Clinicopathologic and molecular characteristics of Chinese lung adenocarcinoma patients with EGFR exon 20 insertion mutations

Lifeng Wang¹#, Zichen Liu²#, Zhiyi Wan³, Dong Jiang³, Min Zhang³, Shuku Liu⁴, Nanying Che²

¹Department of Pathology, Chongqing University Cancer Hospital and Chongqing Cancer Institute and Chongqing Cancer Hospital, Chongqing, China; ²Department of Pathology, Beijing Chest Hospital, Capital Medical University, Beijing Tuberculosis and Thoracic Tumor Research Institute, Beijing, China; ³Beijing Novogene Bioinformatics Technology Co., Ltd., Beijing, China; ⁴Thoracic Minimally Invasive Treatment Center, Beijing Chest Hospital, Capital Medical University, Beijing Tuberculosis and Thoracic Tumor Research Institute, Beijing, China

Contributions: (I) Conception and design: S Liu, N Che; (II) Administrative support: S Liu, N Che; (III) Provision of study materials or patients: L Wang, Z Liu, S Liu, N Che; (IV) Collection and assembly of data: L Wang, Z Liu; (V) Data collection and analysis: L Wang, Z Liu, Z Wan, D Jiang, M Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

These authors contributed equally to this work.

Background: Epidermal growth factor receptor exon 20 insertions (EGFR ex20ins) occur in about 4–14% of lung adenocarcinoma (LUAD) patients with EGFR mutations. Recently some targeted drugs have been approved for the treatment of LUAD patients with EGFR ex20ins. However, the heterogeneity of EGFR ex20ins mutations and resultant challenges in identifying them have led to the underestimation of their frequency.

Methods: We investigated the molecular and clinicopathologic features of EGFR ex20ins in 3,892 Chinese LUAD patients using next-generation sequencing (NGS). The frequency and distribution of EGFR ex20ins mutations between Chinese and Western LUAD patients were also compared by integrating the data of this study and the data of previous studies.

Results: A total of 23 unique EGFR ex20ins were identified in 77 LUAD patients, accounting for 1.98% of all LUAD patients and 3.49% of EGFR mutant LUAD patients. The 2 most common EGFR ex20ins subtypes were S768_D770dup and A767_V769dup, which together accounted for 55.84% of the EGFR ex20ins cases. About 61% (14/23) of the EGFR ex20ins subtypes occurred only once. Additionally, 8 of the EGFR ex20ins subtypes were not recorded in the COSMIC database. These results showed that the EGFR ex20ins mutations were highly heterogeneous. There was no significant difference in the frequency and distribution of EGFR ex20ins mutations between Chinese and Western LUAD patients, but the frequency of EGFR ex20ins mutations was significantly lower in EGFR-mutant Chinese LUAD patients than Western LUAD patients. The co-mutation analysis showed that EGFR ex20ins occurred significantly and exclusively with certain driver genes in LUAD, including ALK fusion ($\chi^2=7.133$, P=0.008), KRAS ($\chi^2=8.468$, P=0.004), and PIK3CA ($\chi^2=5.792$, P=0.016). No gene was observed to be significantly co-mutated with EGFR ex20ins. In general, patients with EGFR ex20ins shared a similar age and gender to patients with other EGFR mutations or without EGFR ex20ins.

Conclusions: Overall, our results revealed the molecular and clinicopathologic features of EGFR ex20ins in Chinese LUAD patients, which will be helpful for drug development and in clinical trials targeting EGFR ex20ins.

Keywords: EGFR exon 20 insertion; lung adenocarcinoma (LUAD); molecular characteristics; next-generation sequencing (NGS)

Submitted Dec 22, 2021. Accepted for publication Feb 21, 2022.
doi: 10.21037/atm-22-383

View this article at: https://dx.doi.org/10.21037/atm-22-383
Introduction

Epidermal growth factor receptor (EGFR) mutations are the most common mutations in lung adenocarcinoma (LUAD), and occur in approximately 40–63% of LUAD patients in China (1-3). The most frequent EGFR mutations are the in-frame exon 19 deletion (19del) and the exon 21 L858R point mutation, which together account for about 85–90% of EGFR mutations (4,5). These mutations represent canonical EGFR mutations, and respond to tyrosine kinase inhibitors (TKIs), such as erlotinib, gefitinib, and osimertinib (6). EGFR exon 20 insertion (EGFR ex20ins) mutations are the third most common mutations in EGFR, represent 4–14% of EGFR mutations (7-11), and are generally resistant to the EGFR TKIs, including osimertinib (7,9,11-13). Only a few EGFR ex20ins, such as D770delinsGY and A763_Y764insFQEA, are sensitive to EGFR TKIs (12,14). Nowadays, platinum-based chemotherapy is usually a standard of care for EGFR ex20ins LUAD. However, platinum-based chemotherapy leads to an average 7 months of progression-free survival for these patients, which is similar to that expected for canonical EGFR mutant LUAD (15).

Fortunately, the development of effective therapies for these patients has made rapid progress in recent years. In 2021, the United States (US) Food and Drug Administration (FDA) approved Exkivity (mobocertinib) and Rybrevant (amivantamab-vnjw) for the treatment of non-small-cell lung cancer (NSCLC) patients with EGFR ex20ins mutations. Poziotinib, a new-generation EGFR TKI, has also been reported to show encouraging antitumor activity for EGFR ex20ins mutations in recent clinical trials (16,17). However, different insertion mutations display divergent responsiveness to EGFR TKIs, including poziotinib (18). Thus, the identification of full insertion sequences of EGFR ex20ins is important.

EGFR ex20ins generally refer to in-frame insertions of 3–21 base pairs at different locations of EGFR exon 20. It contains 2 major regions, the C-helix of the tyrosine kinase domain (AA 762–766) and the loop following the C-helix (AA 767–774) (7,9,10,14). EGFR ex20ins mutations are mainly located in the loop following the C-helix; however, insertion sequences are highly variable due to the heterogeneity of the insertion position and length. The 2 most common insertions (i.e., A767_V769dup and S768_D770dup) comprise only approximately 45% of EGFR ex20ins cases (7,9,11,16). Some studies have reported on the distribution of EGFR ex20ins mutations in different races; however, the racial differences in the genetic and clinical characteristics are not yet well defined. In addition, although a study had reported the distribution of EGFR ex20ins in Chinese lung cancer patients, there was no clear frequency data on LUAD (9). According to recent genomic studies of 3 large cohorts, EGFR ex20ins cases represent 13.9% (237/1,705) of EGFR-mutant LUAD cases in the US (10), but only 4.7% (31/662) to 4.8% (49/1,021) of EGFR-mutant LUAD cases in China (11,19).

We investigated the molecular and clinicopathologic features of EGFR ex20ins in 3,892 Chinese LUAD patients by next-generation sequencing (NGS). We also compared the frequency and distribution of EGFR ex20ins mutations between Chinese and Western LUAD patients by Chi-square test and clustering analyses. To our knowledge, this is the first quantitative comparison of the distribution of EGFR ex20ins in different populations by clustering analysis. We present the following article in accordance with the MDAR reporting checklist (available at https://atm.amergroups.com/article/view/10.21037/atm-22-383/rc).

Methods

Patients and specimens

The data of 4,524 lung cancer patients were extracted from the Beijing Chest Hospital and the internal database at Novogene Co., Ltd., China, which contains tumor targeted NGS data from February 2018 to July 2020. Clinicopathological information, including gender, age at diagnosis, and histologic subtypes, was also extracted. All samples were formalin-fixed paraffin-embedded (FFPE) sections, and were reviewed by experienced pathologists to ensure a tumor content ≥20%. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics and Institutional Review Board for Human Investigation of Beijing Chest Hospital (No. YLQX-2020-005), and individual consent for this retrospective analysis was waived.

DNA and RNA extraction

Deoxyribonucleic acid (DNA) was extracted from the FFPE sections using a TIANamp FFPE DNA Kit (TIANGEN, Beijing, China) in accordance with the manufacturer's protocol. Ribonucleic acid (RNA) was extracted using a RecoverAll™ Total Nucleic Acid Isolation Kit (Life Technologies, Eugene, Oregon, USA). The DNA and
RNA were quantified using a Qubit 3.0 fluorimeter (Life Technologies, Eugene, Oregon, USA). RNA integrity was further evaluated by quantitative polymerase chain reaction (qPCR) and agarose gel electrophoresis.

**NGS analysis**

The DNA and RNA libraries were performed using a human EGFR, KRAS, BRAF, PIK3CA, ALK, and ROS1 genomic mutation test kit (semiconductor sequencing method, Tianjin Novogene Bioinformatics Technology Co., Ltd., Tianjin, China) in accordance with the manufacturer’s protocol. This panel covers hotspot mutations in 23 genes (i.e., AKT1, ALK, BRAF, CTNNB1, DDR2, EGFR, ERBB4, FBXW7, FGFR1, FGFR2, FGFR3, HER2, Kras, MAP2K1, MET, NRAS, NOTCH1, PIK3CA, PTEN, STK11, SMAD4, TP53, and UGT1A1) and fusions in 5 genes (i.e., ALK, MET, NTRK1, RET, and ROS1). The libraries were quantified with a Qubit 3.0 fluorometer (Life Technologies, Eugene, Oregon, USA). Sequencing was performed on the Ion Proton platform using the Ion PI™ Sequencing 200 Kit v3 (Thermo Fisher Scientific, Waltham, MA, USA) in accordance with the manufacturer’s protocol. Ion Reporter software (version 5.6) was used for variant calling. Sequencing data with a mean read depth of 2,000x were considered qualified. Single nucleotide variants (SNV) and insertion-deletion (InDel) mutations with a minimum mutation allele frequency ≥1% were reported.

**Results**

**Mutational spectrum of Chinese LUAD patients with EGFR ex20ins mutations**

A total of 4,524 NSCLC specimens were collected and detected by NGS, yielding a median depth of 10,000×. EGFR ex20ins mutations were detected in 77 cases of LUAD, 3 cases of lung squamous cell carcinoma (LUSC), and 1 case of NSCLC not otherwise specified. Almost all patients (95.06%) with EGFRex20ins were LUAD. Thus, the data of 3,892 LUAD patients were further analyzed to determine the clinicopathological and molecular characteristics of EGFR ex20ins.

A total of 2,208 (56.73%) EGFR-mutant LUAD patients were detected. Among them, 1,059 (47.96%) harbored an exon 20 insertion. Compared to the Western LUAD cohorts [Foundation Medicine [FM] data (10) and Memorial Sloan Kettering Cancer Center [MSKCC] data (7); see Figure 1], the frequency of the EGFR ex20ins mutations in the EGFR-mutant patients was significantly lower in our cohort (3.49% vs. 13.90% in FM, χ²=141.34, P<0.001; 3.49% vs. 8.99% in MSKCC, χ²=23.317, P<0.001); however, the frequency of EGFR ex20ins mutations in our cohort was comparable to that of the FM cohort and MSKCC cohort in terms of the total number of LUAD patients (1.98% vs. 2.31% in FM, χ²=1.408, P=0.235; 1.98% vs. 2.20% in MSKCC, χ²=0.266, P=0.606). Additionally, the frequency of EGFR ex20ins mutations in our cohort was comparable to that of other Chinese LUAD cohorts [Sun Yat-sen University Cancer Center data (11) and Fudan University Shanghai Medical College data (19)] in terms of both total LUAD patients and EGFR-mutant LUAD patients.

A total of 23 unique EGFR ex20ins mutations were identified (see Figure 2A). The most prevalent insertion

---

**Statistical analysis**

The EGFR ex20ins data from the other cohorts were collected from other research (7,9-12,19-21). SPSS 22.0 statistics software (SPSS, Inc., Chicago, IL, USA) was used for the statistical analysis. The Chi-square (χ²) test was used to compare the rates among groups with different features. Hierarchical clustering analyses were performed based on Euclidean distance and the average-linkage method. The statistical tests were 2-sided, and the significance level was set as P<0.05.
The mutation was S768_D770dup (n=22, 28.57%), followed by A767_V769dup (n=21, 27.27%), and N771_H773dup (n=4, 5.19%). The following 8 novel EGFR ex20ins subtypes had not previously been recorded in the COSMIC database: A767_S768insSLD, A767_S768insSVG, D770_N771insNPY, P772_H773dup, H773_V774insAH, D770_P772dup, N771_P772insH, and P772_H773insHT. Almost all the EGFR ex20ins (76/77) occurred in the loop following the C-helix, and only 1 EGFR ex20ins (A763_Y764insFQEA) occurred in the C-helix of the tyrosine kinase domain (see Figure 2A). We also conducted hierarchical clustering analyses to compare the data of the other Chinese and Western cohorts (7,9-12,19-21), and found that the distribution of EGFR ex20ins mutations in our cohort was similar to that of the other cohorts, and there was no significant difference between Chinese and Western populations (see Figure 2B).

The co-mutation analysis showed that EGFR ex20ins were significantly and exclusively associated with certain driver genes in LUAD, including ALK fusion ($\chi^2=7.133$, P=0.008), KRAS ($\chi^2=8.468$, P=0.004), and PIK3CA ($\chi^2=5.792$, P=0.016; see Figure 3). Tp53 was the most common mutant gene in LUAD patients with EGFR ex20ins; however, the Chi-square test results were not significant ($\chi^2=3.056$, P=0.080). No gene was observed to be significantly co-mutated with EGFR ex20ins.
Clinicopathologic features of Chinese LUAD patients with EGFR ex20ins mutations

We also compared the basic clinicopathologic features between LUAD patients with EGFR ex20ins and the other patients (see Table 1). The median age of patients with EGFR ex20ins was 62.72 years, which was significantly older than that of patients with the EGFR 19del mutation (P=0.040). Otherwise, no significant relationship was observed in terms of age and gender between patients with EGFR ex20ins and those with EGFR mutations or without EGFR ex20ins.

Discussion

In general, the ex20ins mutation allows the EGFR protein to form a binding pocket that is inaccessible to canonical EGFR TKIs (22). However, different EGFR ex20ins mutations lead to unique kinase activity and responses to treatment (23). Although EGFR mutation testing is recommended in all advanced LUAD (24), the heterogeneity of EGFR ex20ins mutations and resultant challenges in identifying these mutations in routine clinical genetic testing has led to the underestimation of their real frequency (23). The recent development of NGS has enhanced our ability to identify them in routine clinical genetic testing. In this study, we detected EGFR ex20ins mutations in 3,892 Chinese LUAD patients using the human EGFR, KRAS, BRAF, PIK3CA, ALK, and ROS1 genomic mutation test kit (Novogene), which has been approved by National Medical Products Administration as an NGS companion diagnostic test in China. This test produces more accurate diagnoses than polymerase chain reaction tests, which detect <50% of EGFR ex20ins mutations.

This is the largest cohort study of EGFR ex20ins in Chinese LUAD patients. Qin et al. (9) reported on the variability of EGFR ex20ins in more Chinese lung cancer patients; however, their study lacked definitive data on LUAD. In the present study, a total of 23 unique EGFR ex20ins were identified in 77 LUAD patients, accounting 1.98% of all LUAD patients and 3.49% of EGFR-mutant LUAD patients. This is consistent with previous reports on Chinese LUAD patients (11,19). The frequency of EGFR ex20ins between Chinese and Western LUAD patients was further compared by integrating the data of a total of 18,570 LUAD cases reported in different studies. We found that the EGFR ex20ins mutation frequency of EGFR-mutant Chinese LUAD patients was significantly lower than that of EGFR-mutated Western LUAD patients, but the frequency of EGFR ex20ins mutations in the total number of LUDA patients was similar between Chinese and Western LUAD patients.

Consistent with previous reports (7,9-12,20), we found that the 2 most common EGFR ex20ins subtypes were S768_D770dup and A767_V769dup, which could be major targets for future drug development. However, they only comprised 55.84% of all EGFR ex20ins, and 60.87% (14/23) EGFR ex20ins subtypes occurred only once. Additionally, 8 of the EGFR ex20ins subtypes were not recorded in the COSMIC database. Thus, EGFR ex20ins mutations are
Table 1 Clinicopathological characteristics of lung adenocarcinoma patients with EGFR ex20ins mutations and those with other EGFR mutations or without EGFR ex20ins

<table>
<thead>
<tr>
<th>Variables</th>
<th>Exon 20ins</th>
<th>No exon 20ins</th>
<th>Exon 19 deletion</th>
<th>L858R</th>
<th>Other EGFR mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>N</td>
<td>P†</td>
<td>N</td>
<td>P§</td>
<td>N</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>32</td>
<td>1,904</td>
<td>0.170</td>
<td>359</td>
<td>0.670</td>
</tr>
<tr>
<td>Female</td>
<td>43</td>
<td>1,854</td>
<td>535</td>
<td>652</td>
<td>88</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>62.72</td>
<td>61.27</td>
<td>0.252</td>
<td>59.99</td>
<td>0.040</td>
</tr>
<tr>
<td>SD</td>
<td>11.52</td>
<td>10.84</td>
<td>11.03</td>
<td>9.73</td>
<td>10.13</td>
</tr>
</tbody>
</table>

†, comparison between EGFR ex20ins and no exon 20ins; ‡, comparison between EGFR ex20ins and exon 19 deletion; §, comparison between EGFR ex20ins and L858R; ¶, comparison between EGFR ex20ins and other EGFR mutations. EGFR ex20ins, EGFR exon 20 insertion.

highly heterogeneous, which is consistent with previous reports (8-11,20).

The distribution of EGFR ex20ins mutations between Chinese and Western LUAD patients was also compared using a total of 1,124 LUAD cases with EGFR ex20ins from this study and previous research (7,9-12,19-21). The cluster analysis showed that there was no significant difference in the distribution of EGFR ex20ins mutations between Chinese and Western LUAD patients. The differences in the distribution of EGFR ex20ins reported in different studies are likely due to statistical differences caused by small sample sizes, and have nothing to do with race.

In the present study, the EGFR ex20ins mutations were mutually exclusive with other driver mutations, such as ALK fusion, KRAS, and PIK3CA. This is also consistent with previous reports (9-11,19). However, unlike previous studies (9-11), no other gene was observed to co-mutate with EGFR ex20ins, including TP53.

Chinese LUAD patients with EGFR ex20ins are characterized by a significantly younger age and shorter recurrence-free survival compared to those with typical EGFR mutations (19). However, in here, the patients with EGFR ex20ins were significantly older than those with the EGFR 19del mutation. These patients had a similar age and gender to patients with other EGFR mutations or without EGFR ex20ins. A limitation of this study is the lack of sufficient clinicopathological information, such as TNM stage, smoking history, survival, and EGFR TKIs treatment effects. Thus, the response of different EGFR ex20ins subtypes to EGFR TKIs needs to be explored in future studies.

In conclusion, we described the molecular and clinicopathologic features of EGFR ex20ins in 3,892 Chinese LUAD patients. We found no significant difference in the frequency and distribution of EGFR ex20ins mutations between Chinese and Western LUAD patients; however, the frequency of EGFR ex20ins mutations was significantly lower in EGFR-mutant Chinese LUAD patients. These findings can be used to develop drugs and in clinical trials targeting EGFR ex20ins.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the MDAR reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-22-383/rc

Data Sharing Statement: Available at https://atm.amegroups.com/article/view/10.21037/atm-22-383/dss

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-22-383/coif). ZW, DJ, and MZ are from Beijing Novogene Bioinformatics Technology Co., Ltd. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work, including ensuring that any questions related to the accuracy or integrity of any part of the work...
have been appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics and Institutional Review Board for Human Investigation of Beijing Chest Hospital (No. YLQX-2020-005), and individual consent for this retrospective analysis was waived.

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
