scientific reports

Check for updates

OPEN

Combined utility of genomic breakpoints and frame is a reliable predictor of ALK transcript function

Qiong Yang[1,4], Yumin Guo[3,4], Weihui Guo[3], Dongdong Liu[2], Haitao Wang[2]✉ & Xufan Cai[2]✉

Lung cancer is a major cause of cancer-related deaths globally. Targeted therapies, specifically attacking cancer cells based on genetic mutations, offer promising alternatives. ALK (anaplastic lymphoma kinase) fusions result in aberrant proteins that drive cancer growth. Drugs like crizotinib and ceritinib have shown efficacy in treating ALK-positive NSCLC. Accurate detection of ALK fusions is crucial for guiding these therapies. We conducted a retrospective analysis of a Chinese cohort of 131 ALK rearrangement-positive patients detected by DNA NGS between January 2017 and December 2021. Among those 131 ALK fusions, RNA-NGS confirmed positive transcripts in 88% of canonical ALK fusions and 75% of ALK fusions with rare partners in samples sequenced by both DNA NGS and RNA NGS. The secondary classification approach increased transcript prediction accuracy to 95.4% when combining common breakpoints and inframe fusion analysis in canonical ALK fusions. Combining rare breakpoints and inframe fusion could increase transcript prediction accuracy to 100%. For ALK fusions with rare partners, combining common breakpoints and frameshift improved transcript prediction accuracy to 100%. Additionally, combining rare breakpoints with inframe or frameshift could enhance the prediction accuracy to 100%. Combining DNA NGS and RNA NGS with a secondary classification approach significantly enhances the transcript prediction accuracy at the RNA level. This method optimizes clinical diagnostic and therapeutic strategies for ALK-positive NSCLC, highlighting the importance of advanced sequencing techniques in precision oncology.

Keywords ALK, Breakpoint, Frame, NSCLC, NGS

Lung cancer remains a leading cause of cancer-related mortality worldwide, primarily due to late-stage diagnosis and the limitations of current treatment strategies[1]. Conventional approaches such as surgery, chemotherapy, and radiotherapy often fail to achieve long-term remission and are associated with substantial side effects[2–4]. hese limitations have spurred the pursuit of more effective, targeted treatment options. Targeted therapy, which utilizes drugs designed to specifically attack cancer cells while sparing normal cells, has emerged as a promising and personalized approach. This treatment modality focuses on specific genetic mutations and alterations driving cancer progression, offering improved outcomes for lung cancer patients[5,6].

Among these genetic alterations, ALK (anaplastic lymphoma kinase) fusions play a pivotal role in non-small cell lung cancer (NSCLC)[7,8]. ALK fusions result from the rearrangement and fusion of the ALK gene with another gene, producing an abnormal ALK protein that drives cancer cell proliferation. This discovery has led to the development of several targeted therapies, including ALK inhibitors such as crizotinib, ceritinib, alectinib, brigatinib, and lorlatinib. These drugs have significantly improved outcomes for patients with ALK-positive NSCLC by providing better response rates and extended progression-free survival compared to conventional therapies[9,10].

Accurate detection of ALK fusions is essential for guiding targeted therapy, as patients harboring these fusions benefit significantly from ALK inhibitors. Several methods are available for detecting ALK fusions, including fluorescence in situ hybridization (FISH), immunohistochemistry (IHC), reverse transcription-polymerase chain reaction (RT-PCR), and next-generation sequencing (NGS)[11,12]. Of these, DNA NGS has gained prominence due to its high sensitivity and ability to simultaneously detect a broad spectrum of genetic alterations. DNA NGS is particularly advantageous as it requires only small biopsy samples and can identify both

[1]General Surgery, Cancer Center, Department of Lymphatic surgery and reconstructive microsurgery, Zhejiang Provincial People's Hospital Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China. [2]Cancer Center, Department of Thoracic Surgery, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China. [3]School of Basic Medical Sciences and Forensic Medicine, Hangzhou Medical College, Hangzhou, Zhejiang, China. [4]Qiong Yang and Yumin Guo contributed equally to this work. ✉email: 443292584@qq.com; caixvfan123@163.com

known and novel fusions. However, the complexity of ALK fusion breakpoints presents a diagnostic challenge, as breakpoints can occur in both intronic and exonic regions[13–15]. Breakpoints within intronic regions often result in intronic-breakpoint fusions that generate in-frame chimeric fusion transcripts or proteins due to preserved coding sequences. Conversely, exonic breakpoints complicate the prediction of RNA-level fusion expression, making it challenging to determine whether these fusions will be expressed at the RNA level[15]. Thus, accurately predicting RNA-level fusion expression based solely on DNA breakpoints remains an unmet challenge in the field.

To address this issue, our study analyzed samples that underwent both DNA NGS and RNA NGS sequential sequencing. By employing a secondary classification approach, we identified factors associated to transcript prediction accuracy at the RNA level. This innovative screening method not only enhances the precision of ALK fusion detection but also optimizes clinical diagnostic and therapeutic pathways, paving the way for improved personalized treatment strategies in ALK-positive NSCLC.

Methods and materials

Patients and samples

Data of a Chinese cohort from multi-center hospitals involving 131 patients who were detected ALK rearrangement positive by DNA NGS from January 2017 to December 2021 were retrospectively analyzed. This study was approved by the ethical committee under protocol QT2024193. The study was conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all subjects involved in the study.

DNA NGS

Genomic DNA was isolated from FFPE (Formalin-Fixed, Paraffin-Embedded) samples and detected using the previously disclosed targeted sequencing panel, which consists of all coding exons from 825 genes relevant to cancer and select introns from 44 commonly rearrangement genes. TRIMMOMATIC (version 0.36) was used for quality control on the raw sequencing data to get rid of adapters and poor-quality sections. Using the BURROWS-WHEELER ALIGNER tool (version 0.7.10), read local alignments to the hg19 genome (GRch37) were performed. STRELKA (https://github.com/Illumina/strelka), MUTECT (https://software.broadinstitute.org/cancer/cga/mutect), and GENEFUSE version 0.6.1 (https://github.com/OpenGene/GeneFuse) were used to retrieve somatic single nucleotide variants, somatic insertions and deletions, and structural variations, respectively. Based on recommendations from the Exome Aggregation Consortium, the variations were filtered and eliminated if their population frequency was more than 0.1%. The remaining variations were annotated by Oncotator and Vep.

RNA NGS

RNA was extracted using an AllPrep DNA/RNA FFPE Kit (Qiagen, Hilden, Germany) from FFPE samples. Using a Qubit RNA HS assay (Thermo Fisher Scientific, Waltham, MA), the amount and quality of extracted RNA were quantified. Following strand-specific cDNA synthesis with SuperScript III Reverse Transcriptase (Thermo Fisher Scientific, Waltham, MA, USA), cDNA libraries were constructed and hybridized with capture probe baits from the Fusioncapture panel, which contained the full transcripts of 395 cancer-related common fusion genes. The prepared libraries were subjected to paired-end sequencing on the Illumina NovaSeq 6000 system (San Diego, California). Using Hisat2-2.0.5, sequencing reads were mapped to the human reference genome (hg19). FusionMap was used to recognize gene fusions. At least four unique pairs of supporting readings spanning the breakpoints between the two partners were required to call a fusion.

Statistical analysis

Categorical variables were compared using the chi-square test. Analyses and data presentation were undertaken using R (version 4.2.0).

Results

Screening for ALK fusions using DNA NGS and RNA NGS

Based on the partner type, the ALK fusions were divided into canonical ALK fusions with EML4 partner and rare ALK fusions with other partners (Fig. 1). Moreover, samples with enough specimen were further sequenced by RNA NGS. Totally, 99 canonical ALK fusions and 32 rare ALK fusions were successfully sequenced by RNA NGS. Among the canonical ALK fusions, the positive transcripts were identified in 87 samples (88%). Similarly, 75% samples in rare ALK fusions present functional transcripts (24/32).

The characteristics of ALK fusions

Totally, 30 partners were identified in those samples successfully sequenced by both DNA NGS and RNA NGS. Numerous chromosomes were found to be involved in the ALK fusions, including chr 1–5, chr 7–8, chr 10–11, chr 14, chr 17 and chr 19 (Fig. 2A). Among those chromosomes, chr 2 was found to be the most frequent chromosome involved in the ALK fusions and 82% of partners were identified, including EML4, DCTN, RTN4, PLB1, STRN, SFTPB, RANBP2, CYP27C1 and RGPD1. Those results demonstrated that ALK driver and partners tended to form intrachromosomal rearrangements rather than interchromosomal fusions (Fig. 2B).

Furthermore, a distinct distribution of breakpoints of ALK driver was observed between canonical and rare ALK fusions. In canonical ALK fusions, both intronic breakpoints ($n=92$) and exonic breakpoints ($n=7$) were identified. Notably, the majority of breakpoints located in intron 19, accounting for 86% of the breakpoints in this group. Additionally, sporadic breakpoints were identified in introns 16–18 and 20–21. Among the exonic

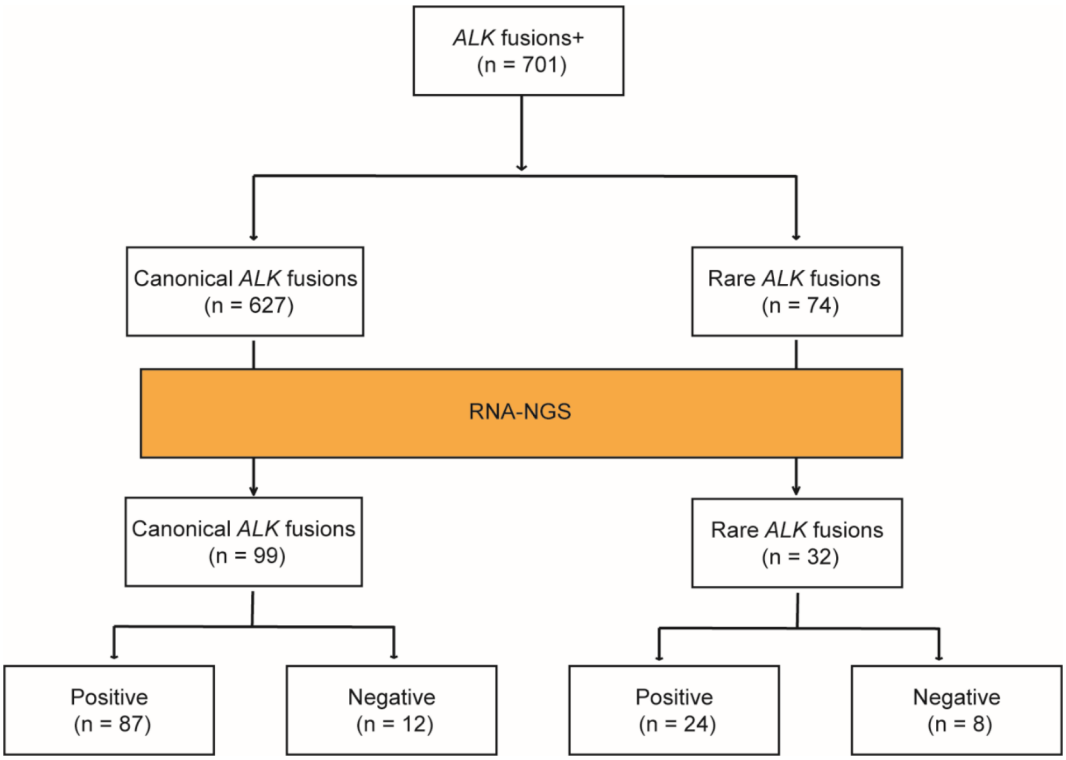

Fig. 1. The ALK fusion screening pipeline by DNA NGS and RNA NGS.

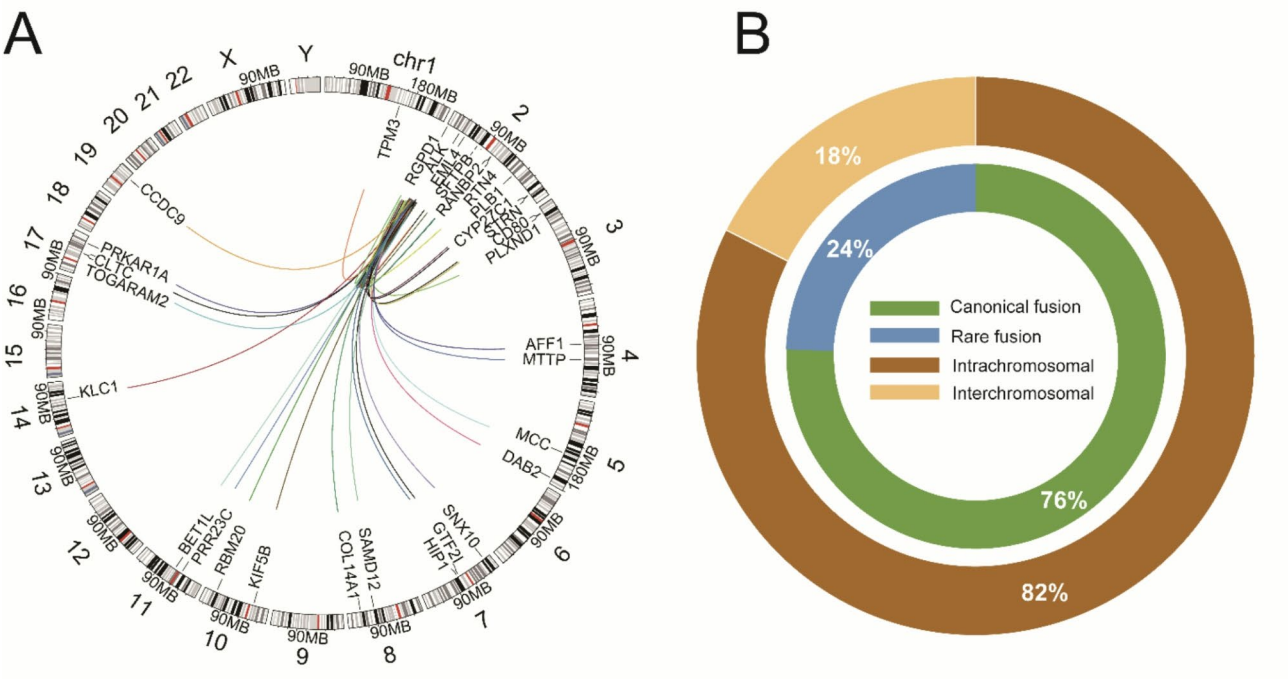

Fig. 2. The ALK fusion characteristics in chromosome distribution. (**A**) Circos plot graphically depicting ALK and partners. (**B**) The interchromosomal and intrachromosomal rearrangements ration.

breakpoints, seven were found, specifically distributed between exons 19 and 20 (Fig. 3A, B). In contrast, ALK breakpoints in rare ALK fusions exhibited a different distribution pattern. While intron 19 remained the most frequent location for breakpoints ($n = 19$, 23, 72%), unique breakpoints were identified exclusively in introns 1–3, intron 6, and intron 9, which were absent in canonical ALK fusions (Fig. 3A, C).

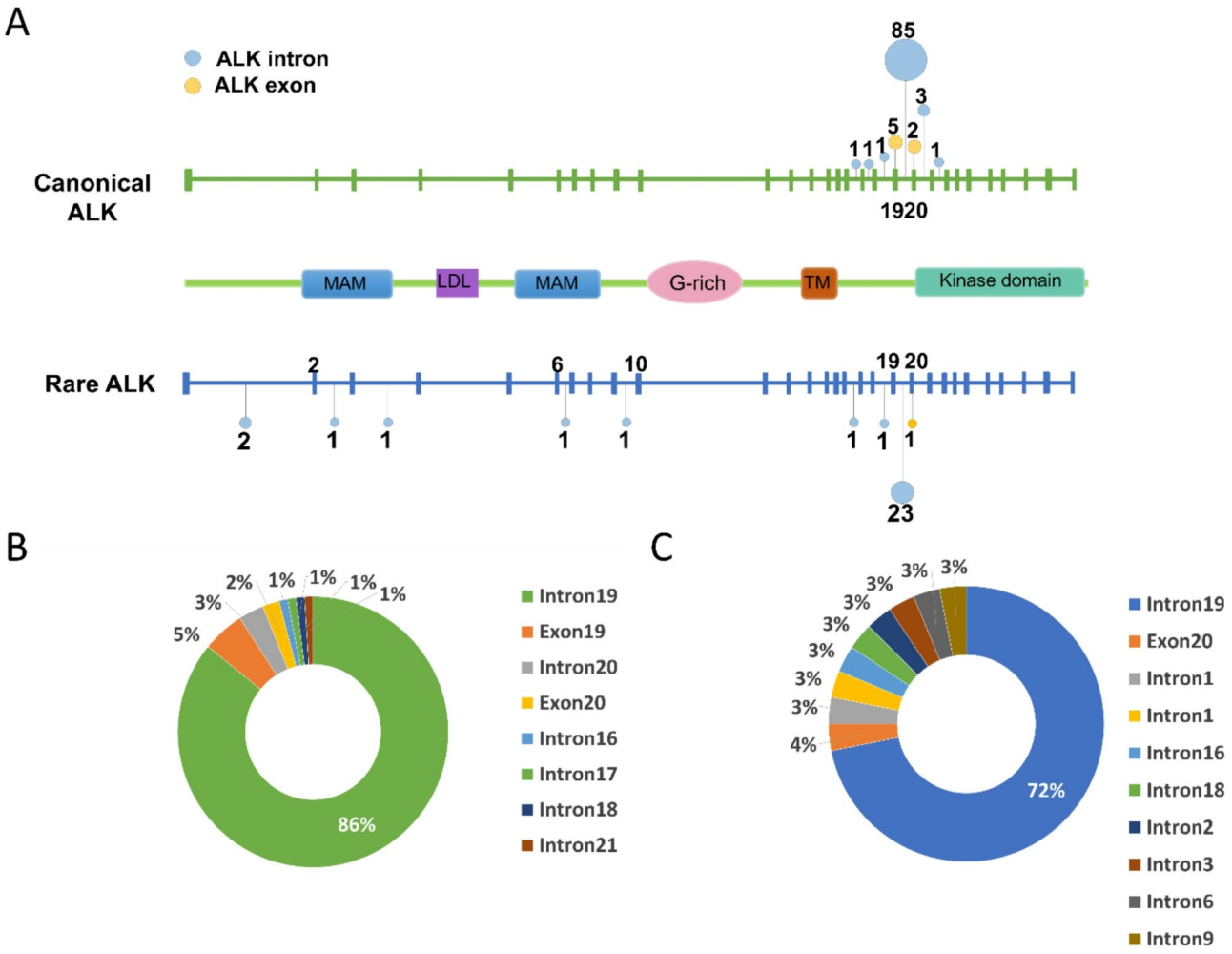

Fig. 3. Distribution pattern of breakpoints in ROS1 rearrangements. (**A**) The breakpoints pattern in ALK fusions. (**B**) the breakpoints proportion in canonical ALK fusion. (**C**) the breakpoints proportion in rare ALK fusion.

Secondary classification approach could increase the transcript prediction accuracy in canonical ALK fusions

As mentioned above, the most frequent breakpoint site of ALK driver was intron 19, which was defined as the common breakpoint, while all other sites were categorized as rare breakpoints, forming the 1st tier classification. Additionally, fusions were further classified into the 2nd tier based on specific characteristics. 1. *sole reciprocal fusion* where ALK is identified only in the 5' position (5' ALK-EML4 3'); 2. *inframe fusion* where breakpoints of ALK driver occur in introns and maintain the reading frame; 3. *frameshift fusion* where intronic breakpoints disrupt the reading frame; 4. *exonic fusion* where the breakpoints locate in exons. When analyzing fusions with common breakpoints, the transcript prediction accuracy increased from 88 to 90% using only the 1st tier classification (Fig. 4A, C). This rate further improved to 95.4% when incorporating both the common breakpoint and inframe characteristics into the analysis. Notably, the transcript prediction accuracy reached 100% when combining the common breakpoint and exonic breakpoint of partner. In contrast, the transcript prediction accuracy decreased when considering both the common breakpoint and frameshift fusions or sole reciprocal type (Fig. 4A, C).

For canonical ALK fusions with rare breakpoints of ALK driver, the transcript prediction accuracy was reduced to 76% when solely including rare breakpoints of ALK driver in the analysis. However, transcript prediction accuracy increased to 100% when combining rare breakpoints of ALK driver with inframe characteristics, and to 90% when combining rare breakpoints with exonic breakpoints of ALK driver (Fig. 4B, C). Conversely, the transcript prediction accuracy reached 100% when combining rare breakpoints of ALK driver with frameshift type. Unfortunately, the combination of rare breakpoints of ALK driver and sole reciprocal type did not improve the transcript prediction accuracy (Fig. 4B, C).

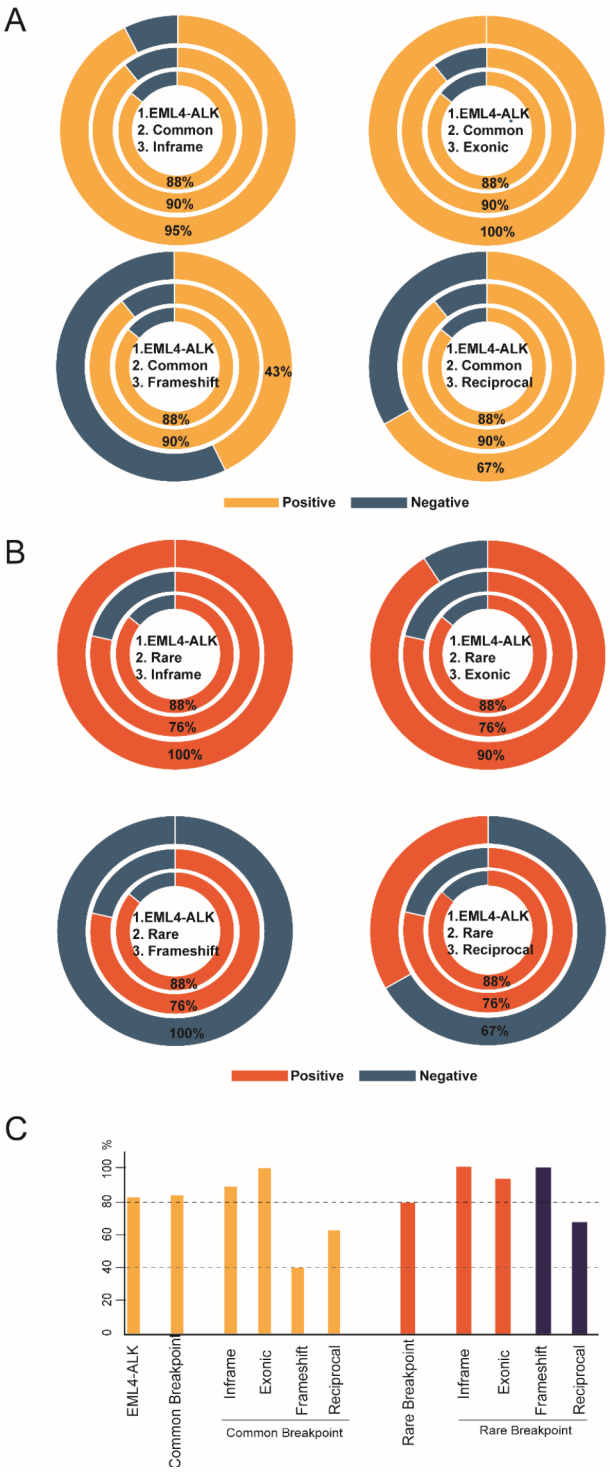

Fig. 4. The transcript prediction accuracy in canonical ALK fusions. (**A**) The transcript prediction accuracy in canonical ALK fusions with common breakpoints, which were further classified by inframe, exonic, frameshift and sole reciprocal structure. (**B**) The transcript prediction accuracy in canonical ALK fusions with rare breakpoints, which were further classified by inframe, exonic, frameshift and sole reciprocal structure. (**C**) The columns represented the positive or negative transcript prediction accuracy. The series numbers displayed in the center of the circle represent the rings: Number 1 corresponds to the inner ring, Number 2 to the middle ring, and Number 3 to the outer ring.

Secondary classification approach could increase the transcript prediction accuracy in rare ALK fusions

The secondary classification approach demonstrated significant improvements in transcript prediction accuracy for rare ALK fusions compared to canonical ALK fusions. Using the common breakpoint of ALK driver classification, the transcript prediction accuracy for rare ALK fusions increased from 75 to 95%. Moreover, combining the common breakpoint of ALK driver with either frameshift or sole reciprocal type further increased the accuracy to 100%. However, including the common breakpoint of ALK driver and inframe characteristics in the analysis slightly decreased the transcript prediction accuracy to 91% (Fig. 5A, C).

When analyzing rare breakpoints of ALK driver alone, the transcript prediction accuracy was only 45% (Fig. 5B, C). A slight improvement was observed when combining rare breakpoints of ALK driver with exonic breakpoints, increasing the accuracy to 50%. However, the accuracy increased dramatically to 100% when including both rare breakpoints of ALK driver and inframe characteristics. Similarly, the transcript prediction accuracy reached 100% when combining rare breakpoints of ALK driver with frameshift type (Fig. 5B, C).

Proposed algorithm for ALK fusion detection in cancers

For optimal detection of ALK fusions, we recommend a testing strategy that combines both a screening and confirmation assay (Fig. 6). DNA NGS, if available, should serve as the primary screening tool due to its broad sensitivity for all ALK fusions. It is essential to ensure coverage of all relevant ALK exons and introns where breakpoints are commonly located, as previously reported. When common breakpoints of ALK driver are detected, whether in canonical or rare ALK fusions, the findings are typically actionable without further analysis. However, when rare breakpoints of ALK driver are identified, a secondary classification approach should be employed to guide subsequent testing and treatment decisions. In the canonical ALK fusions with rare breakpoints, targeted treatment could be proceed if inframe type or exonic breakpoints were detected. Similarly, frameshift can be considered negative results. However, sole reciprocal ALK fusions require further evaluation via RNA sequencing to confirm the presence of an ALK fusion transcript. For rare ALK fusions with rare breakpoints of ALK driver, fusions with inframe can be actionable without additional testing. Similarly, fusions with frameshift indicate negative results. However, fusions with exonic breakpoints need to be further tested by RNA sequencing. An alternative approach could involve performing DNA and RNA sequencing concurrently. While this dual upfront testing may streamline the process, its superiority over sequential testing remains unclear and warrants further evaluation.

Discussion

Lung cancer, particularly non-small cell lung cancer (NSCLC), remains a leading cause of cancer-related deaths worldwide[1]. The emergence of ALK-targeted therapies, such as crizotinib, ceritinib, alectinib, brigatinib, and lorlatinib, has significantly improved patient outcomes by specifically inhibiting the aberrant ALK protein, thereby offering better response rates and prolonged progression-free survival compared to conventional treatments[6,10]. Our study highlights the efficacy of combining DNA NGS and RNA NGS in identifying ALK fusions and accurately predicting their expression at the RNA level. By employing a secondary classification approach, we demonstrated that this method enhances the accuracy of ALK fusion detection, ultimately optimizing clinical diagnostic and therapeutic pathways.

Detecting ALK fusions is critical due to the high incidence of lung cancer, which remains one of the most frequently diagnosed and deadly cancers globally[16]. ALK fusions occur in approximately 3–7% of NSCLC cases, making them a significant subset for targeted therapy[17]. The introduction of ALK inhibitors, such as crizotinib, has led to substantial improvements in patient outcomes, including increased overall survival and progression-free survival. For instance, crizotinib has demonstrated a progression-free survival of 10.9 months compared to 7.0 months with standard chemotherapy[18]. Furthermore, when resistance develops to initial ALK inhibitors, second-generation inhibitors like ceritinib and alectinib have shown efficacy, offering continued therapeutic benefits[19,20]. Research also suggests that targeting other fusion genes can sometimes lead to the emergence of secondary ALK fusions, which are also treatable with ALK inhibitors[21,22]. This adaptability underscores the necessity for robust ALK fusion detection to guide treatment decisions accurately. Comprehensive detection of ALK fusions ensures that patients can benefit from the full spectrum of ALK-targeted therapies, improving clinical outcomes significantly[23]. Thus, incorporating advanced detection techniques, such as DNA NGS and RNA NGS, is essential for optimizing the management and treatment of NSCLC patients.

ALK fusion detection methods such as fluorescence in situ hybridization (FISH) and immunohistochemistry (IHC) have significant limitations. These methods are limited by low throughput, reliance on manual interpretation, and a high susceptibility to errors. Additionally, the IHC could not identify the specific partner genes which can affect the prognosis of ALK-positive patients. For instance, EML4-ALK fusions are associated with better outcomes compared to other fusion partners[8,24,25]. Therefore, accurately identifying the fusion partner is crucial for subsequent clinical treatment and decision-making. DNA NGS offers several advantages over traditional methods. It provides higher sensitivity and the ability to detect a wide range of genetic alterations simultaneously. In our study, we systematically analyzed ALK fusions using both DNA-NGS and RNA NGS. By using RNA NGS as a reference, we explored the potential of predicting RNA-level expression based on DNA NGS results. This approach not only enhances the accuracy of ALK fusion detection but also provides critical insights into the expression of these fusions at the RNA level, which is essential for guiding targeted therapy.

Previous studies and our results both indicate the complexity of ALK fusions[13,26]. For example, various partners were identified which significantly influence patient prognosis[27,28]. Additionally, the distribution of breakpoints of ALK driver spans different introns and even exons. In our study, ALK fusions were categorized into canonical ALK fusions (with EML4 as the partner) and rare ALK fusions (with other partners). The intrachromosomal rearrangements dominate ALK fusions, likely due to the close spatial proximity of the driver and partner genes,

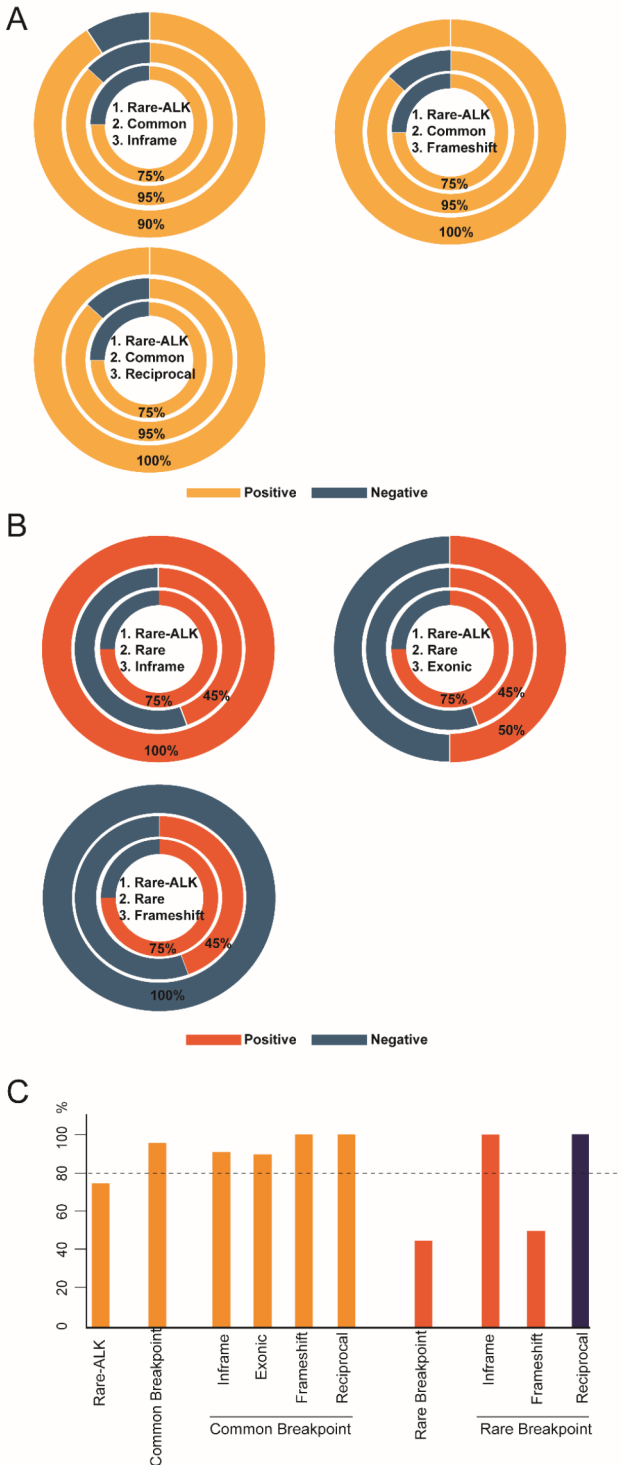

Fig. 5. The transcript prediction accuracy in rare ALK fusions. (**A**) The transcript prediction accuracy in canonical ALK fusions with common breakpoints, which were further classified by inframe, exonic, frameshift and sole reciprocal structure. (**B**) The transcript prediction accuracy in canonical ALK fusions with rare breakpoints, which were further classified by inframe, exonic and frameshift structure. (**C**) The columns represented the positive or negative transcript prediction accuracy. The series numbers displayed in the center of the circle represent the rings: Number 1 corresponds to the inner ring, Number 2 to the middle ring, and Number 3 to the outer ring.

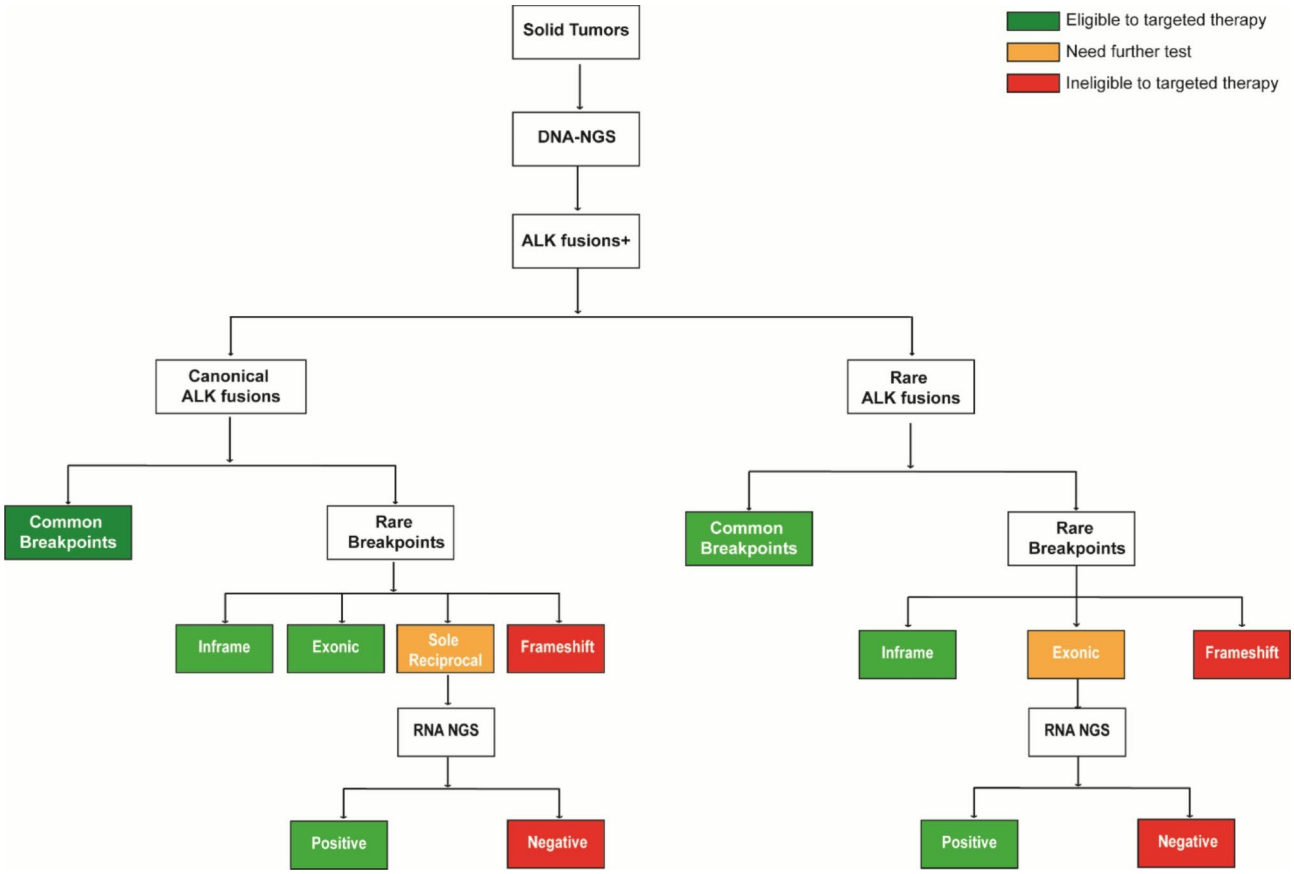

Fig. 6. Proposed algorithm for ALK fusion detection in cancers.

making fusion formation more feasible. This finding contrasts with ROS1 fusions, suggesting that each fusion driver has unique characteristics in fusion formation[29]. Additionally, in both canonical and rare ALK fusions, intron 19 of ALK emerged as the most common breakpoints locate, termed the common breakpoints. This was made possible by the comprehensive coverage of exons and most introns in the hybrid capture DNA NGS panels used in our study. Beyond identifying breakpoints near the kinase domain, breakpoints of ALK driver within the kinase domain (intron 20–21) and in distal introns (intron 1–3) were found in our study. This highlights the importance of using a well-covered panel to identify novel fusions, providing more precise guidance for future therapeutic strategies.

Building on our classification of ALK fusions into canonical and rare types based on their fusion partners, each group was further divided by the common and rare breakpoints of ALK driver. In both canonical and rare ALK fusions, those with common breakpoints of ALK driver exhibited a probability of over 90% for having functional RNA transcripts. This high transcript prediction accuracy underscores the effectiveness and clinical relevance of classifying ALK fusions based on breakpoints location. However, for the ALK driver rare breakpoints subgroup, transcript prediction accuracy was less reliable, indicating that additional factors might influence the expression of these fusions and that more refined methods may be necessary for increase transcript prediction accuracy of those cases.

To address the variability in rare breakpoints of ALK driver, a secondary classification approach was implemented. Initially, ALK drivers were categorized based on their position within the fusion (5' or 3'), and sole reciprocal ALK fusions (5' ALK-X 3') were identified. While previous studies have distinguished between 3' ALK fusions and nonreciprocal/reciprocal ALK fusions. However, sole reciprocal ALK fusions have only been sporadically reported and remain underexplored. In this study, 5' ALK fusion alone was defined as a sole reciprocal fusion, where transcript prediction accuracy is particularly challenging because the ALK driver locates at the 5' end. Subsequently, ALK fusions with exonic breakpoints were classified as the exonic group, while those with intronic breakpoints were further divided into inframe and frameshift groups based on predictive modeling. As expected, both canonical and rare fusions with common breakpoints of ALK driver in the inframe group demonstrated a 100% probability of RNA-level expression. Conversely, all frameshift samples yielded 100% negative results at the RNA level. The subgroup with exonic breakpoint exhibited distinct behaviors between canonical and rare fusions. For canonical ALK fusions, 90% of cases with exonic breakpoints produced transcripts, likely due to exon removal during transcription, resulting in inframe fusions. However, this trend was not observed in rare fusions. It is hypothesized that RNA-level expression of fusions with exonic breakpoints

is significantly influenced by the type of fusion partner involved, highlighting the complexity of rare ALK fusion dynamics.

This study has limitations. First, the retrospective nature of the analysis might introduce selection bias. Second, while our approach enhances prediction accuracy, further prospective studies are needed to validate our findings in diverse clinical settings. Additionally, the influence of different fusion partners on ALK inhibitor response requires more comprehensive exploration to optimize personalized treatment strategies. Finally, the use of hg19 as the reference genome, though appropriate for the study period, is a limitation. Future studies should adopt hg38 to leverage its advantages, including better genome annotations and improved representation of structural variations, to ensure consistency with modern genomic research and clinical applications.

Conclusion

Our study demonstrates that combining DNA NGS and RNA NGS enhances the accuracy of ALK fusion detection and transcript prediction accuracy. This comprehensive approach optimizes clinical diagnostics and therapeutic strategies for ALK fusion-positive cancer, highlighting the importance of advanced sequencing techniques in precision oncology.

Data availability

The datasets generated and/or analyzed during this study are available in the NCBI BioProject repository: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1206772.

Received: 24 October 2024; Accepted: 28 February 2025
Published online: 11 March 2025

References

1. Kratzer, T. B. et al. Lung cancer statistics, 2023. *Cancer* **130**(8), 1330–1348 (2024).
2. Vinod, S. K. & Hau, E. Radiotherapy treatment for lung cancer: Current status and future directions. *Respirology* **25**, 61–71 (2020).
3. De Ruysscher, D. et al. Recommendation for supportive care in patients receiving concurrent chemotherapy and radiotherapy for lung cancer. *Ann. Oncol.* **31**(1), 41–49 (2020).
4. Bryan, D. S. & Donington, J. S. The role of surgery in management of locally advanced non-small cell lung cancer. *Curr. Treat. Options Oncol.* **20**(4), 27 (2019).
5. Zhong, L. et al. Small molecules in targeted cancer therapy: Advances, challenges, and future perspectives. *Signal. Transduct. Target. Therapy* **6**(1), 1–48 (2021).
6. Majeed, U., Manochakian, R., Zhao, Y. & Lou, Y. Targeted therapy in advanced non-small cell lung cancer: Current advances and future trends. *J. Hematol. Oncol.* **14**, 1–20 (2021).
7. Bruno, R. & Fontanini, G. Next generation sequencing for gene fusion analysis in lung cancer: A literature review. *Diagnostics* **10**(8), 521 (2020).
8. Xiang, Y. et al. Therapeutic advances of rare ALK fusions in non-small cell lung cancer. *Curr. Oncol.* **29**(10), 7816–7831 (2022).
9. Shaw, A. T. et al. First-line lorlatinib or Crizotinib in advanced ALK-positive lung cancer. *N. Engl. J. Med.* **383**(21), 2018–2029 (2020).
10. Ou, S-H-I. et al. Efficacy of brigatinib in patients with advanced ALK-positive NSCLC who progressed on alectinib or ceritinib: ALK in lung cancer trial of brigAtinib-2 (ALTA-2). *J. Thorac. Oncol.* **17**(12), 1404–1414 (2022).
11. Letovanec, I. et al. Evaluation of NGS and RT-PCR methods for ALK rearrangement in European NSCLC patients: Results from the European thoracic oncology platform lungscape project. *J. Thorac. Oncol.* **13**(3), 413–425 (2018).
12. Batra, U. et al. IHC versus FISH versus NGS to detect ALK gene rearrangement in NSCLC: All questions answered? *J. Clin. Pathol.* **75**(6), 405–409 (2022).
13. Wang, Z. et al. Molecular characterization of genomic breakpoints of ALK rearrangements in non-small cell lung cancer. *Mol. Oncol.* **17**(5), 765–778 (2023).
14. Li, W. et al. Potential unreliability of uncommon ALK, ROS1, and RET genomic breakpoints in predicting the efficacy of targeted therapy in NSCLC. *J. Thorac. Oncol.* **16**(3), 404–418 (2021).
15. Li, W. et al. Reliability analysis of exonic-breakpoint fusions identified by DNA sequencing for predicting the efficacy of targeted therapy in non-small cell lung cancer. *BMC Med.* **20**(1), 160 (2022).
16. Parvaresh, H., Roozitalab, G., Golandam, F., Behzadi, P. & Jabbarzadeh Kaboli, P. Unraveling the potential of ALK-targeted therapies in non-small cell lung cancer: Comprehensive insights and future directions. *Biomedicines* **12**(2), 297 (2024).
17. Cha, Y. J., Kim, H. R. & Shim, H. S. Clinical outcomes in ALK-rearranged lung adenocarcinomas according to ALK fusion variants. *J. Transl. Med.* **14**, 1–10 (2016).
18. Solomon, B. J. et al. First-line Crizotinib versus chemotherapy in ALK-positive lung cancer. *N. Engl. J. Med.* **371** (23), 2167–2177 (2014).
19. Shaw, A. T. et al. Alectinib in ALK-positive, crizotinib-resistant, non-small-cell lung cancer: A single-group, multicentre, phase 2 trial. *Lancet Oncol.* **17**(2), 234–242 (2016).
20. Gadgeel, S. M. et al. Safety and activity of alectinib against systemic disease and brain metastases in patients with crizotinib-resistant ALK-rearranged non-small-cell lung cancer (AF-002JG): results from the dose-finding portion of a phase 1/2 study. *Lancet Oncol.* **15**(10), 1119–1128 (2014).
21. Doebele, R. C. et al. Mechanisms of resistance to Crizotinib in patients with ALK gene rearranged non–small cell lung cancer. *Clin. Cancer Res.* **18**(5), 1472–1482 (2012).
22. Doebele, R. C. et al. Analysis of resistance mechanisms To ALK kinase inhibitors in ALK + NSCLC patients (American Society of Clinical Oncology, 2012).
23. Iyevleva, A. G. et al. Novel ALK fusion partners in lung cancer. *Cancer Lett.* **362**(1), 116–121 (2015).
24. Kang, J. et al. Complex ALK fusions are associated with better prognosis in advanced non-small cell lung cancer. *Front. Oncol.* **10**, 596937 (2020).
25. Su, Y. et al. Distribution of ALK fusion variants and correlation with clinical outcomes in Chinese patients with non-small cell lung cancer treated with Crizotinib. *Target. Oncol.* **14**, 159–168 (2019).
26. Zhu, V. W. et al. A novel sequentially evolved EML4-ALK variant 3 G1202R/S1206Y double mutation in cis confers resistance to lorlatinib: A brief report and literature review. *JTO Clin. Res. Rep.* **2**(1), 100116 (2021).
27. Zhang, Y. et al. Detection of nonreciprocal/reciprocal ALK translocation as poor predictive marker in patients with first-line crizotinib-treated ALK-rearranged NSCLC. *J. Thorac. Oncol.* **15**(6), 1027–1036 (2020).

28. Childress, M. A. et al. ALK fusion partners impact response to ALK Inhibition: Differential effects on sensitivity, cellular phenotypes, and biochemical properties. *Mol. Cancer Res.* **16**(11), 1724–1736. https://doi.org/10.1158/1541-7786.Mcr-18-0171 (2018). Epub 2018/07/14.
29. Zhou, S. et al. Novel insights into molecular patterns of ROS1 fusions in a large Chinese NSCLC cohort: A multicenter study. *Mol. Oncol.* **17**(10), 2200–2212 (2023).

Acknowledgements
We thank the efforts of all investigators and researchers participating. We also thank the patients' permission to share data.

Author contributions
QY and YMG were involved in conception and design; QY and DDL were involved in administrative support; QY, YMG, DDL, WHG, HTW and XFC were involved in the provision of study materials or patients; QY, YMG and XFC were involved in the collection and assembly of data; all authors were involved in manuscript writing and revising and final approval of manuscript.

Funding
This research was funded by Scientific foundation of Zhejiang Medical and Science Project 2024KY021.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92590-9.

Correspondence and requests for materials should be addressed to H.W. or X.C.