Clinical utility of analyzing circulating tumor DNA in patients with non-small cell lung carcinoma

Haiping Jiang
The Affiliated Hospital of Qingdao University

Ying Zhang
Northeastern University

Lin Zhang
The Affiliated Hospital of Qingdao University

Yongbei Kuang
Northeastern University

Jingyue Li
Northeastern University

Na Sui (snsuina@163.com)
Critical Care Medicine Institute  https://orcid.org/0000-0002-5477-0890

Research Article

Keywords: ctDNA, capture-based targeted sequencing, NSCLC, targeted therapy, mutation, EGFR

Posted Date: April 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1519425/v1

License: ☑️ ☐ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Clinical utility of analyzing circulating tumor DNA in patients with non-small cell lung carcinoma

Haiping Jiang¹,†, Ying Zhang²,†, Lin Zhang³, Yongbei Kuang², Jingyue Li⁴, Na Sui⁵,*

¹ Department of Oncology, The Affiliated Hospital of Qingdao University, Qingdao, Shandong Province 266003, China
² School of Business Administration, Northeastern University, No.195, Chuangxin Road, Shenyang 110167, China
³ Department of Oncology, The Affiliated Hospital of Qingdao University, Qingdao, Shandong Province 266003, China
⁴ College of software, Northeastern University, No.195, Chuangxin Road, Shenyang110167, China
⁵ Department of Critical Care Medicine, The Affiliated Hospital of Qingdao University, Qingdao, Shandong Province 266003, China

†These authors have contributed equally to this work and share first authorship;

*Correspondence:
Na Sui
snsuina@163.com

Keywords: ctDNA, capture-based targeted sequencing, NSCLC, targeted therapy, mutation, EGFR
Abstract

Genotype-directed therapy is critical to the treatment strategy for advanced non-small cell lung carcinoma. Compared with the traditional biopsy sample, ctDNA is easier to sample and monitor in real time. Therefore, we extracted ctDNA from peripheral blood samples of 86 patients with advanced NSCLC for detection. We carried out ultra-high-depth targeted capture sequencing on seven genes. We found that 47.67% (41/86) of patients had at least one potentially manipulable variant. Of the 100 identified variants, 83 were potentially actionable variants. These single nucleotide variations, insertions, and deletions were most common in $EGFR$ (46%), next were $KRAS$ (26%), $MET$ (20%), $PIK3CA$ (15%), $MAP2K1$ (9%), $BRAF$ (4%), $NRAS$ (2%). Twelve new pathogenic mutants were identified by 12 protein function prediction algorithms. These results help us to better understand the mutation of NSCLC.
Introduction

Lung cancer is one of the most common types of human malignancy in the world. Among them, non-small cell lung cancer (NSCLC) is the main type, accounting for >80% of lung cancer cases [1]. Strengthening the understanding of the molecular mechanism of NSCLC and adopting gene detection methods are of great significance and research value for the monitoring and treatment of NSCLC [2]. The diagnostic criteria for NSCLC have been gradually improved, including detection of MET proto-oncogene (MET), epidermal growth factor receptor (EGFR), erb-b2 receptor tyrosine kinase 2 (ERBB2), Kirsten rat sarcoma viral oncogene (KRAS) mutations, and B-Raf proto-oncogene (BRAF) [3].

The heterogeneity of tumors requires both reproducible and invasive biopsy. At present, the difficulty of sampling and dynamic monitoring limits the clinical application of tumor biopsy. In contrast, circulating tumor DNA (ctDNA) has a half-life of only a few hours. Its high turnover rate makes it a valuable tool for providing real-time patient information [4]. In addition, it is easy to obtain and highly invasive. Given these advantages, ctDNA analysis has been used to monitor clinical therapy [5] continuously, identify drug resistance mechanisms [5] and guide targeted molecular treatment [6]. Plasma-derived ctDNA has been proved to be a potential substitute for genome mapping diagnosis, disease surveillance, and drug resistance identification of solid tumors [7]. ctDNA is progressively showing more and more clinical application value. However, ctDNA has not been used in large-scale clinical applications. A few retrospective studies [8-11] have demonstrated the clinical validity of ctDNA in effectively identifying disease progression. However, it is not known how much value ctDNA has in practical clinical applications. We conducted a retrospective study of 86 cases and found 100 mutations in seven genes from ctDNA samples taken from peripheral blood of patients with NSCLC by ultra-high-depth targeted capture sequencing. Potentially actionable variants were most common in EGFR. We also found 12 novel pathogenic mutants.

Materials and methods

Samples collection

Eighty-six plasmas from patients with NSCLC have been gathered. The clinical diagnosis was confirmed by Cytopathology. This study was approved by the Ethics Committee of the Affiliated Hospital of Qingdao University. (approval no.QYFY WZLL 26621) Making sure that every patient signed informed consent and our study strictly followed the guideline released by the National Health and Family Planning Commission of the PRC, our study was launched.

Plasma cell-free DNA extraction

The QIAamp Circulating Nucleic Acid Kit (Qiagen) was used to separate Cell-free DNA from plasma. Agarose gel electrophoresis, the NanoDrop2000 spectrophotometer,
and the Qubit 2.0 fluorimeter (Thermo Fisher Scientific) can be used to evaluate DNA integrity, purity, and concentration. Besides, the library was constructed by DNA samples that conform to specifications.

**Library construction and sequencing**

Following the previous description, library construction was performed [12]. Random fragmentation of DNA was achieved by ultrasound. After that, end repair, phosphorylation, and adaptor ligation were realized step by step. A custom sequence capture-probe (Nimblegen, USA) hybridized the fragments. The amplification was performed through PCR. Sequencing of **EGFR**, **KRAS**, **MET**, **MAP2K1**, **PIK3CA**, **BRAF**, and **NRAS** (Supplementary Table 1) were capture-based ultra-deep targeted and performed on the Illumina Hiseq2500 platform with 2×101 bp paired-end reads (Illumina, San Diego, USA).

**Sequencing data analysis**

In the aftermath of obtaining of raw sequencing data, adaptors and filter low-quality reads were removed by SOAPnuke [13] By the means of bwa-mem2 [14], clean reads were mapped with human genome (hg38). GATK (v 4.1.9.0) [15] was used to mark duplication, call variants, and filter variants. The hard-filtering conditions “**QD < 2.0**”, “**MQ < 40.0**”, “**FS > 60.0**”, “**SOR > 3.0**”, “**MQRankSum < -12.5**”, and “**ReadPosRankSum < -8.0**” were applicable for single nucleotide variants (SNVs) and “**QD < 2.0**”, “**ReadPosRankSum < -20.0**”, “**InbreedingCoeff < -0.8**”, “**FS > 200.0**”, and “**SOR > 10.0**” were applicable for insertions and deletions. After that, VCFtools (v4.2) [16] was used to filter out loci with a depth less than 50. SNV and indel calling required at least five supporting reads. We used ANNOVAR [17] to mark the remaining variants. Stated differently, we excluded polymorphisms, interpreted clinical significance, and predicted the functional impact of sequence variant/splice site changes based on population databases, cancer-specific variant databases, and 12 algorithms, respectively. The specific filtering conditions of SNVs are as follows: 1) Initially, we classified variants with population frequencies, which from Exome Aggregation Consortium (ExAC), ESP6500SI-V2 database and 1000 Genomes Project [18], as single nucleotide polymorphisms (SNPs), which were higher than 1%. Then, we excluded them accordingly. 2) Filter out variants besides splicing region or exonic. Furthermore, the pathogenicity of the SNVs was predicted by SIFT [19], Polyphen2 (HDIV/HVAR) [20], LRT [21], MutationTaster [22], MutationAssessor [23], FATHMM [24], PROVEAN [25], MetaSVM/LR [26], M-CAP [27] and FATHMM-MKL-coding. Thereafter, we evaluated clinical significance of variants through ClinVar [28], Catalog of Somatic Mutations in Cancer (COSMIC) [29], and dbSNP [30]. Eventually, R (v4.0.5) package maftools [31] was used to achieve visualization and analysis of the variants’ mutual exclusivity.

**Statistical analysis**

Frequency and percentage were applied to describe statistics. If P < 0.05, it was reckoned to be statistically significant.
Results

Variant detection and mutation spectrum of 41 NSCLC patients

We conducted a capture-based ultra-high-depth targeted sequencing on peripheral blood samples from 86 patients with NSCLC. On average, we generated 8.5 Gb of sequence for each sample to an average depth of 6,097.4×. To reduce false-positive rates and eliminate common mutations, with comprehensive filtering criteria, we removed variants with population frequencies higher than 0.01 in dbSNP, 1000 Genomes, ESP6500SI-V2, and ExAC database. We identified 100 variants in 86 samples (Supplementary Table 2), including 34 SNVs, 54 insertions, and 12 deletions (Figure 1a). Eighteen of them are categorized as missense SNVs, one is nonsense SNVs, 66 are indels, and two are splicing variants (Figure 1b). Missense SNVs and frameshift indels generally lead to abnormal function of the protein products, rendering NSCLC eventually. Accordingly, these variants may be of highly clinical significance and were included in the subsequent analyses.

In total, 47.67% (41/86) of the patients had at least one potentially actionable variant (Figure 2a). The most common driver mutations were in EGFR (46%) followed by KRAS (26%), MET (20%), PIK3CA (15%), MAP2K1 (9%), BRAF (4%), NRAS (2%). We detected 28 mutations in EGFR, 23 in KRAS, 12 in MET, 11 in MAP2K1, 9 in PIK3CA, 2 in BRAF, and 1 in NRAS (Figure 2b). We observed that C>T alteration was more frequent than other forms (Figure 2c).

Variation functional prediction

We annotated the clinical significance of these mutations based on ClinVar, dbSNP, and COSMIC70 databases. In total, 9, 12, and 6 variants were annotated to the ClinVar, dbSNP, and COSMIC70 databases (Figure 3a). It is widely considered that ClinVar and COSMIC70 database-registered variats (n=12) are functionally important mutations. After comparing with ClinVar, we classified two as drug response, five as benign, and one as variants of uncertain significance (VUS; Supplementary Table 3). Among VUS, there is one had conflicting evidence for pathogenic and benign criteria, while the others did not have enough evidence. Among the 88 novel variants, which were not registered in the dbSNP build 150, ClinVar, and COSMIC70, 12 SNVs were predicted as harmful variants by at least three algorithms. The genes with the highest number of novel deleterious variants were EGFR (n = 3) and MAP2K1 (n = 3), followed by MET (n = 2), PIK3CA (n = 2), and KRAS (n = 2; Figure 3b).

Mutation point on the genes

We summed up the amino acid changes to know which domain is most often affected in this cohort. The domain of the COG1100 protein was greatly mutated (Figure 4a). Specifically, about the seven genes studied, the MAP2K1 variants comprised three missense mutations, four nonframeshift insertions, and three frameshift insertions. The NRAS has only occurred one nonframeshift deletion, and it harbored NRAS exon3 p.V81Yfs*23 mutation. The MET variants included three missense mutations, one frameshift deletions, and 12 frameshift insertions. Among them, 5 (31%, 5/16) samples
harbored MET exon2 p.G181Lfs*59 mutation and exon2 p.A182Rfs*12 mutation, respectively. The BRAF variants consisted of 1 missense mutation and one frameshift deletion. The KRAS variants comprised two missense mutations, one nonframeshift deletions, two nonframeshift insertions, two frameshift deletions, four stopgain, and eight frameshift insertions. The EGFR variants contained seven missense mutations, two nonframeshift deletions, two frameshift deletions, two stopgain, and 15 frameshift insertions. In these, 13 (46%, 2/28) samples harbored EGFR exon16 p.E634Gfs*75 mutation and two (7%, 2/28) samples harbored EGFR exon19 p.E746_A750del mutation. The PIK3CA variants covered two missense mutations, one nonframeshift insertions, and three frameshift insertions. Furthermore, those mutations not detailed about the above seven genotypes occurred only once (Figure 4b-h).

Discussion

We used capture-based ultra-high-depth targeted sequencing to retrospectively investigate peripheral blood samples from 86 patients with NSCLC. From the experiments, we identified 100 variants in all samples, 83 of which were potentially actionable and the most frequent of which were insertion mutations. In addition, 47.67% (41/86) of the patients had at least one potentially actionable variant.

Our mutation spectrums are similar to many mutation spectra of NSCLC that have ever appeared in former papers [32-34]. EGFR (46%) is the most common driver mutations followed by KRAS (26%), MET (20%), PIK3CA (15%), MAP2K1 (9%), BRAF (4%), NRAS (2%) in NSCLC patients. As well, in past non-small-cell lung carcinoma studies, the analysis of mutations in the EGFR gene remained a major focus [35]. Our data also further confirm the importance of these mutations in NSCLC. The EGFR exon16 p.E634Gfs*75, which is a type of frame shift mutation, appears the most frequently in all samples. This mutation results in a structural change in the EGFR gene-directed protein that leads to disease. Another key variation in the sample was p.E746_A750del which also occurred in EGFR. This mutation removes the protein sequence that controls synthesis from 746th E (glutamic acid) to 750th A (alanine), giving rise to no need for EGFR to bind to the ligand, and sustained tyrosine kinase activity, so that some cells grow out of control, causing disease. Besides, because it was the dominant type of EGFR exon19 del, accounting for about 67%, it has mature targeting drugs. Among these drugs, Gefitinib and Erlotinib have a better effect on the deletion of exon 19 mutation in the case of NSCLC [36, 37].

In all, we showed a clear map of the mutation frequency of the associated oncogenes from 86 patients with non-small-cell lung carcinoma, and compared the results of our sample data with the results of previous studies to explore the characteristics and changes of gene mutations in patients who have NSCLC. The data could help clinicians make decisions and provide a potential molecular basis for non-small-cell lung carcinoma treatment.
Data Availability Statement

The data that supports the findings of this study are available in the supplementary material of this article.

Ethics Statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the Affiliated Hospital of Qingdao University. The patients/participants provided their written informed consent to participate in this study.

Author Contributions

Haiping Jiang¹, †, Ying Zhang², †, Lin Zhang³, Yongbei Kuang², Jingyue Li⁴, Na Su⁵,*

NS participated in study conception and design. HJ, YZ and NS enrolled and managed patients. YZ carried out collection and assembly of data. YZ, YK, and JL were involved in data analysis and interpretation. YZ, YK, and JL prepared the manuscript and manuscript figures. HJ and NS edited, critically read, and revised the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

We thank all patients who participated in this study and their families.

Supplementary Material

Supplementary Table 1. Target gene information

Supplementary Table 2. Overview of the variants

Supplementary Table 3. Number of variants categorized as pathogenic, benign and uncertain significance in seven genes for NSCLC.

References


Figure 1. Outcomes of mutation types of patients identified by the seven genes of NSCLC patients. (a) Percentage of SNVs and indels identified by ultra-high-depth targeted capture and sequencing. (b) The number of variants in each category.
Figure 2. Mutations landscapes in NSCLC patients. The characteristics of the genes are summarized in these figures. (a) Significantly mutated genes of 86 patients with NSCLC. The top part shows the number of mutations of each sample. Below, in accordance with the mutation frequency, we ranked the targeted genes. Different types of mutation are also clearly marked by different colors. In all those genes that have mutated, there are some which have mutated multiple times during the process, those particular genes are annotated as Multi_Hit in the above chart. Right, all the mutated genes are shown, result are represented in numbers. At the bottom of the picture, the class of SNVs are shown respectively for all the samples. (b) The proportion of all the genes with mutations. The results are presented in percentage. (c) The chart showcased the SNV class. INS: insertions, DEL: Deletions.
Figure 3. Clinical implications of mutations. (a) Veen plot of database-registered variants. (b) Pie chart of novel harmful variants.
Figure 4. The mutations in protein domain and genes. (a) The Pfam protein domain that is frequently mutated in NSCLC. (b-h) Hemetic representation of mutations in MAP2K1, NRAS, MET, BRAF, KRAS, EGFR, PIK3CA.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.pdf