

Research

Identification of gene variation feature for targeted therapy of non-small cell lung cancer through combined method of DNA and RNA sequencing

Wenguang Pang^{1,2} · Longlong Gong³ · Wangpan Shi⁴ · Hongbo Zheng³ · Min Ye² · Jiarong Chen⁵ · Ronggang Li⁶ · Xin Zhang⁷ · Dong Ren⁸ · Zheng Wang⁹

Received: 10 July 2023 / Accepted: 28 February 2024

Published online: 06 March 2024

© The Author(s) 2024 [OPEN](#)

Abstract

Next generation sequencing (NGS) is typically used to reveal tumor gene variation feature for targeted therapy of various types of human cancers, including non-small cell lung cancer (NSCLC). Here, we report the role and potential applicable value of combining DNA and RNA sequencing in gene variation detection in NSCLC. 386 NSCLC patients with stage II-IV were enrolled and detected using NGS sequencing of DNA and RNA panels that covered all well-documented target driver genes from the Chinese Society of Clinical Oncology (CSCO). The rate of epidermal growth factor receptor (EGFR) single nucleotide variation (SNV)/indel, mesenchymal-epithelial transition factor (MET) copy number variation (CNV) and anaplastic lymphoma kinase (ALK) fusion were 52.1%, 4.1% and 6.0% in the NSCLC cohort. The landscapes of SNV/indel, CNV and gene fusion in the cohort were depicted as well. Meanwhile, we assessed detection efficacy of DNA and RNA sequencing in gene fusion. Detected number and types of gene fusion using the RNA sequencing were better than those using the DNA sequencing. Gene fusion with intergenic region was only detected by DNA sequencing and MET exon 14 skipping (MET Δ ex14) was more easily identified by RNA sequencing. Finally, we investigated clinical correlations of SNV/indel/CNV/fusion with clinicopathologic features in the NSCLC cohort. Taken together, RNA sequencing significantly complements deficiency of DNA sequencing for gene fusion, which cooperatively presents comprehensive and reliable gene variation features and facilitate the identification of potential drug targets for NSCLC patients.

Keywords Next generation sequencing · Non-small cell lung cancer · Single-nucleotide variations · Copy number variations · Gene fusion

Wenguang Pang and Longlong Gong have contributed equally to this work.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12672-024-00915-3>.

✉ Dong Ren, dren3@hs.uci.edu; ✉ Zheng Wang, Dr_wangzheng@163.com; Wenguang Pang, 13631826398@139.com; Longlong Gong, gong.longlong@genecast.com.cn; Wangpan Shi, washi@health.ucsd.edu; Hongbo Zheng, zheng.hongbo@genecast.com.cn; Min Ye, yeminzhangcuiyuan@126.com; Jiarong Chen, garwingchan@163.com; Ronggang Li, lrg760904@163.com; Xin Zhang, zhangx45@mail3.sysu.edu.cn | ¹The First Affiliated Hospital of Jinan University, Guangzhou 510630, China. ²Department of Thoracic Surgery, Jiangmen Central Hospital, Jiangmen 529030, China. ³Genecast Biotechnology Co., Ltd., Wuxi 214104, China. ⁴Department of Pathology, University of California San Diego Health System, San Diego, CA 92037, USA. ⁵Department of Oncology, Jiangmen Central Hospital, Jiangmen 529030, China. ⁶Department of Pathology, Jiangmen Central Hospital, Jiangmen 529030, China. ⁷Clinical Experimental Center, Jiangmen Key Laboratory of Clinical Biobanks and Translational Research, Jiangmen Central Hospital, Jiangmen 529030, China. ⁸Departments of Pathology, University of California Irvine Medical Center, 101 The City Drive South, Orange, CA 92868, USA. ⁹Department of Thoracic Surgery, The 2nd Clinical Medical College of Jinan University, Shenzhen 518020, China.



1 Introduction

Lung cancer remains the leading cause of cancer-related death worldwide, where non-small cell lung cancer (NSCLC) accounts for approximately 85 percent of lung cancer cases [1]. With substantial development and extensive application of targeted drugs against driver mutations, outcome and survival of NSCLC patients have been significantly improved [2, 3]. Notably, the Chinese Society of Clinical Oncology (CSCO) guideline (Version 2022) distributes recent treatment recommendations for NSCLC, including detection and target therapy of single-nucleotide variations (SNV)/insertions/deletions (indel) of epidermal growth factor receptor (EGFR), Kirsten rat sarcoma viral oncogene homolog (KRAS), human epidermal growth factor receptor 2 (HER2) and serine/threonine-protein kinase BRAF genes, copy number variations (CNV) of mesenchymal-epithelial transition factor (MET) and HER2, gene fusions of anaplastic lymphoma kinase (ALK), ROS proto-oncogene 1 (ROS1), rearranged during transfection (RET) and neurotrophin receptor kinase (NTRK), and MET exon 14 skipping (MET Δ ex14). NSCLC patients with driver gene mutations, such as EGFR mutation, presented favorable outcome and survival after treatment of Icotinib, Osimertinib and Gefitinib [4–6]. Alectinib and ceritinib significantly enhanced progression-free survival (PFS) in ALK-rearranged NSCLC patients [7, 8]. Capmatinib showed substantial anti-tumor activity in advanced NSCLC patients with a MET Δ ex14 or MET amplification, particularly in patients without previous treatment history [9]. Therefore, it is conceivable that accurate and comprehensive investigation of specific gene mutation for NSCLC patients might significantly improve the outcome and prognosis of NSCLC patient through using target-drug treatment.

Next generation sequencing (NGS) technology has been undoubtedly recommended by the National Comprehensive Cancer Network (NCCN) and CSCO guidelines for the detection of high-throughput gene variations, which will guide target therapy for NSCLC patients [10]. Genomic DNA (gDNA) and whole exome sequencing (WES) are currently and typically used for analysis of SNV, indel and CNV in tumor tissue samples in NGS-based target therapy selection, and messenger RNA (mRNA) has been more widely adopted as input template for identification of fusions and gene expression analysis. However, mRNA is rarely used to detect somatic mutations such as SNVs or indels due to variable expression levels due to limited accuracy. 46–49% of pathogenic variants identified by DNA sequencing can only be detected through RNA sequencing analysis, especially all telomerase reverse tranase (TERT) mutations failed to be identified by RNA sequencing [11]. Thus, combination of DNA and RNA sequencing could comprehensively and accurately explore gene variation feature as mutual complement for NSCLC patients.

The present study aimed at developing an inexpensive and scalable panel DNA/RNA-seq method capable of detecting multiple types of clinically actionable mutations with high accuracy in NSCLC. 386 NSCLC patients with stage II-IV were enrolled and detected using NGS sequencing of DNA and RNA panels. The DNA panel (769 genes) and RNA panel (29 genes) covered all well-documented target driver genes from the CSCO guideline for NSCLC. We systematically analyzed gene variation feature and explored gene fusion feature between DNA panel and RNA panel sequencing in the NSCLC cohort. Clinical correlations of these gene variations/fusion with clinicopathological feature have been further elucidated in the NSCLC cohort. Therefore, combination of DNA and RNA panel sequencing will facilitate comprehensive identification of gene variation and potential drug targets so as to improve the outcome and prognosis of every single NSCLC patient according to the specific gene mutation. Meanwhile, fewer tumor samples, more detecting depth and less cost of time and material with DNA and RNA panels provided the foundation for wide clinical application in NSCLC.

2 Materials and methods

2.1 Patients and clinical data

386 NSCLC patients with stage II-IV were obtained from Jiangmen Central Hospital and Genecast Clinical Laboratory Center from June 2020 to October 2022. The detailed clinical information was summarized in Table 1. In the cohort, cases with stage II, III and IV occupied 16.3%, 32.9% and 50.8%, respectively. 339 patients (87.8%) were lung adenocarcinoma (LUAD) and 47 patients (12.2%) were lung squamous cell carcinoma (LUSC). The study was reviewed and approved by the Ethics Committee of Jiangmen Central Hospital (2022–119) and all tests were performed according to relevant guidelines.

Table 1 Clinical characteristics

Factor	NSCLC (n = 386)
Age (Median, quartiles)	63 (31–87)
Gender (n, %)	
Male	207 (53.6%)
Female	179 (46.4%)
Smoking (n, %)	
Yes	132 (34.2%)
No	212 (54.9%)
NA	42 (10.9%)
Clinical stage (n, %)	
II	63 (16.3%)
III	127 (32.9%)
IV	196 (50.8%)
Tumor subtype (n, %)	
LUAD	339 (87.8%)
LUSC	47 (12.2%)

2.2 DNA sequencing

Detail methods of DNA sequencing was performed as our previous study [12]. DNA panel was 2,189 kb in length, which covered full coding regions or hotspot mutation regions in 769 genes. The genomic DNA from formalin-fixed paraffin-embedded (FFPE) tumor samples was extracted using the MagPure FFPE DNA Kit B (Magen, China), fragmented into DNA pieces of approximately 200 bp using an enzymatic method (5×FEA Enzyme Mix; Qiagen, CN), subsequently constructed

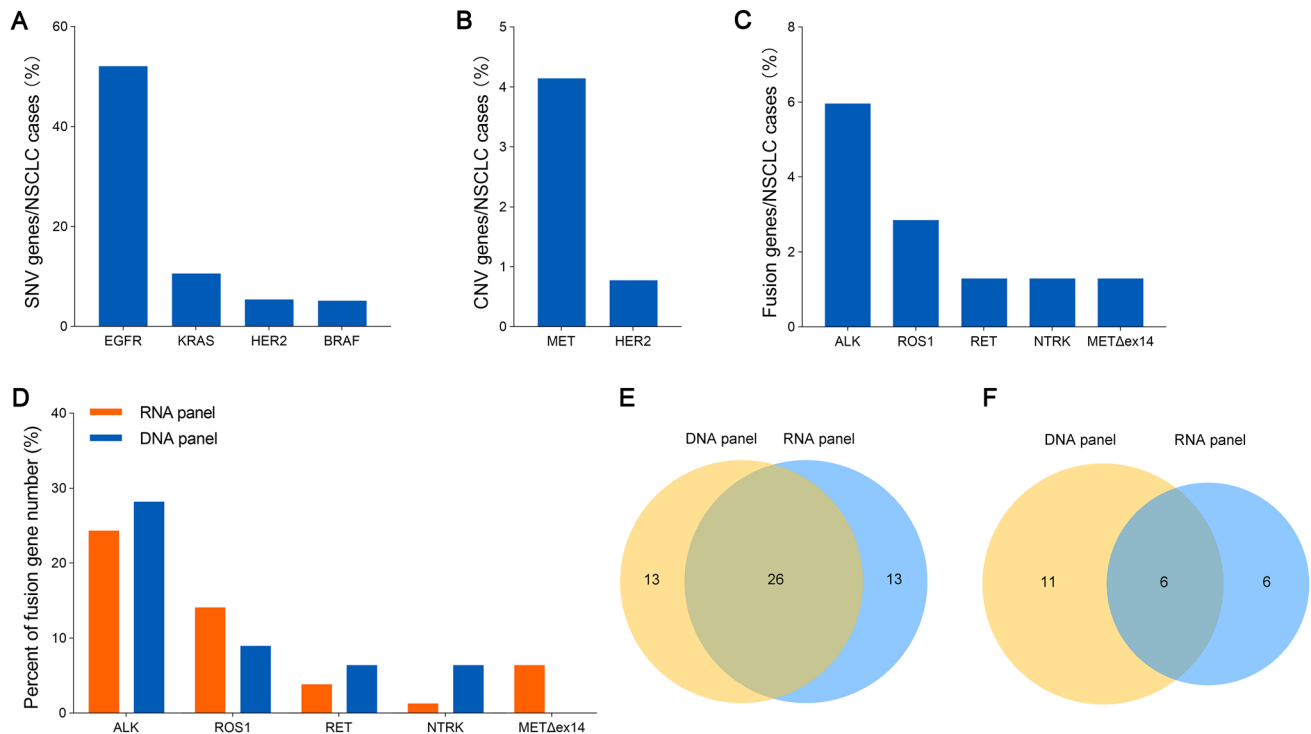


Fig. 1 Roles of NGS detection in diagnostic and treatment of NSCLC according to the Chinese Society of Clinical Oncology (CSCO) clinical guidelines. **A** SNV/indel mutations of EGFR, KRAS, HER2 and BRAF in NSCLC cases (n = 386) detected by DNA panel sequencing. **B** CNV mutations of MET and HER2 in NSCLC cases (n = 386) detected by DNA panel sequencing. **C** ALK, ROS1, RET and NTRK fusion and MET Δ ex14 in NSCLC cases (n = 386) detected by DNA and RNA panel sequencing. **D** Percent of ALK, ROS1, RET and NTRK fusion and MET Δ ex14 detected in NSCLC cases by DNA and RNA sequencing, respectively. **E-F** Gene fusion number **E** and fusion types **F** of ALK, ROS1, RET, NTRK and MET Δ ex14 detected by DNA or RNA panel sequencing

Fig. 2 Landscape of SNV/indel, CNV and gene fusion in the NSCLC cohort. **A** SNV/indel landscape of top 20 genes using DNA panel sequencing in NSCLC cases (n=365). Gene variations contained SNV, insertion, deletion and complex mutations. **B** CNV landscape using DNA panel sequencing in NSCLC cases (n=172). There were 58 genes with CNV in the cohort. Gene copy number was showed using $\log_2(\text{CNV} + 1)$. **C** Gene fusion using DNA and RNA panels sequencing in NSCLC cases (n=386). Orange, blue and green presented fusion genes of RNA panel, DNA panel and common detection with RNA and DNA panels, respectively. **D** Venn diagram of the number of fusion genes detected by DNA panel and RNA panel in NSCLC cases (n=386). **E** Venn diagram of the number of gene fusion types detected by DNA panel and RNA panel in NSCLC cases (n=386)

DNA libraries using a VAHTS Universal DNA Library Prep Kit (Vazyme, China). The libraries of genomic DNA were captured with a 769 gene panel (Table S1) and sequenced using a HyperCap Target Enrichment Kit (Roche, Switzerland) and the NovaSeq 6000 system (Illumina, USA) according to the manufacturer's protocols, respectively.

2.3 RNA sequencing

RNA panel was 126 kb in length, which covered whole transcript sequences of 29 genes. RNA was extracted from FFPE samples using the MagPure FFPE RNA/DNA Kit (Magen, China), and FFPE RNA (50–300 ng) was employed for library construction using mRNA-seq Lib Prep Module for Illumina (ABclonal, China). The RNA panel (29 genes, Table S2) was synthesized by IDT (Integrated DNA Technologies, USA). Hybridization and washing were performed using the xGen Hybridization and Wash Kit and xGen Universal Blockers-TS Mix-96rxn (IDT, USA) according to manufacturer's instructions. The captured products were sequenced on Illumina NovaSeq 6000.

2.4 Statistical analysis

All statistical analysis was performed using GraphPad Prism 8.4.2 and R4.2.1. Analysis SNV/indel and CNV landscapes was performed using a Complex Heatmaps package (2.12.1). Sequence alignment was performed using Integrative Genomics Viewer (IGV) analysis. The Fisher's exact test was used to analyze the frequency difference of SNV and gene fusion between the two/three groups. The unpaired *t* test and One Way ANOVA were used to analyze the difference of CNV between two and three groups, respectively. $p < 0.05$ was considered statistically significant.

3 Results

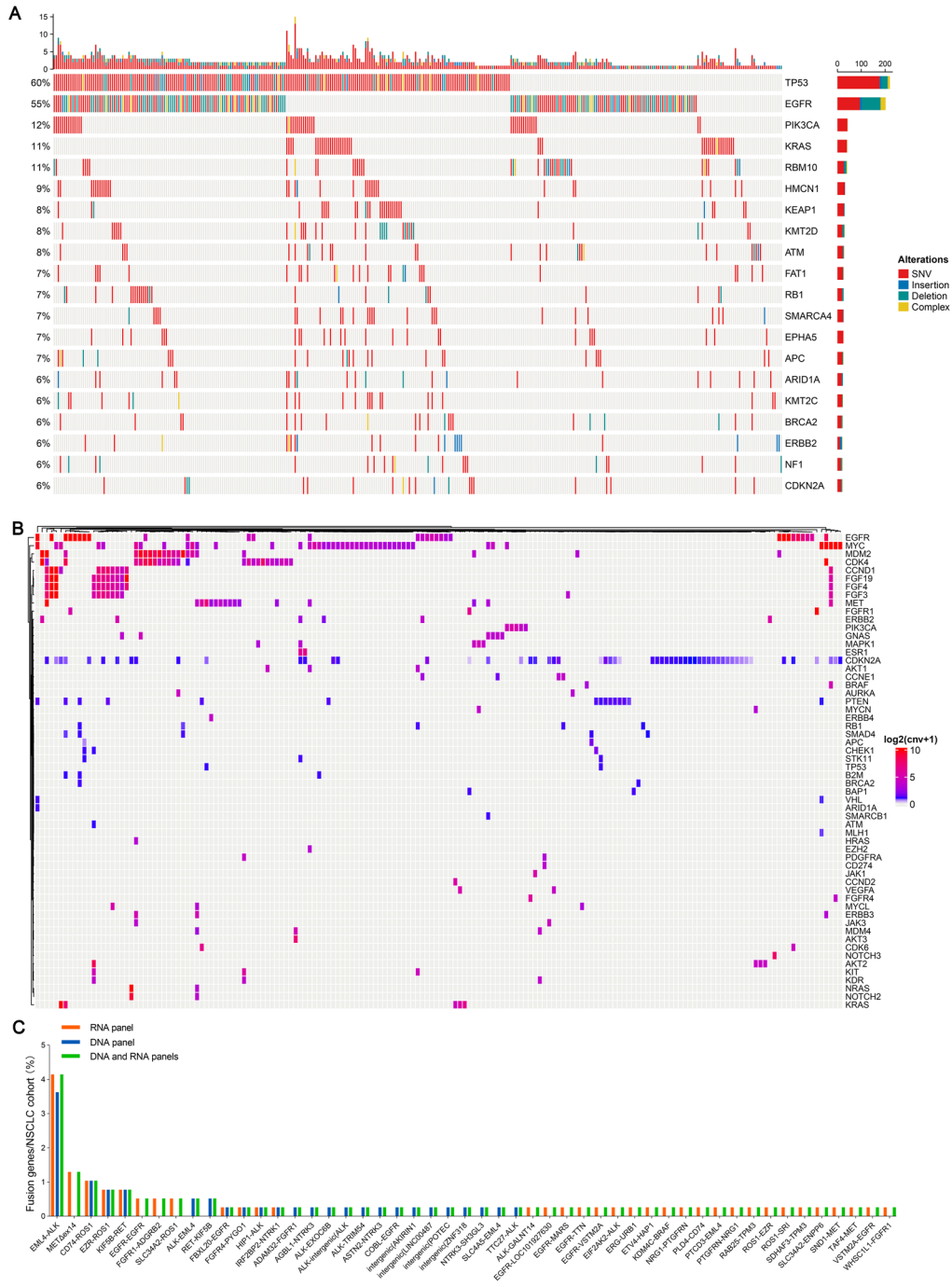
3.1 Potential well-documented target driver genes from the CSCO guideline for NSCLC using NGS

According to the NSCLC CSCO 2022 guideline, detection of SNV/indel of EGFR, KRAS, HER2 and BRAF genes, CNV of MET and HER2, gene fusions of ALK, ROS1, RET and NTRK, and MET exon 14 skipping (MET Δ ex14) would facilitate clinical application of target drugs for NSCLC patients based on these gene variations. DNA panel sequencing (769 genes) covering all aforementioned driver genes was used for the SNV/indel detection of EGFR, KRAS, HER2 and BRAF genes (Fig. 1A), and CNV of MET and HER2 genes (Fig. 1B). The number of SNV/indel of EGFR was 201 cases (52.1%), and the number of CNV of MET was 16 cases (4.1%) in the NSCLC cohort (Fig. 1A, B). Gene fusions of ALK, ROS1, RET, NTRK, and MET Δ ex14 were detected by sequencing of DNA and RNA panels (Fig. 1C).

Gene fusions were further explored using DNA panel or RNA panel sequencing in the NSCLC cohort. However, fisher's exact test showed that of the DNA and RNA panel sequencing was not significantly different in detection of ALK, ROS1, RET, NTRK, and MET Δ ex14 genes ($p > 0.05$, Fig. 1D). The number of gene fusion containing ALK, ROS1, RET, NTRK or MET Δ ex14 was comparable in the NSCLC cohort (Fig. 1E). 50% (26/52) gene fusion was simultaneously detected by both strategies and the remaining 50% was complementarily found by DNA or RNA panel sequencing (Fig. 1E). Surprisingly, detection rate of gene fusion types with ALK, ROS1, RET, NTRK or MET Δ ex14 by DNA panel sequencing were higher than that using RNA panel (Fig. 1F). These data indicated that DNA sequencing combined with RNA sequencing could more comprehensively and reliably exhibit variations of genes in the NSCLC cohort.

3.2 Gene variation landscapes in the NSCLC cohort with DNA and RNA panel sequencing

We further displayed SNV/indel landscape of top 20 genes using DNA panel sequencing in NSCLC cases, which indicated that the SNV/indel frequencies of TP53 and EGFR were 60% and 55%, respectively (Fig. 2A). CNV landscapes of all



genes in the NSCLC cases were showed in Fig. 2B. CNV rates of CDKN2A, myelocytomatosis oncogene (MYC), EGFR and cyclin-dependent kinases 4 (CDK4) genes in the NSCLC cohort were 15.9%, 12.4%, 8.2% and 7.4%, respectively. CDKN2A gene was loss of copy number, and MYC, EGFR and CDK4 genes were gain of copy number. Gene fusion landscapes was showed in Fig. 2C in the NSCLC cohort using the DNA and RNA panel sequencing. Echinoderm microtubule-associated protein-like 4 (EML4)-ALK fusion detected by NGS was most in all gene fusions and the rate was 8.2% in the NSCLC cohort. Although fusion detection rate of genes from the CSCO guideline with DNA and RNA panels was similar (Fig. 1E, F), the number and types of gene fusion detected using the RNA panel sequencing were more than those through the DNA panel sequencing (Fig. 2D, E). Thus, sequencing of the RNA panel greatly complemented deficiency of the DNA panel for gene fusion in the NSCLC cohort, which cooperatively revealed a more comprehensive gene variation feature in NSCLC.

3.3 Detection of rare gene fusion with DNA and RNA panel sequencing

By definition, fusion gene that was detected only once with DNA and/or RNA panels in the NSCLC cohort, was regarded as a rare gene fusion. As shown in Fig. 3A, there were 40 rare fusions identified in the cohort. Among them, 4 gene fusions, including F-box and leucine-rich repeat protein 20 (FBXL20)-EGFR, fibroblast growth factor receptor (FGFR4)-PYGO1, hedgehog-interacting protein 1(HIP1)-ALK and interferon regulatory factor-2-binding protein-2 (IRF2BP2)-NTRK1, were simultaneously discovered using DNA and RNA panel sequencing. 14 (35%) rare fusions were detected by the DNA panel and 22 (55%) rare fusions were detected by the RNA panel. The gene fusion with intergenic region was only detected by DNA panel, including intergenic (AKIRIN1, NDUFS5)-NTRK2, intergenic (ZNF318, ABCC10)-EZR, intergenic (POTEC, ANKRD30B)-PDGFB, intergenic (LINC00487, NRIR)-ALK and ALK-intergenic (ALK, YPEL5) (Fig. 3A). Gene fusion with driver genes such as TAF4-MET and SND1-MET detected by the RNA panel sequencing might provide a target-drug direction (Fig. 3B, C). In addition, the gene variants outside the spectrum of panels were not measured, which needs to be further assessed using whole genetic sequencing or transcriptome sequencing.

3.4 MET Δ ex14 detection using DNA and RNA sequencing

In the NSCLC cohort, we found five patients with MET Δ ex14 using DNA or RNA panel sequencing. DNA sequencing showed that 11 NSCLC cases with MET mutation and only one case was MET mutation at a c.3028G > A site, which predicted MET exon 14 skipping in mRNA processing. The MET Δ ex14 was also confirmed using RNA panel sequencing (Fig. 4A). Four cases with MET Δ ex14 were only found through RNA panel sequencing (Fig. 4B). Thus, MET Δ ex14 was easier detected using RNA sequencing compared to DNA sequencing.

3.5 Correlation of SNV/indel/CNV/fusion with clinical characteristics

We further analyzed correlation of SNV/indel/CNV/fusion with clinical characteristics in the NSCLC cohort. As shown in Fig. 5A–C, SNV/indel rates of TP53, EGFR and KRAS were significantly correlated with gender (Fig. 5A), smoking (Fig. 5B) and cancer subtype (Fig. 5C). Of course, some genes, such as EGFR were significantly different in different age (Figure S1A) and clinical stage (Figure S1B). CNV including cyclin D1 (CCND1) and FGF3/4/19 genes in male, smoking and LUSC was significantly enhanced than no-smoking and LUAD, respectively (Fig. 5D–F). CNV of CCND1 and FGF3/4/19 genes were dramatically different between stage II and III/IV patients (Fig. 5G). But there was no statistical correlation in CNV of many genes with age (Figure S1C). Fusion rates of most genes in different groups according to age (Figure S1D), gender (Figure S1E), smoking (Figure S1F), tumor subtype (Figure S1G) and clinical stage (Figure S1H) exhibited subtle difference in the NSCLC cases. Overall, SNV/indel and CNV of many genes were correlated with some clinicopathological features, such as gender, smoking, cancer subtypes and clinical stages in NSCLC patients.

4 Discussion

DNA sequencing for detection of gene fusion has proved to be challenging due to the complexed nature of genomic rearrangements and existence of a variety of intron non-coding regions [13]. Combination of RNA sequencing could comprehensively reveal gene variation features, including SNV/indel, CNV and gene fusion. In the current study, DNA and RNA panels covered all well-documented driver genes, including SNV/indel of EGFR, KRAS, HER2 and BRAF, CNV of MET and HER2, gene fusions of ALK, ROS1, RET and NTRK, and MET Δ ex14, from the CSCO. The rates of EGFR SNV/indel, MET

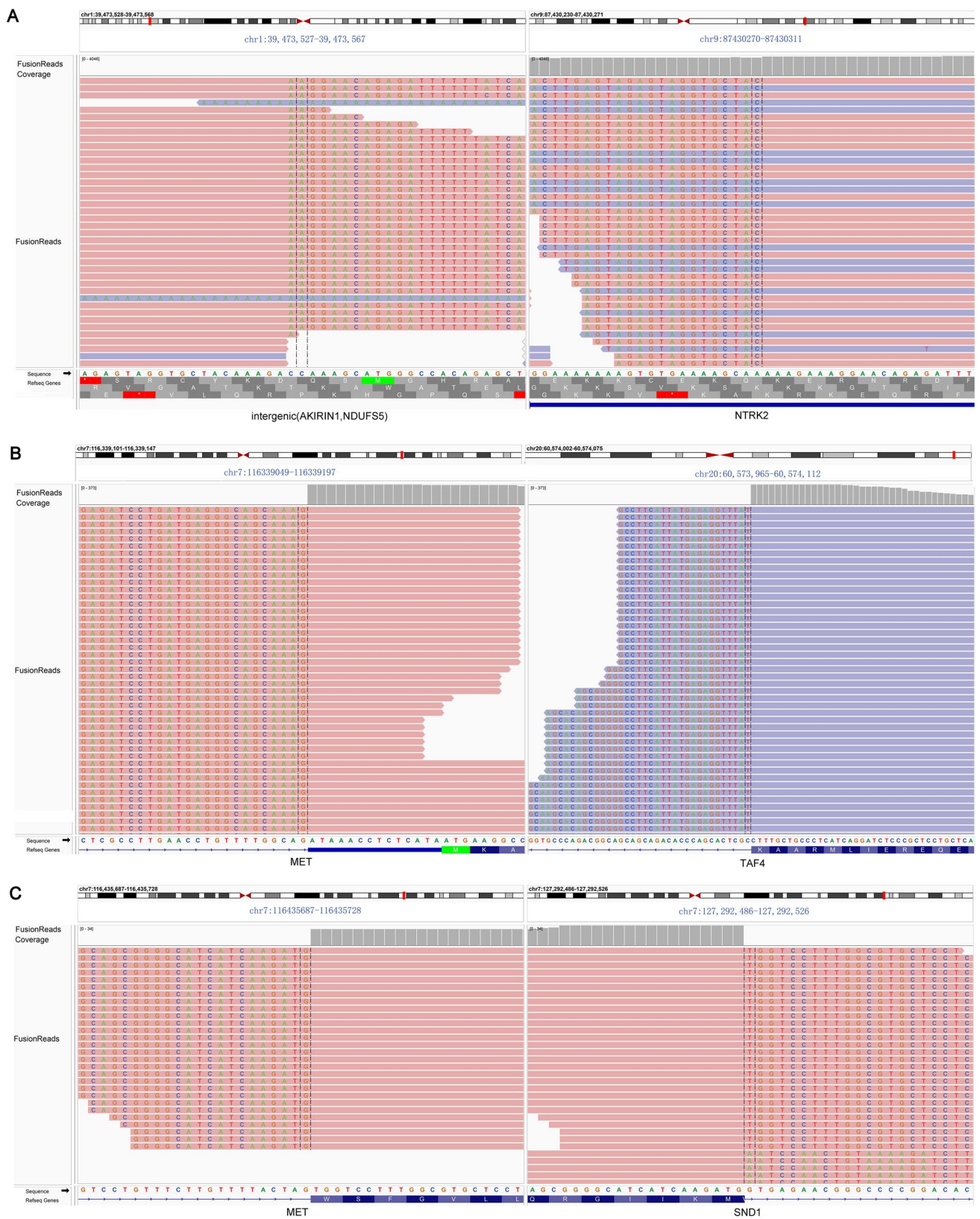


Fig. 3 Detection of rare gene fusion with sequencing of the DNA and RNA panels. **A** Intergenic(AKIRIN1,NDUF5)-NTRK2 fusion was only detected by sequencing of DNA panel. **B** TAF4-MET fusion was only detected by sequencing of RNA panel. **C** SND1-MET fusion was only detected by sequencing of RNA panel

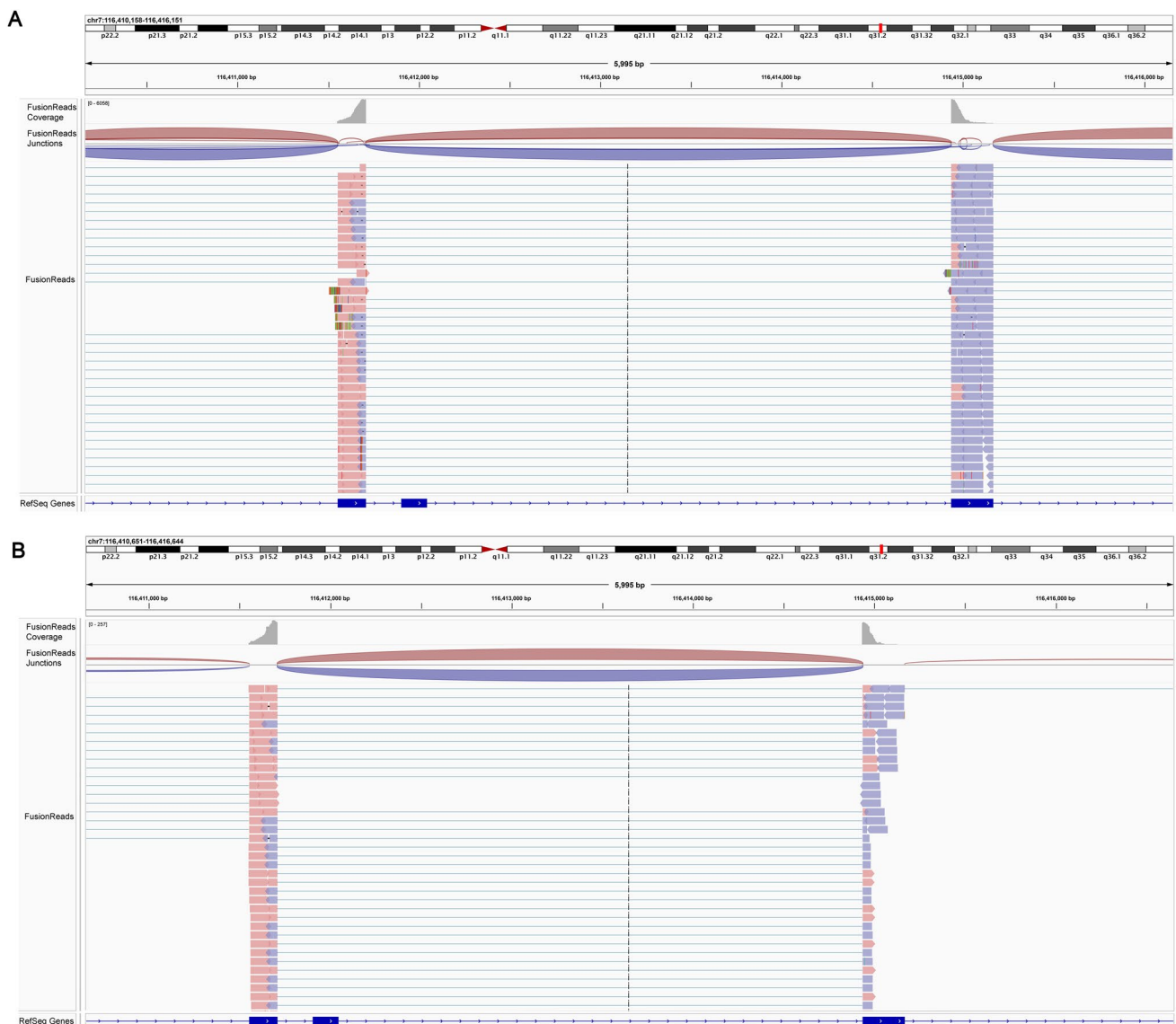
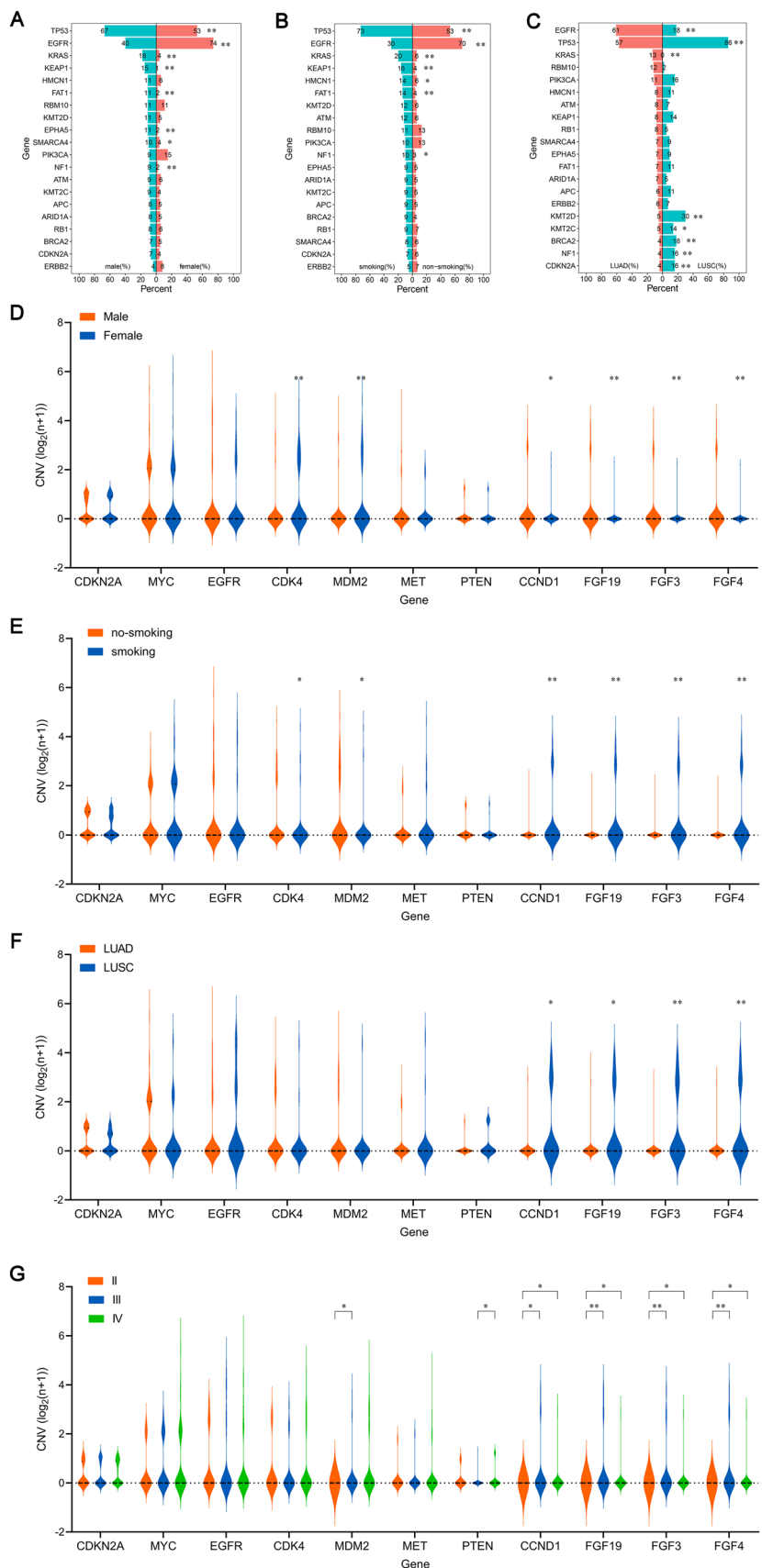


Fig. 4 MET Δ ex14 detection using DNA and RNA sequencing. **A** Sequencing of DNA panel showed MET mutation at c.694_695insA (protein: p.S232Yfs*13) and sequencing of RNA panel showed MET Δ ex14 in the patient. **B** Representative MET Δ ex14 data was only detected with sequencing of RNA panel in the patient

CNV and ALK fusion were 52.1%, 4.1% and 6.0% in 386 NSCLC patients, respectively. DNA panel sequencing accurately revealed somatic SNV/indel and CNV information of tumor genes and RNA panel sequencing was superior in identifying gene fusion compared to DNA panel sequencing. MET Δ ex14 was more easily detected by RNA sequencing, but gene fusion with intergenic region was only detected by DNA sequencing. Thus, RNA sequencing greatly complemented deficiency of DNA sequencing for gene fusion and vice versa, which together provided comprehensive and reliable gene variation features for NSCLC.

RNA sequencing used to detect gene somatic SNV/indel variation is considered to be inaccurate because of high error rate and degeneration of RNA. Only 11% of SNVs with VAF between 5 and 10% in BRAF and RAS genes was detected by whole-transcriptome sequencing compared with WES [11]. Similar findings were reported by another independent study as well [14]. Generally, FISH and IHC are two commonly used methods to detect CNVs in clinical tumor tissue samples. As expedited development of NGS technology in the past decade, DNA NGS has been extensively applied for CNV detection, although there has not been any NGS-based companion diagnostic method approved for CNV detection in NSCLC. So far, there is no consensus on the criteria for determining MET gene amplification with susceptibility or resistance to targeted

Fig. 5 Correlation of gene SNV/indel/CNV/fusion and clinical characteristics in the NSCLC cohort. **A** SNV/indel rates of top 20 genes between male and female in NSCLC cases. Statistics based on the Fisher's exact test. * $p < 0.05$, ** $p < 0.01$. **B** SNV/indel rates of top 20 genes between smoking and non-smoking patients with NSCLC. Statistics based on the Fisher's exact test. * $p < 0.05$, ** $p < 0.01$. **C** SNV/indel rates of top 20 genes in patients with LUAD or LUSC. Statistics based on the Fisher's exact test. * $p < 0.05$, ** $p < 0.01$. **D** CNV difference analysis of NSCLC cases between male and female. Statistics based on the unpaired t test. * $p < 0.05$, ** $p < 0.01$. **E** CNV difference analysis between smoking and non-smoking patients with NSCLC. Statistics based on the unpaired t test. * $p < 0.05$, ** $p < 0.01$. **F** CNV difference analysis in patients with LUAD or LUSC. Statistics based on the unpaired t test. * $p < 0.05$, ** $p < 0.01$. **G** Difference analysis of CNV in NSCLC patients with stage II, III and IV. Statistics based on the One-Way ANOVA test. * $p < 0.05$, ** $p < 0.01$



therapy [15]. In the current study, the expression of EGFR and MET genes were characterized by both parallel DNA and RNA sequencing analysis, which may increase the possibility of identifying appropriate targeted drugs for these patients.

Gene CNV are considered to be potential tumor biomarkers for NSCLC therapy. Apolipoprotein B mRNA editing catalytic polypeptide-like 3B (APOBEC3B) CNV gains were beneficial to immunotherapy response in NSCLC cases [16]. Patients with somatic CNV amplification in myeloid cell leukemia 1 (MCL1) gene exhibited unfavorable OS of NSCLC in a southern Chinese population [17]. CNV of MET and HER2 were recommended by the CSCO guideline for target therapy of NSCLC. In this study, we found that CNV including CCND1, FGF3/4/19 genes in male, smoking and LUSC cases was significantly enhanced than female, no-smoking and LUAD, respectively. Consistently, Heo and the colleagues have reported significant associations between CNV and smoking, such as acyl-CoA thioesterase 1 (ACOT1), N-terminal acetyltransferase F (NAA60), gasdermin-D (GSDMD) and HLA-DPA1 genes [18].

Overall, sequencing of DNA and RNA panels, due to features of fewer tumor samples, more detecting depth and less cost of time and material provided the foundation for wide clinical application in NSCLC. Meanwhile, combination of DNA and RNA panels revealed more gene variations, which might discover potential drug targets and benefit for patients with NSCLC.

5 Conclusion

Our findings demonstrated that combination method of DNA and RNA sequencing analysis in detecting multiple types of clinical driver mutations will facilitate the possibility of identifying appropriate targeted drugs for NSCLC patients. More solid conclusion is warranted for further validation of the clinical utility of this combined method in target therapy selection for NSCLC patients in a series of following prospective studies.

Author contributions WP, DR and ZW designed the research and supervised the study. WP, MY and JC collected clinical data and sample. LG and HZ performed experiments. WP, RL and XZ analyzed the data and designed the figures. WP and LG wrote the paper. DR and ZW revised the paper. All authors read and approved the final manuscript.

Funding This study was supported by the grants from the National Natural Science Foundation of China (82203748, 82173284), the China Postdoctoral Science Foundation Grant (2019M660206), the Science and Technology Project of Guangdong Province (2019A1515011565, 2022A1515010055), and the Science and Technology Project of Jiangmen (2020030103140008978, 2019030102430012905, 2019030102480013011).

Data availability Raw data is not publicly available to preserve individuals' privacy.

Declarations

Ethical approval and consent to participate This work was approved by the Ethics Committee of Jiangmen Central Hospital (2022–119), and informed consent was provided in accordance with the Declaration of Helsinki.

Competing interests The authors declared that no competing interest exists.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–49.
2. Sholl LM, Aisner DL, Varella-Garcia M, Berry LD, Dias-Santagata D, Wistuba II, et al. Multi-institutional oncogenic driver mutation analysis in lung adenocarcinoma: the lung cancer mutation consortium experience. *J Thorac Oncol.* 2015;10(5):768–77.

3. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature*. 2018;553(7689):446–54.
4. He J, Su C, Liang W, Xu S, Wu L, Fu X, et al. Icotinib versus chemotherapy as adjuvant treatment for stage II-IIIa EGFR-mutant non-small-cell lung cancer (EVIDENCE): a randomised, open-label, phase 3 trial. *Lancet Respir Med*. 2021;9(9):1021–9.
5. Wu YL, Tsuboi M, He J, John T, Grohe C, Majem M, et al. Osimertinib in resected EGFR-mutated non-small-cell lung cancer. *N Engl J Med*. 2020;383(18):1711–23.
6. Zhong WZ, Wang Q, Mao WM, Xu ST, Wu L. Gefitinib versus vinorelbine plus cisplatin as adjuvant treatment for stage II-IIIa (N1–N2) EGFR-mutant NSCLC: final overall survival analysis of CTONG1104 phase III trial. *J Clin Oncol*. 2021;39(7):713–22.
7. Zhou C, Kim SW, Reungwetwattana T, Zhou J, Zhang Y, He J, et al. Alectinib versus crizotinib in untreated Asian patients with anaplastic lymphoma kinase-positive non-small-cell lung cancer (ALESIA): a randomised phase 3 study. *Lancet Respir Med*. 2019;7(5):437–46.
8. Soria JC, Tan DSW, Chiari R, Wu YL, Paz-Ares L, Wolf J, et al. First-line ceritinib versus platinum-based chemotherapy in advanced ALK-rearranged non-small-cell lung cancer (ASCEND-4): a randomised, open-label, phase 3 study. *Lancet*. 2017;389(10072):917–29.
9. Wolf J, Seto T, Han JY, Reguart N, Garon EB, Groen HJM, et al. Capmatinib in MET Exon 14-mutated or MET-amplified non-small-cell lung cancer. *N Engl J Med*. 2020;383(10):944–57.
10. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. Non-small cell lung cancer, version 3.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2022;20(5):497–530.
11. Kaya C, Dorsaint P, Mercurio S, Campbell AM, Eng KW, Nikiforova MN, et al. Limitations of detecting genetic variants from the RNA sequencing data in tissue and fine-needle aspiration samples. *Thyroid*. 2021;31(4):589–95.
12. Xu E, Su K, Zhou Y, Gong L, Xuan Y, Liao M, et al. Comprehensive landscape and interference of clonal haematopoiesis mutations for liquid biopsy: a Chinese pan-cancer cohort. *J Cell Mol Med*. 2021;25(21):10279–90.
13. Li W, Guo L, Liu Y, Dong L, Yang L, Chen L, et al. Potential unreliability of uncommon ALK, ROS1, and RET genomic breakpoints in predicting the efficacy of targeted therapy in NSCLC. *J Thorac Oncol*. 2021;16(3):404–18.
14. O'Brien TD, Jia P, Xia J, Saxena U, Jin H, Vuong H, et al. Inconsistency and features of single nucleotide variants detected in whole exome sequencing versus transcriptome sequencing: a case study in lung cancer. *Methods*. 2015;83:118–27.
15. Lee HJ, Xu X, Choe G, Chung DH, Seo JW, Lee JH, et al. Protein overexpression and gene amplification of epidermal growth factor receptor in nonsmall cell lung carcinomas: comparison of four commercially available antibodies by immunohistochemistry and fluorescence in situ hybridization study. *Lung Cancer*. 2010;68(3):375–82.
16. Wang S, Jia M, He Z, Liu X-S. APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. *Oncogene*. 2018;37(29):3924–36.
17. Yin J, Li Y, Zhao H, Qin Q, Li X, Huang J, et al. Copy-number variation of MCL1 predicts overall survival of non-small-cell lung cancer in a Southern Chinese population. *Cancer Med*. 2016;5(9):2171–9.
18. Heo Y, Heo J, Han S-s, Kim WJ, Cheong HS, Hong Y. Difference of copy number variation in blood of patients with lung cancer. *Int J Biol Markers*. 2020;36(1):3–9.

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