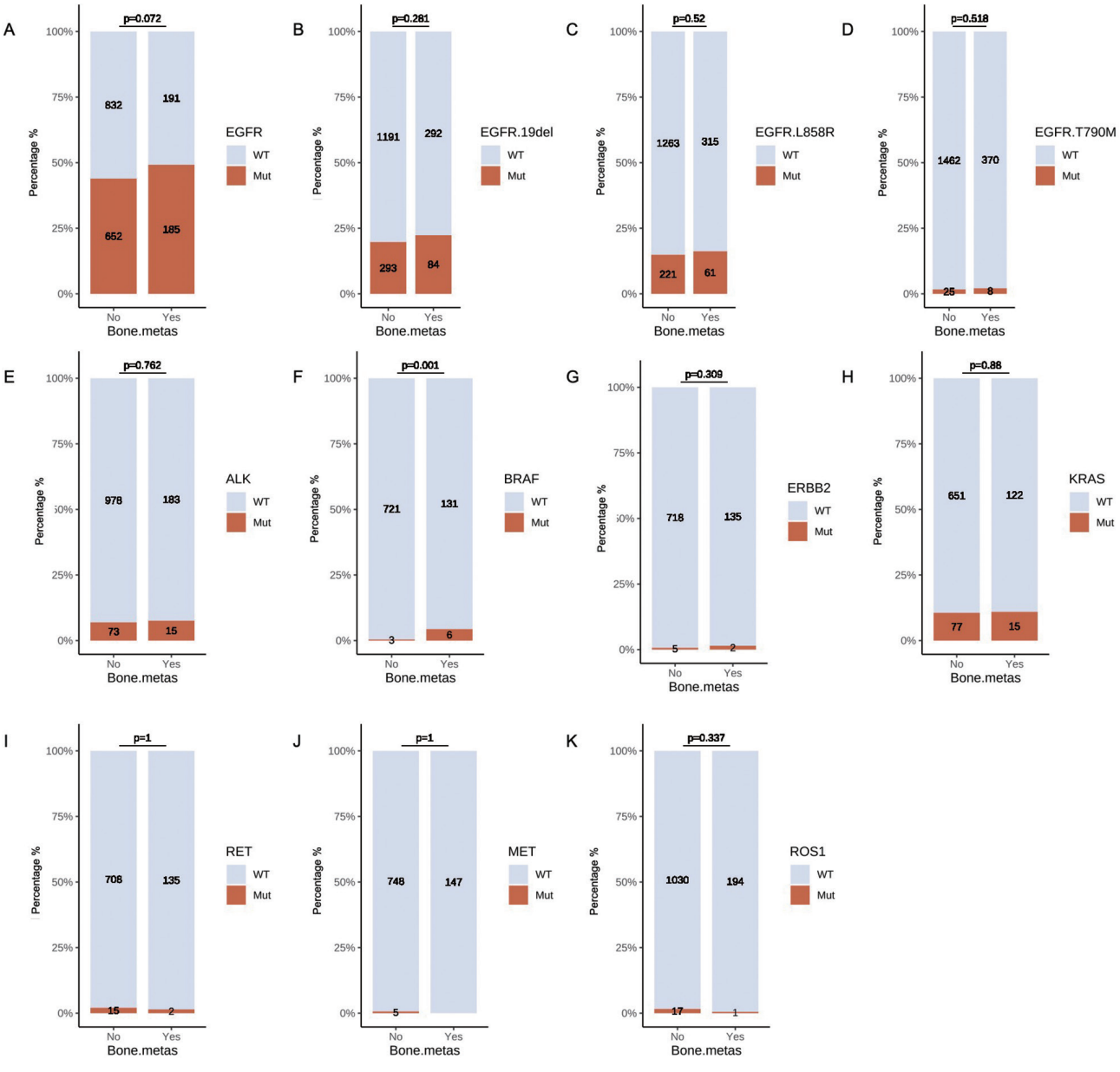

Figure S1 Comparison of the prevalence of major driver mutations between patients with bone metastasis and patients without bone metastasis. (A-K) *EGFR*, *EGFR* 19 deletion, *EGFR* L858R, *EGFR* T790M, *ALK*, *BRAF*, *ERBB2*, *KRAS*, *RET*, *MET*, and *ROS1* mutation. WT, wild type; Mut, mutation.

https://dx.doi.org/10.21037/atm-22-987

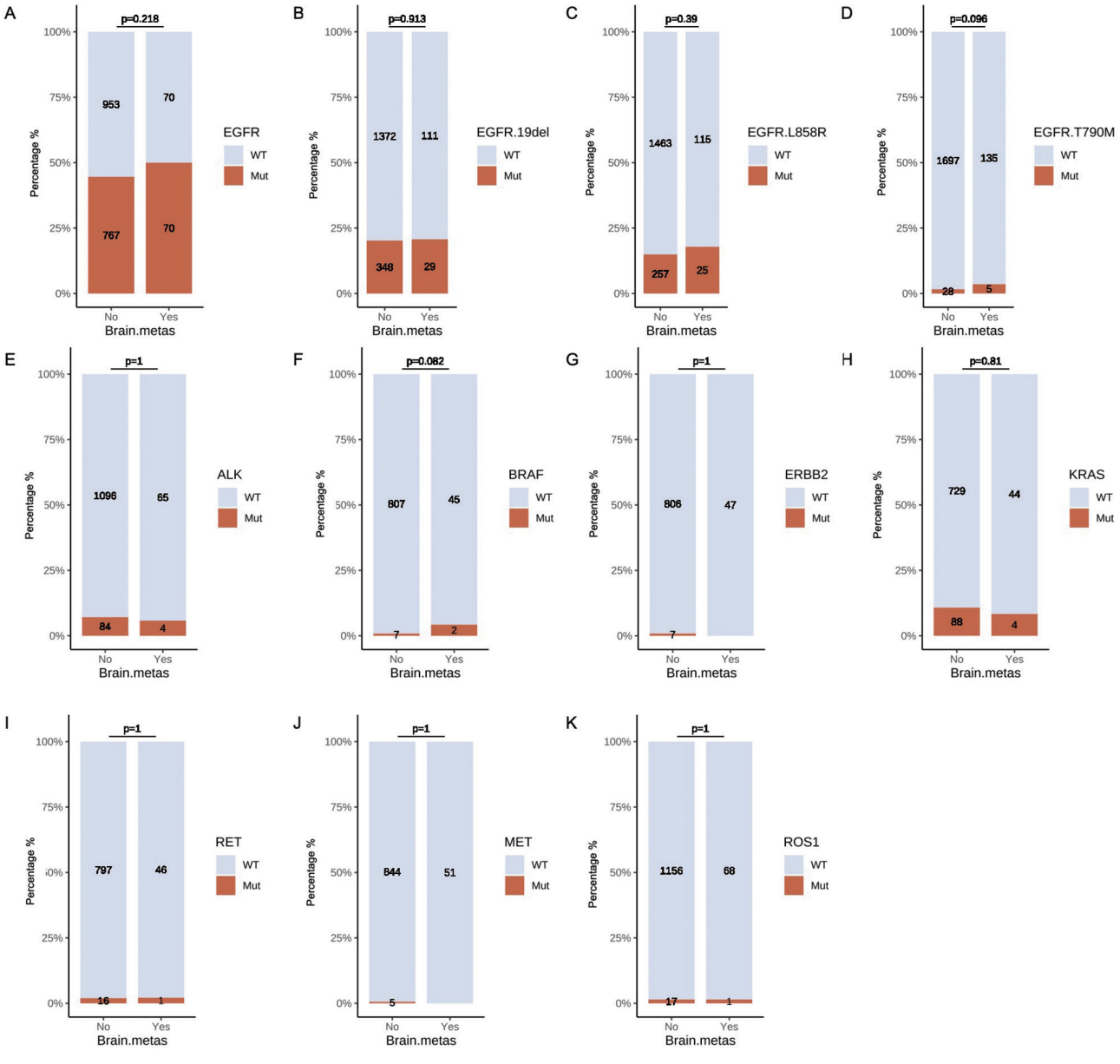

Figure S2 Comparison of the prevalence of major driver mutations between patients with brain metastasis and patients without brain metastasis. (A-K) *EGFR*, *EGFR* 19 deletion, *EGFR* L858R, *EGFR* T790M, *ALK*, *BRAF*, *ERBB2*, *KRAS*, *RET*, *MET*, and *ROS1* mutation. WT, wild type; Mut, mutation.

https://dx.doi.org/10.21037/atm-22-987

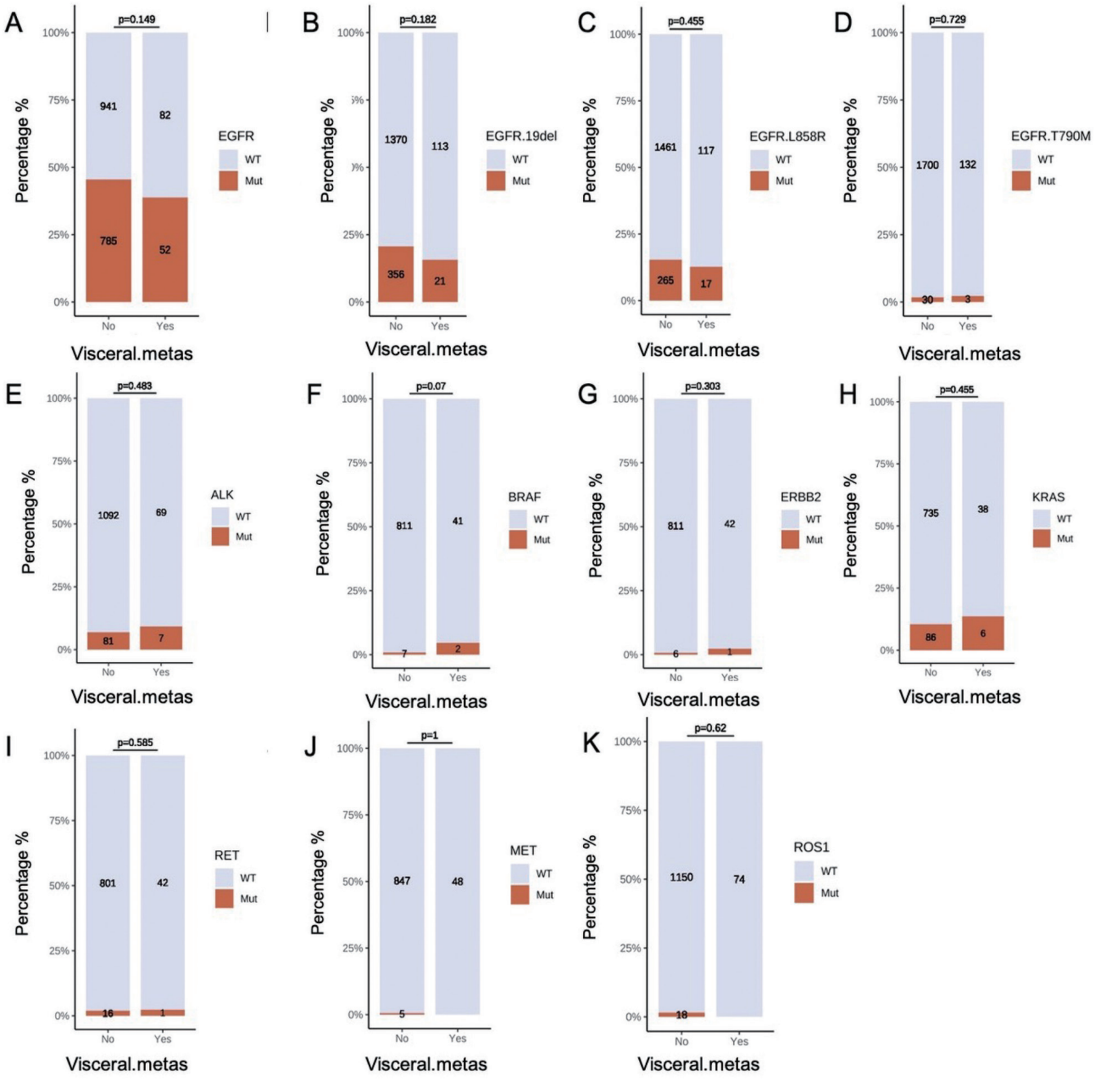

Figure S3 Comparison of the prevalence of major driver mutations between patients with visceral metastasis and patients without visceral metastasis. (A-K) *EGFR*, *EGFR* 19 deletion, *EGFR* L858R, *EGFR* T790M, *ALK*, *BRAF*, *ERBB2*, *KRAS*, *RET*, *MET*, and *ROS1* mutation. WT, wild type; Mut, mutation.

Table S1 The detection terms of the study population after PSM

Detection terms	ARMS	NGS
8 drivers	376	484
EGFR	618	NA
EGFR, *ALK*, *ROS1*	342	NA
EGFR, *ALK*, *ROS1*, *MET*	40	NA

PSM, propensity score matching; ARMS, amplification refractory mutation system; NGS, next-generation sequencing.

https://dx.doi.org/10.21037/atm-22-987

Table S2 Baseline characteristics of lung cancer patients from ethnic groups in Yunnan

Characteristics	Han (n=3,664)	Yi (n=175)	Bai (n=100)	Hui (n=50)	Hani (n=48)	P value
Gender, n [%]						0.353
Female	1,745 [48]	73 [42]	45 [45]	23 [46]	18 [38]	
Male	1,919 [52]	102 [58]	55 [55]	27 [54]	30 [62]	
Age (years)						0.547
Mean ± SD	57.5±10.4	56.7±10.3	57.9±11.0	55.4±9.75	57.0±11.6	
Median (min, max)	57.0 (17.0, 92.0)	56.0 (26.0, 81.0)	59.5 (28.0, 80.0)	54.0 (28.0, 82.0)	57.5 (31.0, 82.0)	
Smoking, n [%]						0.948
No	2,389 [65]	112 [64]	62 [62]	34 [68]	31 [65]	
Yes	1,275 [35]	63 [36]	38 [38]	16 [32]	17 [35]	
Brain metastasis, n [%]						0.645
No	3,288 [90]	159 [91]	90 [90]	48 [96]	44 [92]	
Yes	376 [10]	16 [9]	10 [10]	2 [4]	4 [8]	
Bone metastasis, n [%]						0.326
No	3,048 [83]	137 [78]	82 [82]	38 [76]	39 [81]	
Yes	616 [17]	38 [22]	18 [18]	12 [24]	9 [19]	
Visceral metastasis, n [%]						0.343
No	3,413 [93]	166 [95]	90 [90]	44 [88]	44 [92]	
Yes	251 [7]	9 [5]	10 [10]	6 [12]	4 [8]	
Stage, n [%]						0.262
I	824 [22]	37 [21]	17 [17]	14 [28]	8 [17]	
II	201 [5]	8 [5]	6 [6]	0 [0]	1 [2]	
III	629 [17]	29 [17]	16 [16]	3 [6]	6 [12]	
IV	2,010 [55]	101 [58]	61 [61]	33 [66]	33 [69]	
Histology, n [%]						0.817
LUAD	3,245 [89]	153 [87]	86 [86]	45 [90]	46 [96]	
LUSC	310 [8]	15 [9]	11 [11]	5 [10]	1 [2]	
NSCLC nos.	22 [1]	3 [2]	1 [1]	0 [0]	0 [0]	
ADSC	30 [1]	1 [1]	1 [1]	0 [0]	1 [2]	
LCLC	8 [0]	2 [1]	0 [0]	0 [0]	0 [0]	
Others	5 [0]	0 [0]	0 [0]	0 [0]	0 [0]	
SCLC	44 [1]	1 [1]	1 [1]	0 [0]	0 [0]	
Sample type, n [%]						0.637
PLA	1,307 [36]	57 [33]	37 [37]	16 [32]	13 [27]	
TIS	2,357 [64]	118 [67]	63 [63]	34 [68]	35 [73]	
Xuanwei, n [%]						<0.001
No	2,998 [82]	168 [96]	99 [99]	48 [96]	48 [100]	
Yes	666 [18]	7 [4]	1 [1]	2 [4]	0 [0]	

PSM, propensity score matching; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; NSCLC nos., non-small cell lung cancer not otherwise specified; ADSC, adipose-derived stem cell; LCLC, large-cell lung cancer; SCLC, small-cell lung cancer; PLA, plasma; TIS, tissue.

https://dx.doi.org/10.21037/atm-22-987