Actionable Mutation Profiles of Non-Small Cell Lung Cancer patients from Vietnamese population

Anh-Thu Huynh Dang, Vu-Uyen Tran, Thanh-Truong Tran, Hong-Anh Thai Pham, Dinh-Thong Le, Lam Nguyen, Ngoc-Vu Nguyen, Thai-Hoa Thai Nguyen, Chu Van Nguyen, Ha Thu Le, Mai-Lan Thi Nguyen, Vu Thuong Le, Phuc Huu Nguyen, Binh Thanh Vo, Hong-Thuy Thi Dao, Luan Thanh Nguyen, Thien-Chi Van Nguyen, Quynh-Tram Nguyen Bui, Long Hung Nguyen, Nguyen Huu Nguyen, Quynh-Tho Thi Nguyen, Truong Xuan Le, Thanh-Thuy Thi Do, Kiet Truong Dinh, Han Ngoc Do, Minh-Duy Phan, Hoai-Nghia Nguyen, Le Son Tran, & Hoa Giang

Comprehensive profiling of actionable mutations in non-small cell lung cancer (NSCLC) is vital to guide targeted therapy, thereby improving the survival rate of patients. Despite the high incidence and mortality rate of NSCLC in Vietnam, the actionable mutation profiles of Vietnamese patients have not been thoroughly examined. Here, we employed massively parallel sequencing to identify alterations in major driver genes (EGFR, KRAS, NRAS, BRAF, ALK and ROS1) in 350 Vietnamese NSCLC patients. We showed that the Vietnamese NSCLC patients exhibited mutations most frequently in EGFR (35.4%) and KRAS (22.6%), followed by ALK (6.6%), ROS1 (3.1%), BRAF (2.3%) and NRAS (0.6%). Interestingly, the cohort of Vietnamese patients with advanced adenocarcinoma had higher prevalence of EGFR mutations than the Caucasian MSK-IMPACT cohort. Compared to the East Asian cohort, it had lower EGFR but higher KRAS mutation prevalence. We found that KRAS mutations were more commonly detected in male patients while EGFR mutations were more frequently found in female. Moreover, younger patients (< 61 years) had higher genetic rearrangements in ALK or ROS1. In conclusions, our study revealed mutation profiles of 6 driver genes in the largest cohort of NSCLC patients in Vietnam to date, highlighting significant differences in mutation prevalence to other cohorts.

Lung cancer is the most common malignancy and the leading cause of cancer related deaths worldwide (18.4% of total cancer deaths), with non-small cell lung cancer (NSCLC) being the most common subtype, accounting for approximately 85% of all diagnosed cases. The majority of NSCLC patients display advanced disease when diagnosed and thus have poor prognosis. It is well established that acquired genetic alterations in certain driver genes result in tumour growth and invasiveness, and that patients harboring certain mutations may benefit from targeted therapies. Indeed, a randomized clinical trial reported that advanced NSCLC patients harboring activating mutations in EGFR, one of the major driver genes of NSCLC, exhibited longer progression-free period when treated with a tyrosine kinase inhibitor (TKI), gefitinib, compared to those treated with standard platinum based chemotherapy. However, those who were treated with TKI drugs can acquire secondary resistant mutations, in which case a new treatment regimen is needed to maintain therapeutic effect. In addition to EGFR, NSCLC patients carrying ALK or ROS1 rearrangement were shown to respond well to a different TKI drug.
crizotinib, while BRAF mutated NSCLC patients can be treated with a combination of BRAF inhibitors, dabrafenib and trametinib. These findings suggest that the identification of mutation profiles of NSCLC is critical in order to prescribe suitable TKI therapy as well as elucidate the molecular basis of drug resistance to provide timely treatment adjustment.

Since 2018, the American Society of Clinical Oncology (ASCO) has recommended routine mutation testing for driver genes including EGFR, ALK, ROS1 and BRAF in clinical practice for NSCLC patients. Although there are currently no targeted drugs for KRAS or NRAS mutated NSCLCs, mutation testing for these genes has also been recommended due to their proven impact on clinical outcomes of NSCLC patients. Hence, simultaneous mutation profiling of these six driver genes has been recommended in current clinical practice. Currently, information in publicly available database such as The Cancer Genome Atlas (TCGA) was obtained mostly from prospective studies in Caucasians and European cohorts. Therefore, the impact of heterogeneous genetic constitution of NSCLC patients across different populations might be underestimated.

Vietnam displays higher incidence rate of lung cancer than the average worldwide incidence (14.5% versus 11% of all new cases) despite having the same level of mortality rate according to the latest Globocan data. Also, in Vietnam lung cancer is the most aggressive type of cancer in men and ranks as the second leading cause of cancer deaths in women. Therefore, it is important to assess the mutation landscape as well as their association with unique clinicopathological features of Vietnamese NSCLC patients. In the present study, we employed massively parallel sequencing to detect genetic changes in six major driver genes including EGFR, KRAS, NRAS, BRAF, ALK and ROS1 in tumour tissues from 350 NSCLC patients in Vietnam. Interestingly, we found that the Vietnamese cohort had a significantly higher frequency of KRAS mutations as compared to Caucasians and East Asian cohorts. We further identified significant associations between the prevalence of these mutations with patients’ clinical features in the Vietnamese cohort. Our data provide valuable information for guiding targeted therapy and drug development for Vietnamese NSCLC patients.

Results

Clinical characteristics of Vietnamese NSCLC patients. The cohort in this study comprised of 350 patients diagnosed with NSCLC by clinical histology from 4 hospitals in Vietnam, with higher percentage of male compared to female (60.6% versus 37.4%, p < 0.05) and the median age of 61 years, ranging from 24 to 89 years (Table 1). The majority of patients (289 cases, 82.6%) were classified into advanced stages (III-IV), while 12 patients (3.4%) were in early stages (I-II) and 49 cases (14%) missing information on clinical stages. Among 350 patients, 185 cases (52.8%) had smoking status recorded including 54 smokers (15.4%) and 131 non-smokers (37.4%) while 165 cases (47.1%) had missing smoking status. Treatment details were obtained for 306 patients (87.4%), including 285 patients (81.4%) who had naïve to treatment, 5 (1.4%) resection, 10 (2.9%) chemotherapy/radiation and 6 (1.7%) TKI. Table 1. Major clinicopathological features of 350 Vietnamese NSCLC patients: N: number of cases; AC: adenocarcinoma; SCC: squamous cell carcinoma.
never taken any treatment at the time of diagnosis and 21 patients (6%) experienced either tumour resection (5 cases, 1.4%), chemotherapy combined with radiotherapy (10 cases, 2.9%) or TKI therapy (6 cases, 1.7% for each treatment) (Table 1).

**Mutation profiles of driver genes in Vietnamese NSCLC patients.** In the present study, we developed a targeted capture sequencing assay to analyse genetic alterations in formalin-fixed paraffin-embedded (FFPE) tissue biopsy specimens of NSCLC patients. We first validated our assay by comparing its performance with a commercial droplet digital PCR (ddPCR) assay (Bio-rad) for detecting three major **EGFR** mutations (L858R, del19 and T790M) in 40 tissue samples randomly selected from our cohort. When ddPCR results were used as reference standard, our targeted capture sequencing assay exhibited sensitivity of 81.8% (11/13), specificity of 100% (27/27) and concordance rate of 95% (38/40) (Table 2, Table S2). The two cases (LBL021 and L10021) that were positive for del19 mutation by ddPCR but missed by our assay had relatively low variant allele frequency (VAF) of 0.5% and 3.9%, below the limit of detection of our assay (4%) (Table S2). Hence, these results confirmed that our targeted capture sequencing assay achieved precise identification of mutations with VAF >4% in FFPE tissue samples. Therefore, this assay was subsequently used to identify genetic alterations in six major driver genes of NSCLC including **KRAS**, **EGFR**, **NRAS**, **BRAF**, **ALK** and **ROS1** for the cohort of 350 NSCLC patients.

Among 350 patients successfully sequenced, 232 (66.3%) cases were found to carry at least one clinically relevant genetic alteration (according to ClinVar) in the tested driver genes while the remaining 118 cases (33.7%) were negative for these mutations (Fig. 1A). **EGFR** (32.3%) and **KRAS** (20%) were the most frequently mutated driver genes, followed by **ALK** (5.4%), **ROS1** (2.9%), **BRAF** (1.1%) and **NRAS** (0.6%) (Fig. 1A). Although mutations in driver genes such as **EGFR** and **KRAS and ALK** were reported to be mutually exclusive in majority of NSCLC patients20, we detected 14 cases (4%) carrying mutations in more than one driver gene (Fig. 1B). Of those, the co-occurrence of mutations in **EGFR** and **KRAS** was the most common (6 cases), including one case carrying concurrent mutations in 3 driver genes (**EGFR**, **KRAS** and **BRAF**). **EGFR** mutation was also found in 3 cases with

| ddPCR | NGS | Mutation | Wild type | Total | Results  |
|-------|-----|----------|-----------|-------|----------|
|       |     |          |           |       | Sensitivity (%) | Specificity (%) | Concordance (%) |
| Mutation | 11  | 2        | 13        |       | 84.6% (95% CI:54.5%–98.1%) | 100% (95% CI:87.2%–100%) | 95% (95% CI:83.1%–99.4%) |
| Wild type | 0   | 27       | 27        |       |           |           |              |
| Total | 11  | 29       | 40        |       |           |           |              |

**Figure 1.** The mutation composition in six major driver genes in 350 Vietnamese NSCLC patients. (A) Prevalence of mutations in 6 major driver genes determined by targeted capture sequencing. Cases with mutations occurring in more than one driver gene were counted as "co-mutation". (B) Mutation frequency in cases harbouring co-mutations.
ALK rearrangement and 2 cases with BRAF mutation. In addition, KRAS mutations were also detected in patients carrying BRAF (2 cases), ALK (1 case) and ROS1 (1 case) mutations (Fig. 1B).

We next explored the distribution of mutation sites and subtypes across the six driver genes. EGFR mutations were predominantly detected in exon 19 and exon 21, with activating mutations del19 (55 cases) and L858R (48 cases) accounting for the majority of EGFR mutations (83.1% of all EGFR mutation events) (Fig. 2A). In contrast, mutations reported to be associated with resistance to TKI therapy, including T790M and ins20 in exon 20, were detected in 16 patients (12.9% of total patients, 8 cases for each mutation type) (Fig. 2A). All ins20 mutations

Figure 2. Distribution of mutation subtypes across 6 driver genes in Vietnamese NSCLC. (A–F) The mutation frequencies in particular subtypes of EGFR (A), KRAS (B), BRAF (C), NRAS (D), ALK (E) and ROS1 (F) genes were calculated as percentage of mutant cases in the total number of cases carrying mutations in the corresponding driver gene.
were later confirmed by a real time-PCR based commercial assay Cobas® EGFR Mutation Test (Roche) (Data not shown). Additionally, other rare mutations with prevalence of less than 1% were also detected in exon 3 (R108K) and exon 18 (G719A, G719C, G719D and E709K) (Fig. 2A). The most common KRAS mutations occurred in exon 2 at G12 and G13 residues, accounting for 62/79 cases (78.4% of total KRAS mutation cases, Fig. 2B). Mutations were also detected in exon 3, residues Q61 (11 cases, 13.9%) and exon 4, residue K117 (11 cases, 13.9%, Fig. 2B). There were 8 cases carrying more than one KRAS mutation subtype and such concurrent mutations tend to arise at the same residue, with 6 out of 8 cases carrying co-mutations at the same residue (1 case at G12, 4 cases at G13 and 1 case at Q61) (Fig. 2B). Mutations in NRAS were detected in exon 3, residue Q61 (Fig. 2C); and mutations in BRAF were detected in exon 4, 11 and 12 (2 cases each) and exon 15, residue V600 (3 cases) (Fig. 2D).

Our assay was able to detect genetic rearrangements in ALK and ROS1. ALK-EML4 translocation was identified as the most common mutation subtype (22/23 cases, 95.6%, Fig. 2E). All ROS1 mutations (11 cases) identified in this cohort were the results of fusion with the following genes: LIRG3 (1 case), SLC34A2 (3 cases), SDC4 (2 cases), CD74 (4 cases) and ERZ (1 case) (Fig. 2F).

Collectively, our data indicated that mutations in EGFR, KRAS are the two most common events occurring in more than half of all tested NSCLC patients in Vietnam, followed by less frequent mutations in other tested genes.

### Table 3. Comparison of mutation frequency among the three cohorts.

| Characteristics | Vietnam (N = 220) | MSK-IMPACT (N = 764) | East Asia (N = 361) |
|-----------------|------------------|----------------------|---------------------|
|                 | N %              | N %                  | N %                |
| **Sex**         |                  |                      |                    |
| Female          | 90 40.9          | 447 58.5             | 201 55.7           |
| Male            | 128 58.2         | 317 41.5             | 160 44.3           |
| **Unsure**      | 2 0.9            | 0 0.0                | 0 0.0              |
| **Age at diagnosis** |                |                      |                    |
| Min             | 31 19            | 0.9                  | 0 0.0              |
| Max             | 86 43            | 93 36                | 84 23              |
| **Median**      | 61 64            | 58 58                |                    |
| **EGFR**        |                  |                      |                    |
| Mutant          | 83 37.7          | 222 29.1             | 265 73.4           |
| WT              | 137 62.3         | 542 70.9             | 96 26.6            |
| **KRAS**        |                  |                      |                    |
| Mutant          | 47 21.4          | 188 24.6             | 33 9.1             |
| WT              | 173 78.6         | 576 75.4             | 328 90.9           |
| **NRAS**        |                  |                      |                    |
| Mutant          | 1 0.5            | 10 1.3               | Not tested         |
| WT              | 219 99.5         | 754 98.7             |                    |
| **BRAF**        |                  |                      |                    |
| Mutant          | 5 2.3            | 26 3.4               | 5 1.4              |
| WT              | 215 97.7         | 738 96.6             | 356 98.6           |
| **ALK**         |                  |                      |                    |
| Mutant          | 17 7.7           | 30 3.9               | 31 10.0            |
| WT              | 203 92.3         | 734 96.1             | 330 90.0           |
| **ROS1**        |                  |                      |                    |
| Mutant          | 5 2.3            | 23 3.0               | 6 1.7              |
| WT              | 215 97.7         | 741 97.0             | 355 98.3           |

In summary, the cohort of Vietnamese NSCLC patients showed specific characteristics that set it apart from the two other cohorts. It had higher prevalence of EGFR mutations than the Caucasian MSK-IMPACT cohort but lower than the East Asian cohort while its KRAS mutation prevalence is higher than the East Asian cohort.
Correlation between mutation prevalence and clinicopathological features of Vietnamese NSCLC patients. Previous studies have reported significant association between prevalence of driver mutations and patients’ clinicopathological features. However, the results are often inconsistent across different studies. Here, we examined such associations in the Vietnamese NSCLC cohort.

Gender. Gender status was available in 343 patients including 131 female and 212 male patients (1:1.6 ratio); the 7 patients with unknown sex were excluded from gender association analysis. We performed Chi-squared ($\chi^2$) test to investigate the association between patients’ gender and mutation prevalence. Consistent with previous studies, we found that $EGFR$ mutations were more commonly detected in Vietnamese female patients than in male patients (48.1% versus 26.9%, $p < 0.00001$, Table 4). Conversely, $KRAS$ mutation frequency was significantly higher in male than in female patients (30.7% versus 9.2%, $p < 0.0001$, Table 4). Other driver genes including $NRAS$, $BRAF$, $ALK$ and $ROS1$ did not show any significant correlation.

Age. Patient age was available in 344 patients, ranging from to 24 to 89 years old. When the median age (61 years) was used as a cutoff value, $KRAS$ mutations were more frequently detected in elderly patients aged over 61 years than younger individuals (25.3% versus 20.1%, $p < 0.05$, Table 4). In contrast, the younger group showed higher prevalence of $ALK$ (9.8% versus 3.5%, $p < 0.05$) and $ROS1$ (5.2% versus 1.2%, $p < 0.05$) mutations (Table 4). There was no significant correlation between patient age and mutation prevalence of other driver genes including $EGFR$, $NRAS$ and $BRAF$ (Table 4).

Smoking status. Of the 185 cases with smoking status, we could detect a statistically significant correlation between $EGFR$ and $KRAS$ mutation prevalence and smoking status (Table 4). Our data indicated that non-smokers showed significantly higher frequency of $EGFR$ mutation (51.9% versus 29.6%, $p < 0.001$) but lower frequency of $KRAS$ mutation (6.9% versus 29.6%, $p < 0.001$) than smokers.

Histology. Adenocarcinoma (AC) were diagnosed in 241 patients, accounting for the most common histological type (81%), while squamous cell carcinoma (SCC) were identified in 25 cases (8.4%). We detected a significant association between histology type and $EGFR$ mutation with higher prevalence in AC group than SCC group (35.7% versus 16%, $p < 0.05$) (Table 4).
Table 4. Association between clinical factors and mutation frequencies of NSCLC driver genes. N: total case number; n: number of particular group; %: percentage of particular cases in total number cases. Chi squared ($\chi^2$) test (sample size $\geq 5$) or Fisher's exact test (sample size $< 5$) was performed to estimate p value.

| Clinical characteristics | Total | EGFR Not mutated | Mutated p value | KRAS Not mutated | Mutated p value | NRAS Not mutated | Mutated p value |
|--------------------------|-------|------------------|----------------|------------------|----------------|-----------------|----------------|
| Sex                      |       |                  |                |                  |                |                 |                |
| Female                   | 131   | 68               | 119            | $<0.0001$        |                |                 |                |
| Male                     | 212   | 155             | 147            | $<0.0001$        |                |                 |                |
| Unknown                  | 7     | 3               | 5              | NS               |                |                 |                |
| Age                      |       |                  |                |                  |                |                 |                |
| $\leq$61                 | 174   | 112             | 139            | $0.05$           |                |                 |                |
| >61                      | 170   | 112             | 127            | $<0.0001$        |                |                 |                |
| Smoking status           |       |                  |                |                  |                |                 |                |
| Yes                      | 54    | 38              | 38             | NS               |                |                 |                |
| No                       | 131   | 63              | 122            | $<0.01$          |                |                 |                |
| Unknown                  | 165   | 53              | 93             | NS               |                |                 |                |
| Histology                |       |                  |                |                  |                |                 |                |
| ACC                      | 241   | 155             | 185            | $<0.05$          |                |                 |                |
| SCC                      | 25    | 21              | 19             | NS               |                |                 |                |
| Unknown                  | 84    | 50              | 67             | NS               |                |                 |                |
| Clinical characteristic  | Total | BRAF Not mutated | Mutated p value | ALK Not mutated | Mutated p value | ROS1 Not mutated | Mutated p value |
| Sex                      |       |                  |                |                  |                |                 |                |
| Female                   | 131   | 128             | 120            | NS               |                |                 |                |
| Male                     | 212   | 207             | 201            | NS               |                |                 |                |
| Unknown                  | 7     | 7               | 6              | NS               |                |                 |                |
| Age                      |       |                  |                |                  |                |                 |                |
| $\leq$61                 | 174   | 171             | 157            | $<0.05$          |                |                 |                |
| >61                      | 170   | 165             | 164            | $<0.05$          |                |                 |                |
| Smoking status           |       |                  |                |                  |                |                 |                |
| Yes                      | 54    | 54              | 53             | NA               |                |                 |                |
| No                       | 131   | 127             | 122            | NS               |                |                 |                |
| Unknown                  | 165   | 159             | 144            | NS               |                |                 |                |
| Histology                |       |                  |                |                  |                |                 |                |
| ACC                      | 241   | 235             | 223            | NS               |                |                 |                |
| SCC                      | 25    | 24              | 25             | NS               |                |                 |                |
| Unknown                  | 84    | 83              | 79             | NS               |                |                 |                |

common mutation sites: *EGFR* exon 19 deletion (del19) and exon 21 (L858R)\(^{35,36}\), *KRAS* exon 2 (G12C)\(^{37}\), *BRAF* V600E\(^{38,39}\) and *ALK-EML4* fusion\(^{40}\). The frequencies of *ALK* (5.4%) and *ROS1* (2.9%) mutations determined in our study (Fig. 1) are comparable to previously published studies reporting the frequencies of 5.05% and 1%-2% for *ALK* and *ROS1*, respectively\(^{40,41}\). *EGFR* del19, *EGFR* L858R, *BRAF* V600E, *ALK-EML4* and *ROS1* fusion mutations in combined (139 cases) accounted for 39.7% of cases in the Vietnamese cohort. They were known as activating mutations and clinically proven to be sensitive to treatments with available TKI drugs\(^{5,6,11,39}\). Thus, our findings suggested that approximately 40% of Vietnamese patients would carry such mutations and therefore would benefit from available targeted drugs. Among 6 patients currently known to be on TKI therapy (Table 1), 5 cases carrying activating *EGFR* mutations (*EGFR* del19, *EGFR* L858R) and one case positive for both *EGFR* L858R and *ALK-EML4* fusion.

In contrast, some patients with activating *EGFR* mutations were found to develop an acquired resistance mutation, T790M\(^{42,43}\). Consistent with previous studies, we found that 7 out of 8 T790M cases were also positive for either L858R or del19 mutations and we suspected that these patients were under treatment with TKI drugs although we could not obtain treatment data for these cases. In addition to T790M, ins20 mutations were also known as resistant mutations\(^{44,45}\) and were detected in 8 cases in our cohort and most of them (6/8 cases) did not co-exist with any activating *EGFR* mutations L858R and del19, suggesting that ins20 mutations are likely primary inactivating mutations rather than acquired resistant mutations. Although a significant proportion of Vietnamese NSCLC patients were reported to harbour concomitant driver mutations\(^{46,47}\) while 5% of patients in another cohort of 1,000 NSCLC patients at The NCI's Lung Cancer Mutation Consortium were reported to harbour concomitant driver gene mutations\(^{48}\). However, these studies together with our study consistently reported the *EGFR/KRAS* (6/14 cases in our study) as the most common co-mutation event in NSCLC patients\(^{33,48}\). The identification of patients...
with such co-mutations is of clinical importance since these concurrent mutations represent a distinct subset of patients and may have significant impact on treatment outcomes. In this regard, previous studies showed that patients carrying EGFR/ALK co-mutations varied in their sensitivity to and that the choice between these two classes of TKI drugs as first-line treatment for these patients is still being debated\(^5\). Hence, further studies are required to investigate clinical activity and drug sensitivity of different co-mutation subsets in order to develop suitable treatment approaches.

To identify Vietnamese-specific mutation profiles in NSCLC patients, we selected patients with comparable histology (AC) and tumour stage (stage III-IV) from the East Asian cohort (China)\(^23\) and the MSK-IMPACT cohort mainly consisting of Caucasian patients\(^21,22\). The prevalence of EGFR mutations among Vietnamese NSCLC patients was markedly lower than East Asia cohort (37.7% versus 73.4%, \(p < 0.00001\), Table 3) but significantly higher than MSK-IMPACT cohorts (37.7% versus 29.1%, \(p < 0.05\), Table 3), confirming previous reports that EGFR mutations are more prevalent in Asian patients than in Caucasian patients\(^36\). Of note, the percentage of Vietnamese patients with KRAS mutation including those with concurrent mutations was comparable to the MSK-IMPACT cohort but significantly higher than the published data in East Asian (21.4% versus 9.1%, \(p < 0.0001\)). Hence, our results demonstrated that NSCLC patients from Vietnamese population exhibit a unique mutation constitution, suggesting that ethnic composition might contribute to the observed variation in mutation profiles. Furthermore, Nguyen et al.\(^32\) reported a remarkably higher frequency of KRAS mutations in Vietnamese patients living in Vietnam than in those living in the USA (24.4% versus 4.5%), suggesting that geographic and socioeconomic disparities might also contribute to the variation in mutation frequencies in different cohorts. Although the clinical significance and mechanisms driving these variations are unclear, the unique mutation profiles should be taken into consideration for prioritizing research programs aiming to develop new treatment strategies for Vietnamese NSCLC patients.

KRAS mutations were detected in 30% of NSCLC patients who were non-responsive to TKI treatment\(^31\). Hence, the high prevalence of KRAS mutations in the Vietnamese NSCLC cohort might have negative impacts on clinical outcomes. However, the use of KRAS mutation status as a negatively predictive marker of TKI therapy remains controversial due to inconsistent results obtained from subsequent meta-analysis studies\(^52–55\). Interestingly, KRAS mutations, particularly the most prevalent subtype G12C, when co-existing with PD-L1 expression in patients’ tumour were shown to have poor prognosis\(^56\), supporting the potential benefit of KRAS mutation testing in selecting patients for immunotherapy using check-point inhibitor.

We further investigated correlations between mutation prevalence and major patients’ clinical characteristics. Consistent with previous studies\(^25,50\), we observed that EGFR mutations were more prevalent in female patients, non-smoker and those with histological subtype of AC. Unlike EGFR mutations, KRAS mutations were more commonly detected in male patients and showed significant correlations with patients’ age, with higher prevalence in elder patients. It is possible that the high prevalence of KRAS mutations in Vietnamese male patients may be responsible for their higher mortality rate. Previous studies reported that KRAS mutations more frequently arise in smokers than in non-smokers\(^25,57\). Consistently, we observed such correlation in our study, indicating that the high frequency of KRAS mutation in our cohort could be attributable to high prevalence of smoking in Vietnam population. In addition, we found that ALK and ROS1 rearrangement mutations were more common in younger patients as compared to elderly patients, which is consistent with previous studies\(^58,59\). Together, our data revealed several significant correlations between driver gene mutation prevalence and patients’ clinical characteristics.

There are a few limitations in our study. Firstly, although the panel of driver genes used in this study was chosen based on ASCO guidelines\(^5\), we did not take into account mutations in other driver genes such as PIK3CA\(^38\), AKT1\(^39\) and ERBB2\(^60\), previously reported to co-exist with those detected mutations and possibly having significant clinical impacts. Secondly, when comparing the mutation profiles of the Vietnamese cohorts with MSK-IMPACT and East Asian cohort, we selected patients with AC in late stages (stage III-IV) to exclude the confounding effect of histological subtypes and tumour stage that are varied among three cohorts. However, there were still differences in patients’ age at diagnosis and gender ratio between the Vietnamese and the other two cohorts, which were identified be significant factors associated with EGFR and KRAS mutation prevalence, thus might have impact on the analysis of the mutation profiles. Future large-scale studies are required to assess whether these confounders contribute to the variations in mutation profile between the of Vietnamese population and other races.

In conclusions, our study revealed the mutation profiles of multiple driver genes in the largest cohort of NSCLC patients in Vietnam to date. Our data highlighted several subsets of Vietnamese NSCLC patients carrying specific mutations that would benefit from future studies to provide more suitable treatment options.

Material and Methods

Tumour tissues. We studied 350 formalin fixed, paraffin-embedded tumour specimens from NSCLC patients treated at Pham Ngoc Thach hospital, Cho Ray hospital, Ha Noi Oncology hospital and Vietnam National cancer hospital. The tumour-rich areas of the tissues that contain at least 20% of tumour cells identified by a hematoxylin and eosin staining were micro-dissected. Written informed consents were obtained from all patients. Clinical characteristics of all patients were summarised in Table S1. This study was approved by The Ethic Committee of University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam (Ethic number: 027/DHYD-HD) and The Medical Genetics Institute. All methods were performed in accordance with the relevant guidelines and regulations.

DNA isolation. DNA was extracted from FFPE samples using QIaamp DNA FFPE Tissue Kit (Qiagen, USA) following the manufacturer’s instructions and then quantified using the QuantiFlour dsDNA system (Promega, USA).
Massively parallel sequencing. DNA fragmentation and library preparation were performed using the NEBNext Ultra II FS DNA library prep kit (New England Biolabs, USA) following the manufacturer’s instructions. DNA library concentrations were quantified with a QuantiFlour dsDNA system (Promega, USA). Equal amounts of libraries (150 pg per sample) were pooled together and hybridized with xGen Lockdown probes for six targeted genes EGFR, KRAS, NRAS, BRAF, ALK and ROSI (IDT DNA, USA). For ALK and ROSI, customized probes (Table S3) for intron regions were designed and mixed with probes for exon regions at equal concentration. Sequencing was run using NextSeq. 500/550 High output kits v2 (150 cycles) on Illumina NextSeq. 550 system (Illumina, USA) with minimum target coverage of 100×. In cases where the mean coverage in the targeted regions is lower than 100×, extra sequencing was performed to increase the mean coverage to the expected range. The mean coverage in the target regions for all samples is approximately 129×.

Variant calling using Mutect2 and Factera. Each FFPE sample was barcoded with dual indexes in the P7 and P5 primer. The PE reads were generated by bcl2fastq package (Illumina) and aligned to human genome (hg38) using BWA package68. Duplicate reads were marked using MarkDuplicates from Picard tools (http://broadinstitute.github.io/picard/). Somatic variants were called using Mutect2 package62. A custom pipeline with call to BWA, Picard, and Samtools packages were built to perform the above-mentioned analysis steps68. For detection of ALK and ROSI rearrangement, fusion variant calling was analyzed using Factera v1.4.4 with default parameters66.

ddPCR method. A four-step ddPCR procedure was performed using reagents and equipment from Bio-Rad (unless otherwise stated) following the manufacturer’s instruction66. Briefly, the PCR mix was first prepared by mixing 1 x ddPCR Supermix for Probes, primers and probes (IDT DNA) and DNA template (0.8 or 1.6 ng). Next, 20 μl of the PCR mix was transferred into the Droplet Generator DG8TM Cartridge followed by 70 μl of the Droplet Generation Oil before placing in a QX100TM Droplet Generator to generate droplets. Subsequently, the droplets were transferred to a 96-well plate before placing in a thermal cycler (C1000 Touch, Bio-Rad) for PCR amplification. The PCR thermal program was performed as follows: 95 °C for 10 min, then 40 successive cycles of amplification (94 °C for 30 sec; 55 °C for 60 sec) and 98 °C for 10 min. Lastly, the droplet reading was acquired by the QX 200 Droplet reader and analyzed using the QuantaSoft Software. Positive and negative droplets are assigned based on the fluorescence threshold that was set as previously described by Deprez et al.66.

To detect T790M and L858R mutations in exon 20 and 21 of the EGFR gene, one reaction of ddPCR was used with two sets of primers and probes as follows: T790M primer F-GGCTGGCTGGGATCTTG; T790M primer R-TCTTTTGTTCGGCAGATGTG; T790M mutation probe FAM-ATAGACTGCA TGATGAG-ZEN/3‘IBFQ; L858R primer F-GGACAGATGTCAGCATGAGAGT; L858R primer R-CCTCCCTT CTGCGATGATTCTTCTTCT; L858R mutation probe HEX-AGTTTGGCCGCGCA-ZEN/3‘IBFQ. For detection of 15 deletion sites in exon 19 (del19) of the EGFR gene, a commercially available ddPCR reaction (Bio-rad) was used (ddPCR66 EGFR Exon 19 Deletions Screening Kit #12002392).

Statistical analysis. Pearson’s chi-squared (χ²) test (sample size >5) or Fisher’s exact test (sample size <5) was performed on the web page ‘Social Science Statistics’ (http://www.socscistatistics.com) to assess the association between two categorical variables (Tables 3 and 4). Bonferroni correction was applied when multiple comparisons were performed (Table 3).

Received: 5 September 2019; Accepted: 9 January 2020; Published online: 17 February 2020

References
1. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer J. clinicians 68, 394–424, https://doi.org/10.3322/caac.21492 (2018).
2. Sher, T., Dy, G. K. & Adjei, A. A. Small cell lung cancer. Mayo Clin. Proc. 83, 355–367, https://doi.org/10.1016/j.mayocp.2017.09.015 (2018).
3. Birring, S. S. & Peake, M. D. Symptoms and the early diagnosis of lung cancer. Thorax 60, 268–269, https://doi.org/10.1136/thrx.2004.032698 (2005).
4. Tafe, L. J. et al. Clinical Genotyping of Non-Small Cell Lung Cancers Using Targeted Next-Generation Sequencing: Utility of Identifying Rare and Co-mutations in Oncogenic Driver Genes. Neoplasia 18, 577–583, https://doi.org/10.1016/j.neo.2016.07.010 (2016).
5. Lindeman, N. L. 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. 8, 823–859, https://doi.org/10.1097/JTO.0000000000000515 (2013).
6. Mok, T. S. et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N. Engl. J. Med. 361, 947–957, https://doi.org/10.1056/NEJMoa0810699 (2009).
7. Sequist, L. V. et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. J. Clin. Oncol. 31, 3327–3334, https://doi.org/10.1200/JCO.2012.44.2806 (2013).
8. Yu, H. A. et al. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 153 patients with EGFR-mutant lung cancers. Clin. cancer research: an. Off. J. Am. Assoc. Cancer Res. 19, 2240–2247, https://doi.org/10.1158/1078-0432.ccr-12-2246 (2013).
9. Horn, L. & Pao, W. EML4-ALK: honing in on a new target in non-small-cell lung cancer. J. Clin. Oncol. 26, 4232–4235 (2009).
10. Shaw, A. et al. Clinical features and outcome of patients with non-small-cell lung cancer who. J. Clin. Oncol. 26, 4247–4253 (2009).
11. Oudgouw, L. et al. FDA Approval Summary: Dabrafenib and Trametinib for the Treatment of Metastatic. Oncologist. 6, 740–745 (2018).
12. Shaw, A. et al. Crizotinib in ROSI-rearranged advanced non-small-cell lung cancer (NSCLC). Ann. Oncol. 30, 1121–1126 (2019).
13. Kalemkerian, G., Narula, N. & Kennedy, E. Molecular Testing Guideline for the Selection of Lung Cancer Patients for. Oncol. Pract. 14, 323–327 (2018).
14. Soda, K., Tomizawa, K. & Mitsudomi, T. Biological and clinical significance of KRAS mutations in lung cancer. Cancer Metastasis Rev. 29, 49–60 (2010).
58. Chaft, J. E. et al. Coexistence of PIK3CA and other oncogene mutations in lung adenocarcinoma-rationale for comprehensive mutation profiling. Mol. Cancer Ther. 11, 485–491 (2012).
59. Rao, G. et al. Inhibition of AKT1 signaling promotes invasion and metastasis of non-small cell lung cancer cells with K-RAS or EGFR mutations. Sci Rep. 7, 2066. 7010.1038/s41598-41017-0612-41599 (2017).
60. Chuang, J. C. et al. ERBB2-Mutated Metastatic Non-Small Cell Lung Cancer: Response and Resistance to Targeted Therapies. J. Thorac. Oncol. 12, 833–842 (2017).
61. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. 25, 1754–1760 (2009).
62. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. 31, 213–219 (2013).
63. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25, 2078–2079 (2009).
64. Newman, A. M. et al. FACTERA: a practical method for the discovery of genomic rearrangements at breakpoint resolution. Bioinforma. 30, 3390–3393 (2014).
65. Bio-Rad Laboratories, I. Rare Mutation Detection Best Practices Guidelines.
66. Deprez, L. et al. Validation of a digital PCR method for quantification of DNA copy number concentrations by using a certified reference material. Biomol. Detect. Quantif. 9, 29–39 (2016).

Acknowledgements
This research is funded by Gene Solutions (GS-004). The funder (Gene Solutions) provided support in the form of salaries for authors V.U.T., H.A.T.P., T.T.T., B.T.V., L.T.N., T.C.V.N., L.H.N., N.H.N., H.N.D., L.S.T. and H.G. but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions
A.T.H.D., V.U.T., H.A.T.P., T.T.T., B.T.V., H.T.T.D., L.T.N., T.C.V.N., Q.T.N.B. and L.H.N. performed experiments. A.T.H.D., D.T.L., L.N., N.V.N., T.H.T.N., C.V.N., H.T.L., M.L.T.N., V.T.I. and P.H.N. recruited patients and performed pathological analysis. N.H.N., Q.T.T.N., T.X.L., T.T.T.D., K.T.D. and H.N.D. designed experiments and analyzed data. M.D.P. revised the manuscript, H.N.N. supervised the project; L.S.T. analyzed the data and wrote the manuscript; H.G. designed experiments and analyzed sequencing data.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-59744-3.

Correspondence and requests for materials should be addressed to H.-N.N., L.S.T. or H.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020