REVIEW

Challenges and prospects in utilizing technologies for gene fusion analysis in cancer diagnostics

Xinglei Su^{1,2,3} · Qiangting Zheng⁴ · Xuehao Xiu² · Qiong Zhao⁵ · Yudong Wang⁶ · Da Han³ · Ping Song²

Received: 23 April 2024 / Revised: 29 July 2024 / Accepted: 23 August 2024 © The Author(s) 2024

Abstract

Gene fusions are vital biomarkers for tumor diagnosis and drug development, with precise detection becoming increasingly important. This review explores the links between gene fusions and common tumors, systematically evaluating detection technologies like fluorescence in situ hybridization (FISH), polymerase chain reaction (PCR), immunohistochemistry (IHC), electrochemiluminescence (ECL), and next-generation sequencing (NGS). FISH is the gold standard for DNA-level rearrangements, while PCR and NGS are widely used, with PCR confirming known fusions and NGS offering comprehensive genome-wide detection. Bioinformatic tools like STAR-Fusion, FusionCatcher, and Arriba are assessed for diagnostic accuracy. The review highlights how artificial intelligence (AI), particularly deep learning (DL) technologies like convolutional neural networks (CNNs) and recurrent neural networks (RNNs), is transforming gene fusion research by accurately detecting and annotating genes from genomic data, eliminating biases. Finally, we present an overview of advanced technologies for gene fusion analysis, emphasizing their potential to uncover unknown gene fusions.

Xinglei Su and Qiangting Zheng contributed equally to this work.

- ☑ Yudong Wang wangyudong@shsmu.edu.cn
- ☑ Da Han dahan@sjtu.edu.cn

Ping Song songpingsjtu@sjtu.edu.cn

Xinglei Su suxinglei@shu.edu.cn

Qiangting Zheng qiangtingz@tongji.edu.cn

Xuehao Xiu xiuxuehao@sjtu.edu.cn

¹ School of Life Sciences, Shanghai University, Shanghai 200444, China

² School of Biomedical Engineering, Zhangjiang Institute for Advanced Study and National Center for Translational Medicine, Shanghai Jiao Tong University, Shanghai 200240, China

- ³ Institute of Molecular Medicine Shanghai Key Laboratory for Nucleic Acid Chemistry and Nanomedicine Renji Hospital School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China
- ⁴ Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, Shanghai Key Laboratory of Signaling and Disease Research, Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
- ⁵ College of Sciences and Technology, University of Houston Downtown, Houston, TX 77002, USA
- ⁶ Department of Gynecologic Oncology, School of Medicine, The International Peace Maternity and Child Health Hospital, Shanghai Jiao Tong University, Shanghai 200030, China



Graphical Abstract



Highlights

• This review explores the intricate relationship between gene fusions and prevalent tumors, comprehensively reviewing techniques for gene fusion testing.

• It evaluates the strengths and limitations of key assays like fluorescence in situ hybridization, polymerase chain reaction, and next-generation sequencing, and explores the emerging use of electrochemiluminescence for high-sensitivity gene fusion detection.

• It emphasizes the diagnostic accuracy of bioinformatics tools such as STAR-Fusion, FusionCatcher, Arriba, and underscores the transformative role of artificial intelligence, particularly deep learning technologies, in revolutionizing gene fusion research.

Keywords Gene fusion · Cancer · Diagnostic technologies · Bioinformatics · Biological analysis · Artificial intelligence

Introduction

Gene fusion is the combination of two or more distinct gene coding regions at the DNA or RNA level, typically as a result of chromosomal rearrangements such as translocations, deletions, inversions, or tandem duplications [1]. The formation of hybrid chimeric genes is facilitated by a common collection of regulatory regions, including promoters, enhancers, ribosome-binding sequences, and terminators. Within the field of oncology, gene fusion is a crucial process that initiates unregulated cell proliferation and is a key factor in the development and progression of different types of malignancies [2]. The earliest recorded discovery of gene fusions was in 1960 by Peter Nowell and David Hungerford [3], who identified a unique microchromosome in chronic myeloid leukemia known as the Philadelphia chromosome. This landmark discovery marked the first time a recurrent chromosomal rearrangement was identified in cancer. In 1973, Janet Rowley's groundbreaking research further revealed that the Philadelphia chromosome resulted from the rearrangement of segments from chromosomes 9 and 22, which created a gene fusion [4]. In 1985, Shtivelman et al. [5] conclusively demonstrated that the BCR-ABL gene fusion is the primary oncogenic driver in chronic myelogenous leukemia (CML).

Gene fusion and cancer relationship

Types of gene fusions

Gene fusion may occur at both the DNA and RNA levels. Cancer pathogenesis is typically caused by two mechanisms: gene deregulation and fusion events that result in the formation of chimeric proteins. Gene fusions can cause gene truncation, resulting in the loss of protein-coding regions. This leads to a decrease in the expression of tumor suppressor genes, ultimately leading to cancer (Fig. 1a) [6].

A comprehensive analysis was conducted on Whole Exome Sequencing (WES) data from 268 tumor samples. The researchers focused on regions containing breakpoints and annotated them genomically. They categorized the observed gene fusion events into three main types: gene-gene fusion, gene-intergenic fusion, and intergenicintergenic fusion. Among the 13,698 gene fusion events analyzed, approximately 38% were classified as gene–gene fusions, 28% as intergenic fusions, and the remaining 34% as intergenic-intergenic fusions. These findings offer deeper insight into the complexity of tumor genomes (Fig. 1b) [8]. Based on the analysis conducted by Gao et al. [9] of the 25,664 gene fusion breakpoints examined, the majority were found within the coding sequence (CDS) of the two chaperone genes (as shown in Fig. 1c).

Tumor-associated gene fusion

Overview

Gene fusion represents a significant somatic alteration in cancer [9] and plays a pivotal role in the development of various cancer types. Many gene fusion mutations exhibit a remarkable degree of specificity, making them key targets in molecularly targeted cancer therapies [10].

In the realm of cancer research, a considerable number of ongoing studies have concentrated on the identification of distinct variations in the DNA sequence, which are referred to as single nucleotide polymorphisms (SNPs), as well as mutations involving the insertion, deletion, or duplication of genetic material within the cancer genome [11]. Thorough investigations of changes in genetic material have greatly enhanced our understanding of the fundamental processes driving the development of tumors. It is evident that gene fusions play a considerable role in complex tumor environments. Fusion events occur in different types of tumors and are important biomarkers for specific cancers, such as anaplastic lymphoma kinase (ALK) and c-ros oncogene 1 kinase (ROS1), and are rearranged during transfection (RET). Proto-oncogene tyrosine-protein kinase fusions have become key targets for tyrosine kinase inhibitors (TKIs) in the context of non-small cell lung cancer (NSCLC) [12]. Extensive evidence shows that gene fusions contribute to the progression of cancer in 16.5% of patients and play a significant role in promoting cancer development in more than 1% of patients [9]. Picco et al. [10] used CRISPR gene editing to reveal the essential functions of several gene fusion categories in promoting the proliferation of cancer cells. In addition, their investigation revealed a unique gene fusion, YAP1-MAML2, which offers promising therapeutic targets for a wide range of malignancies, including brain and ovarian cancers.

Identification and treatment progress of ALK gene fusions in NSCLC

Significant strides have been made in the field of NSCLC research over the past few decades, particularly in the identification of ALK fusion events and their targeted treatments. ALK, a transmembrane receptor tyrosine kinase part of the insulin receptor superfamily, is gene-located in the chromosomal region 2p23 and encodes a receptor tyrosine kinase within the insulin receptor family (Fig. 2a). Since the first documentation of the TPM3-ALK fusion by Lamant et al. [13] in 1999 within a specific subtype of anaplastic large cell lymphoma, and the 2007 report by Soda et al. [14] on the presence and oncogenic role of ALK gene rearrangements in NSCLC patients, ALK fusion has become a key focus in NSCLC research. Table 1 integrates the numerous known ALK gene fusion forms within NSCLC, highlighting the complexity and diversity of this research area.

Echinoderm microtubule-associated protein-like 4 (EML4) is the most prevalent partner gene of ALK, with fusion occurring at the N-terminal end. Disruption of EML4 occurs approximately 3.6 kilobases downstream of exon 13 and is related to a location 297 base pairs upstream of exon 21 of ALK [14]. The detection rate of the EML4-ALK gene fusion is estimated to be between 3 to 7% among NSCLC patients. Various EML4 breakpoints can lead to the formation of over ten EML4-ALK gene fusion variants with ALK, with variants 1 and 3 being the most common (Fig. 2c) [24]. Approximately 5% of NSCLC patients test positive for ALK fusion, primarily with EML4 as the fusion partner. Initially, two EML4-ALK variants [14], known as variant 1 and variant 2, were identified, with further research unveiling variants 3a and 3b (Fig. 2b and c) [25, 26]. Takeuchi et al.



Fig. 1 Mechanism and types of gene fusions. **a** Gene fusion is caused by structural chromosomal rearrangements. Translocations, which involve mutual translocations that create two derived chromosomes possibly containing the causative gene [left], and nonreciprocal translocations accompanied by the loss of short chromosomal arms and chromosome numbers [right]. Inversions, including inversions near centromeres occurring on one arm of the chromosome, not involving centromeres [left]; and inversions around the centromere center with two breakpoints on each arm, involving centromeres [right]. Tandem duplication involves the duplication of genes or gene segments within a specific region of a chromosome. With deletions, genes or gene segments on the chromosome are deleted or lost. A and B denote the locations of breakpoints, marked by small blue arrows, while the large black arrows indicate the resulting rearrangement of chromosomes. Reproduced with permission [7]. Copyright 2018, Elsevier. **b** The distribution patterns of various fusion types (gene-gene, gene-intergene, intergene-intergene) [8]. Copyright 2018, Springer Nature. **c** Schematic diagram of gene fusion breakpoints. The purple symbol indicates the 5' gene partner, and the green symbol indicates the 3' gene partner. For gene fusion, breakpoints can appear in the following genomic regions: 5' UTR (first column from left to right), CDS (second column), 3' UTR (third column), and noncoding region (fourth column). The dotted line indicates the junction between the breakpoints of the gene fusion, and the circle size and number indicate the number of occurrences of the corresponding fusion type [9]. Copyright 2018, Elsevier



Fig.2 ALK gene fusion types in tumors. **a** ALK sequencing results from the full reference sequence of the Genome Browser (https://genome. ucsc.edu/). **b** Rearranged form of ALK with EML4 in NSCLC. **c** ALK fusion in patients with NSCLC [15]. Copyright 2020, De Gruyter. **d** The fusion positions of ALK are indicated by vertical bars, and five variants (V) of EML4-ALK have been identified, namely, 1, 2, 3, 4 and 5. Among these variants, exons 13, 20, 6, 14 and 2 of EML4 cDNA were fused with exon 20 (e20) of ALK cDNA. Reproduced with permission [16]. Copyright 2008, John Wiley and Sons

Table 1 ALK rear	rangements in NSCLC
------------------	---------------------

Cancer type	Frequency of ALK rearrangements	ALK fusion partner gene	Location of fusion partner	Reference
NSCLC	3%-7%	TPR	1q31.1	[17–23]
		CRIM1	2p22.2	
		EML4a	2p21	
		STRN	2p22.2	
		TFG	3q12.2	
		HIP1	7q11.23	
		PTPN3	9q31	
		KIF5B	10p11.22	
		KLC1	14q32.3	
		CLTC	17q23.1	

[16] developed a single-tube multiplex reverse transcription polymerase chain reaction (RT-PCR) method [25], successfully detecting all known and new EML4 and ALK fusion variants, including variants 4 and 5 (Fig. 2d).

Recent years have seen significant breakthroughs in the treatment of ALK fusion-positive NSCLC [27]. Crizotinib, the first-generation FDA-approved small molecule ALK inhibitor, marked a new era in this field. With a deeper understanding of resistance mechanisms, second-generation (such as ceritinib, alectinib, brigatinib) [28] and third-generation (such as lorlatinib) ALK inhibitors were subsequently developed and introduced, significantly enhancing treatment outcomes and offering new hope to ALK fusion-positive NSCLC patients [29–33]. Table 2 provides an overview of several FDA-approved ALK inhibitors currently utilized in clinical practice. These advancements not only improved patient survival rates but also propelled a shift towards personalized medicine and precision treatment strategies.

Breakthroughs in targeted therapies for cholangiocarcinoma (CCA): focusing on fibroblast growth factor receptor (FGFR) fusions

The field of cholangiocarcinoma (CCA), a subtype of liver cancer related to the biliary tract, has witnessed significant advancements in targeted therapies, particularly with the focus on fibroblast growth factor receptor (FGFR) fusions. FGFR, part of the receptor kinase family with four subtypes (FGFR1, FGFR2, FGFR3, and FGFR4), has become a key target in developing treatments for CCA. These receptors interact with 18 distinct fibroblast growth factors (FGFs) [42], crucial for cell survival, proliferation, differentiation, and motility, as well as playing vital roles in embryogenesis, wound healing, angiogenesis, and carcinogenesis [43–45]. Among these, FGFR2, especially noted in intrahepatic cholangiocarcinoma (iCCA), where its fusion occurs in about 10-15% of patients [46–49], has been identified as a significant biomarker and therapeutic target. This fusion typically involves FGFR2 and a chaperone gene, leading to the production of a functional protein that promotes dimerization and oligomerization [42, 46], with BICC1 among its common fusion partners [50, 51].

In recent years, the FDA approval of four targeted drugs for CCA has ushered in a new dawn of hope for patients. Pemigatinib, an orally administered inhibitor, became the first targeted medication approved for CCA treatment in the United States in April 2020 [52] and in China in April 2022. Following this, truseltiq (infigratinib), an ATPcompetitive inhibitor of the FGFR tyrosine kinase family, was approved in May 2021 for patients with FGFR2 alterations [53]. Ivosidenib, the first isocitrate dehydrogenase-1 (IDH1) inhibitor, was approved in August 2021, introducing precision treatment options for patients with advanced CCA with IDH1 mutations [54]. In September 2022, futibatinib, targeting adults with FGFR2 gene fusions or rearrangements, further expanded the arsenal of targeted therapies for CCA. These advancements not only signify a leap forward in the personalized treatment of CCA but also highlight the growing importance of FGFR fusions as biomarkers and therapeutic targets, offering new pathways for research and patient care [55]. Table 3

Table 2	Several FDA-a	approved AL	K inhibitors	have been	used in	clinical	settings

Study Drug	Research and development company company	FDA approval time to market	Targets other than ALK	Phase of testing	References
Crizotinib	Pfizer	2011	ROS1 MET	Phase III complete	[34, 35]
Ceritinib	Novartis	2014	ROS1 IGF1R IR	III	[35]
Alectinib	Genentech	2015	GAK LTK RET	III	[35, 36]
Brigatinib	Ariad	2017	ROS1	III	[35, 37–39]
Lorlatinib	Pfizer	2018	ROS1	III	[35]
Entrectinib	Ignyta	2019	NTRK1 NTRK2 NTRK3 ROS1	Π	[40]
Ensartinib	Xcovery	2020	ROS1 MET AXL	III	[40, 41]

ALK TKI	Research and develop- ment company company	FDA approval time to market	Listed in China	Pathway targeted	ORR (%)	DCR (%)	Reference
Pemigatinib	Incyte	2020	Yes	FGFR2	36	82	[56]
Infigratinib	BridgeBio	2021	No	FGFR2	23	83.6	[57]
Ivosidenib	Agios Pharmaceuticas	2021	No	IDH1	2	51	[58]
Futibatinib	Taiho Oncology	2022	No	FGFR	25.4	71.9	[59]

Table 3 FDA-approved targeted drugs for the treatment of CCA

summarizes the FDA-approved targeted therapies for the treatment of CCA.

Gene fusions identified in other malignancies

Gene fusions are common not only in NSCLC and CCA but also in a variety of cancers. Gene fusions, including prominent examples such as HNRPA2B1-ETV1 [60], TMPRSS2-ERG, and Breakpoint Cluster Region (BCR)-Abelson Murine Leukemia Viral Oncogene Homolog 1 (ABL), are highly prevalent in cancers like prostate cancer and CML. Notably, the TMPRSS2-ERG fusion is detected in approximately 50% of prostate cancer patients [61], and the BCR-ABL gene fusion, a hallmark of CML, is found in the majority of CML cases. In colorectal cancer, the DNA repair gene RAD51C can fuse with ATXN7 at a fusion rate of 36% [62], these examples highlight the critical importance of gene fusions in various cancers as the list of identified gene fusions continues to grow, further establishing their link to cancer development.

Comparison of RNA and DNA level gene fusion assays and their advantages

Overview

DNA and RNA gene fusions are two distinct molecular events at the genomic and gene expression levels. At the DNA level, gene fusion typically refers to the merging of DNA segments or chromosomes from different sources. Such fusion events may result in novel genomic structures, including gene rearrangements, insertions, or deletions. At the RNA level, gene fusion involves the merging of two distinct RNA molecules that commonly occur during transcription and splicing processes. These fusion events play crucial roles in regulating the diversity of genomic structures and expression at the genetic level.

DNA-based gene fusion detection offers a more comprehensive view of the genome, yielding extensive sequencing data. This approach not only identifies gene fusion events across the entire genome but also uncovers complex structural changes, providing a deeper understanding of the genomic landscape. However, gene fusion breakpoints predominantly manifest within intron regions and tend to be longer than within exon regions [63]. This poses a challenge for creating breakpoint capture probes using DNA next-generation sequencing (NGS) technology, as these long intronic regions contain GC-rich and highly repetitive sequences, along with some complex rearrangements. This complexity increases the detection cost, and it also affects probe capture efficiency and sequence alignment accuracy.

In contrast, RNA-based NGS can directly detect exonexon junctions, indicating fusion events. This approach aids in precise probe or primer design, increasing the accuracy and precision of recovery from high-GC content and repetitive regions. Additionally, when the tumor purity of a sample is low, the variant allele frequency (VAF) of mutations may be below detectable levels. However, at the transcript (RNA) level, the theoretical overexpression of fusions can potentially compensate for the low VAF. In many cases, RNAbased methods have the potential to overcome the limitations of DNA-based gene fusion detection techniques, providing a more effective approach for identifying gene fusions with significant advantages (Fig. 3a) [64, 65].

Advantages of RNA-based gene fusion detection

As previously mentioned, RNA sequencing (RNA-seq) probes are more straightforward than other probes and offer an impartial assessment of larger transcripts across a dynamic range [68, 69]. In regard to accuracy, sensitivity, and cost-effectiveness, RNA-level detection outperforms DNA-level methods.

In a study conducted by Seo et al. [70] in 2012, significant upregulation of RNA expression was observed for gene fusions (ALK, ROS1, and RET). Additionally, Beaubier et al. [71] highlighted that RNA-seq surpasses DNA sequencing (DNA-seq) in detecting gene fusions across pancancer populations. Sheikine et al. [72], in their study, identified the potential for false-negative results in DNA-based NGS, particularly when fusion breakpoints reside within extensive intron regions containing numerous repetitive elements. Li et al. [34] conducted comprehensive studies involving DNA-NGS and RNA-NGS



Fig. 3 Comparison of RNA and DNA gene fusion detection. **a** Schematic diagram of the potential genomic complexity that could lead to falsenegative gene fusion results in DNA-based NGS analysis. In some cases, RNA-based approaches may overcome the limitations of DNA-based detection [66]. Copyright 2019, American Association for Cancer Research. **b** A schematic diagram illustrating the primary NGS-targeting methods used for the gene fusion assays. Hybridization capture: Gene-specific enrichment is achieved using biotinylated DNA or RNA probes that target the region of interest through hybridization steps. Classical amplification methods: Target enrichment is accomplished through multiplex PCR employing fusion variant-specific primers. Anchor multiplex PCR (AMP): This method focuses exclusively on one fusion partner. By anchoring one side of the target-specific primer while the other end is randomly connected to the universal adapter, AMP enriches the region of interest with information from only one end [67]. Copyright 2020, Diagnostics

with 3,787 patients with NSCLC. Notably, among the 140 patients initially classified as negative by DNA-NGS, RNA-NGS using the Hengte gene identified gene fusions/ rearrangements in 10%, thus highlighting potential limitations of DNA-NGS in detecting specific fusion variants.

In summary, RNA-level detection offers a solution to challenges related to the presence of large intron regions and the concurrent involvement of multiple exon fusions in the identification of gene fusions [73]. Furthermore, in terms of predicting the most suitable targeted medication, RNA-level-based tests demonstrate exceptional efficacy [34].

This analysis clearly demonstrates that RNA-level detection is superior in identifying a diverse array of rare gene fusions compared to DNA-level methods. It emphasizes the critical importance of RNA-based techniques in achieving accurate and comprehensive gene fusion detection in practical testing environments.

Challenges and considerations in RNA-based gene fusion detection

Although RNA-level gene fusion testing offers heightened sensitivity and accuracy compared to DNA-level testing, it presents its own set of challenges. RNA-based targeted methods require direct analysis and quantification of fusion transcripts, yet their effectiveness can be limited by the availability and quality of the RNA [67].

Given its inherent single-stranded nature, RNA is more susceptible to instability compared to the double-stranded

Page 9 of 23 14

structure of DNA. This vulnerability is particularly pronounced when working with RNA extracted from formalin-fixed, paraffin-embedded (FFPE) specimens, which presents additional challenges due to its tendency towards degradation. Consequently, conducting comprehensive quality assessments becomes crucial prior to NGS analysis to ensure accurate and reliable results [74].

During RNA extraction, extremely small or excessively large sample sizes present challenges. Small samples may yield limited RNA from cell tissues, whereas large volumes can cause incomplete lysis and low yields. Clinically, tumor samples, especially FFPE specimens after pathological examination, often contain partially degraded and lowerquality RNA. Low-quality RNA may result in uneven gene coverage and higher false positive rates [75]. The ongoing challenges in RNA gene fusion detection involve low RNA content and compromised quality.

Methods for gene fusion detection

The reasonable selection of gene fusion diagnosis methods is crucial for effectively predicting patient benefit, and accurate tumor-targeted therapy depends on accurate gene fusion detection. Currently, commonly used gene fusion detection methods include fluorescence in situ hybridization (FISH), RT-PCR, immunohistochemistry (IHC), electrochemiluminescence (ECL) [76], NGS, and several NGS-based approaches [77]. These methods vary in their techniques and applications, providing different ways to detect and analyze gene fusions.

FISH uses fluorescently labeled DNA probes to hybridize with target DNA sequences within the cell nucleus, thereby obtaining information about the chromosome or gene status within the nucleus [78, 79]. IHC detects antigens within tissue cells through antigen–antibody specific binding and chromogenic reactions, enabling localization, qualitative, and relative quantitative analysis of gene fusions at the protein level [80]. PCR analysis of gene fusions is based on designing oligonucleotide primers on both sides of the breakpoint fusion regions, ensuring that the PCR product contains the tumor-specific fusion sequences [81, 82]. NGS detects gene fusions by using high-throughput sequencing and alignment analysis to identify reads spanning gene boundaries [83].

FISH gene fusion testing gold standard

FISH has long been considered the gold standard for detecting ALK and ROS1 rearrangements in NSCLC patients, as it detects gene rearrangements at the DNA level (Fig. 4a). Primarily, isolation probes that do not require prior knowledge of fusion partners can distinguish rearrangements from polyploidy and amplification but cannot be used to identify the specific fusion variant precisely. In lung cancer, FISH detects both known and unknown fusion variants. Erin E. and colleagues assessed gene fusion detection in two lung cancer biopsies diagnosed via FISH. Its reliability and stability make it essential for validating other detection methods [84, 85].

Polymerase chain reaction (PCR) techniques for gene fusion detection

PCR is a molecular biology technique employed to amplify specific DNA fragments, significantly increasing the quantity of DNA. PCR technology is frequently applied in the detection of gene fusions.

RT-PCR is also used to detect gene fusions, as it boasts short cycling times and high sensitivity. However, new gene fusions partners or complex structural rearrangements cannot be identified. Figure 4b illustrates the detection principle of the RT-PCR method (Fig. 4b) [88]. Currently, digital PCR (dPCR) is emerging as the primary method for detecting gene fusions. However, traditional PCR techniques often struggle to detect new gene fusions when used alone. With the development of the GeneChip exon array, comprehensive genome-wide exon expression analysis has become feasible. Wada et al. [89] introduced an innovative approach involving exon arrays for detecting abnormal gene structures. By combining RT-PCR methods, they successfully discovered 3 new gene fusions in breast cancer and pancreatic cancer cell lines.

Multiplex polymerase chain reaction (MPCR) is a variant of PCR that simultaneously amplifies multiple nucleic acid fragments using several primer pairs in a single reaction. The principles, reagents, and operational procedures of MPCR are similar to those of conventional PCR. MPCR enhances throughput, conserves samples, and reduces costs, albeit at the expense of additional time and effort. However, this approach produces more comprehensive and valuable data. Figure 4d shows a schematic of the conventional MPCR principle [86].

With ongoing research and the advancing maturity of detection technologies, a growing array of innovative PCR methods has emerged. Notably, techniques such as anchored multiplex PCR and single-primer extension (SPE) will be discussed in detail below, especially when it comes to their application combined with NGS. These advancements signify significant progress in enhancing the precision, sensitivity, and efficiency of PCR-based assays.

NGS for comprehensive gene fusion detection

NGS, also known as high-throughput sequencing, is characterized by its high output and high resolution. This Fig. 4 Common methods for gene fusion testing. a Gold standard for gene fusion detection. In the lung cancer sample MO-16-000393, FISH revealed rearrangement of the ROS1 gene. Positive signals consisted of a fused set of red and green dots in each cell, along with at least one isolated green dot. Fused dots are indicated by white arrows, while isolated green dots are indicated by gray arrows [78]. Copyright 2019, Springer Nature. b RT-PCR detection of gene fusions. c Hybrid-capture approach [78]. Copyright 2019, Springer Nature. d Conventional MPCR utilizes gene-specific primers (GSPs) to selectively amplify known fusion junctions [86]. Copyright 2020, John Wiley and Sons. e Design of a singleend duplex unique molecular identifier (UMI) adapter [87]. Copyright 2019, Springer Nature



technology excels in large-scale parallel sequencing, offering ultra-high throughput, excellent scalability, and rapid data processing. Additionally, NGS can be used for rapid wholegenome sequencing (WGS) and provides deep sequencing of targeted regions.

NGS-based gene fusion detection methods identify fusion events at both DNA and RNA levels. Library preparation is crucial for successful NGS, achievable through hybrid capture or amplicon methods. Hybrid capture uses sequencespecific probes to target specific genomic regions, while amplicon methods rely on multiplex PCR to enrich target sequences. Although hybrid capture provides detailed information about adjacent regions, it is complex, time-consuming, and prone to errors. In contrast, amplicon methods are simpler and faster.

Currently, RNA-based NGS is the most widely used method for detecting multi-target fusions. Commonly used fusion detection platforms include Illumina's AmpliSeq, ArcherDX's AMP technology, and QIAGEN's QIAseq. All of these platforms use amplicon-based methods for library preparation.

AmpliSeq technology facilitates multiplex amplification of 12 to > 24,000 amplicons, enabling the simultaneous capture of multiple targets in a single reaction. This method allows for comprehensive transcriptome assessment and represents an innovative targeted whole-transcriptome RNA-seq approach for gene expression analysis [90]. AmpliSeq proves highly sensitive and cost-effective, particularly for large-scale gene expression analysis and precise mRNA marker screening [15]. ArcherDX's patented AMP technology not only enhances the specificity of PCR amplification but also captures unknown ends of gene fusions (using universal primer ends). This method enables the discovery of more unknown gene fusions in cancer samples [67]. When used in conjunction with NGS, Qiagen's SPE technology can amplify and analyze DNA sequences linked to potential gene fusion events. Unlike traditional PCR amplicon methods, SPE allows up

to 20,000 multiplex PCR reactions in a single tube while maintaining over 95% specificity and uniformity [87].

Hybrid-capture approach in NGS-based gene fusion detection

As previously discussed, although RNA-seq has many advantages for gene fusion identification, traditional RNAseq methods may overlook low-expressed transcripts and fusion transcripts in samples with low tumor content or low tumor cell fractions due to sequencing depth limitations. To overcome the limitations of traditional RNA-seq in terms of resolution and yield, Mercer et al. [91, 92] developed a new targeted RNA-seq method using hybridcapture approach. This approach uses biotinylated oligonucleotide probes to enrich RNA transcripts of interest and integrates the gold standard FISH technique (Fig. 4c). By targeting and capturing hundreds of genes in a single assay, this method increases sequencing coverage, enabling sensitive detection of rare or low-abundance transcripts. Compared to traditional FISH and RT-PCR methods, targeted RNA-seq has improved the diagnostic rate from 63-76%.

AMP Technology: advancements in NGS-based gene fusion detection

As researchers explore the field, various PCR-based technologies have been developed. In a 2014 report, the introduction of single-primer PCR technology, specifically for AMP, offered a rapid target enrichment method for NGS (Fig. 3b). The technology begins by fragmenting the sample DNA and attaching molecular barcodes and universal primer sites, which significantly reduces PCR amplification bias. It then designs two PCR primers (GSP1 and GSP2) for the target region and performs nested PCR using the universal primer sites. Using AMP technology, a single-end primer design is sufficient to detect all fusion events, regardless of whether the gene fusions are known or novel (Fig. 4d). AMP has demonstrated adaptability to FFPE specimens with low nucleic acid content, cost effectiveness, and suitability for large-scale sequencing approaches, such as WGS or comprehensive transcriptome sequencing [93].

Haas et al. [94] describe AMP as a rapid NGS enrichment method for RNA and DNA. In their study, the library remained functional after just two rounds of PCR, demonstrating its cost-effectiveness and suitability for large-scale sequencing approaches such as WGS or comprehensive transcriptome sequencing.

Advancing targeted NGS analysis with unique molecular identifiers (UMIs) and SPE technology

Amplicon-based NGS analysis provides advantages in targeted enrichment. However, as the number of PCR cycles and primer multiplexing increase, the risk of bias and errors also increase. To address these issues, QIAseq NGS plates incorporate UMIs for PCR bias correction. Additionally, they utilize SPE technology to enhance design flexibility and precisely enrich target regions. Notably, SPE accommodates variable amplicon sizes without predefined limits, providing an efficient solution for specific enrichment needs [87].

UMI technology ensures precise labeling of each molecule in the sample library, aiding in error correction during sequencing and greatly improving accuracy. Peng et al. [87] achieved both duplex sequencing and enhanced MPCR using a unique duplex UMI adapter. This method enables the accurate detection of SNVs at allele fractions down to 0.1-0.2%. In comparison to existing targeted duplex sequencing methods, this new targeted sequencing approach not only leverages duplex UMIs to eliminate NGS artifacts but also streamlines the workflow (Fig. 4e).

Despite the development of numerous RNA-based sequencing methods, the cost of RNA-seq remains relatively high, at 980 euros per case [95]. It is crucial to address this issue, as researchers are striving to develop faster, more precise, and affordable alternatives.

Electrochemiluminescence (ECL) for high-sensitivity gene fusion detection: exploring new frontiers

ECL, though not commonly employed for gene fusion detection, stands at the forefront of technological innovation. It utilizes the power of luminescent signals generated through electrochemical reactions to accurately identify specific biomolecules, thus offering a novel, albeit less conventional, approach for the detection of gene fusions. This technique is executed via ECL emitters in a process where light radiation is initiated electrochemically through the energy relaxation of excited species, leading to highly sensitive detection of a diverse array of biomarkers [96–98]. By integrating these advanced processes, ECL provides groundbreaking advantages in the field of bioanalytical detection, making it a valuable tool for identifying gene fusions and various other biomolecular interactions. ELC requires crafting and labeling bespoke probes that identify gene fusion sequences, which then specifically hybridize with corresponding target sequences. By applying an electrical potential, the tagged luminescent reporter molecules on the probes are excited, emitting a detectable light signal. This signal is captured by a detection apparatus and subjected to software analysis, effectively confirming the existence of gene fusions.

Fig. 5 Electrochemiluminescence method for gene fusion testing. a Schematic illustration of an 'on-off' ECL biosensor based on configuration-entropy driven DT-SDR and CeO2/ MXene heterojunction for highly sensitive detection of BCR-ABL gene fusion [76]. Copyright 2022, Elsevier. b Schematic Diagrams of the Ultrasensitive 'on-off' ECL Biosensing for BCR-ABL Gene fusion Determination on Basis of the Well-Regulated Track-Based DNA Walker and Au@g-C₃N₄ NHs [97]. Copyright 2020, ACS publication



Cheng et al. [76] have developed an innovative ECL biosensor for detecting the BCR-ABL gene fusion, which utilizes a CeO₂/MXene heterojunction and leverages a dual-toehold strand displacement reaction to amplify the signal (Fig. 5a). Created through a one-step hydrothermal method, the CeO₂/MXene heterojunction enhances ECL emission and serves as an effective electrode material. The presence of the BCR-ABL gene initiates a sequence of strand displacements, culminating in an 'on-off' ECL signal controlled by quenching labels attached to Pt nanoparticle-functionalized polydopamine composites. This biosensor showcases a wide detection range from 1 fM to 100 pM and an impressively low detection limit of 0.27 fM, offering a promising approach for the molecular diagnosis of chronic myelogenous leukemia. Lv et al. [97] reported another ECL biosensor capable of sensitively detecting the BCR-ABL gene fusion by employing a highly efficient DNA walking strategy on electrode surfaces (Fig. 5b). This method constructs regulated DNA tracks using a supersand wich hybridization chain reaction to generate linear double-stranded DNA. These tracks, assembled on an electrode modified with an Au nanoparticle-enhanced g-C3N4 nanohybrid, facilitate the movement of bipedal DNA walkers. These walkers, triggered by the presence of the BCR-ABL gene fusion, replace quenched folic acid-labeled strands, thereby enhancing ECL signal for amplified detection. This precise control over the DNA walking process significantly improves signal amplification efficiency, achieving a detection limit of 0.18 fM for the BCR-ABL gene fusion. This breakthrough demonstrates substantial promise for clinical molecular diagnostics applications.

ECL bioanalysis merges the advantages of electrochemical and luminescence technologies, making it highly effective for clinical diagnostics, biomarker discovery, and environmental monitoring due to its sensitivity and low noise. Despite its promise in gene fusion detection, ECL encounters several significant challenges, including the necessity for highly specific probes, complex sample preparation processes, and a heavy reliance on bioinformatics. Addressing these hurdles effectively requires leveraging advanced technologies to enhance ECL's capabilities and overcome its limitations. These challenges indicate that while ECL is powerful, its application in gene fusion detection requires further innovation and strategic adaptation. Thus, ECL's significant potential in detecting gene fusions will likely depend on tailored approaches and integration with other technologies.

Enhanced fusion transcript identification: bioinformatics tools for high-throughput gene fusion detection

Gene fusions are detectable via both DNA-seq and RNAseq, although RNA-seq is preferred because of its simplicity and cost-effectiveness. As a result, several tools have been developed for detecting fusion transcripts from RNAseq data, and many studies have compared these RNA-seq techniques [99–103]. However, these studies often focus on specific RNA-seq analysis steps, and workflow analysis is usually limited to only a few stages [104, 105].

The accurate identification of fusion transcripts is essential for obtaining a comprehensive understanding of the cancer transcriptome. Numerous bioinformatics tools, such as FusionSeq, deFuse, Fusion-Hunter, TopHat-Fusion, and STAR-Fusion, can be used to detect gene fusions. In Table 4, an overview of frequently used bioinformatics analysis software methods is presented.

In gene fusion analysis, high-throughput sequencing methods are frequently used to detect potential gene rearrangements. Software tools such as STAR-Fusion, Arriba, and FusionCatcher are part of the category of programs designed for discovering gene fusions from transcripts [113]. These software applications typically initiate the process by mapping reads to reference genomes through internal alignment. They subsequently identified soft-clipped reads and compared their sequences to the reference genome to identify potential fusion partners.

FuSeq: a computational breakthrough for high-throughput gene fusion detection

With research advancements, the introduction of paired-end RNA-seq has simplified the identification of gene fusions. However, current methods often involve extensive computational requirements, limiting their utility for routine analysis of large sample sets. To address this challenge, Vu et al. [114] have introduced FuSeq, a mapping-based approach tailored for the swift and precise detection of gene fusions. FuSeq efficiently aligns reads, extracts initial fusion candidates from segmented read groups, and subsequently applies multiple filters and statistical tests to pinpoint the final candidates. FuSeq stands out for its computational efficiency, speed, and ability to enable high-throughput gene fusion studies.

STAR-Fusion: fast fusion transcript detection

STAR-Fusion, designed for the rapid and precise identification of fusion transcripts from RNA-Seq data, utilizes the output of STAR for detection [110, 115]. It offers highly accurate and sensitive gene fusion assays, making it adept at detecting low-frequency or rare fusion events. This makes STAR-Fusion a valuable tool in various research domains, including cancer genomics and other disease-related studies. This approach capitalizes on the computational efficiency of STAR for sequence alignment, boasting an impressive endto-end processing rate of 550 million Illumina reads using 12 threads. However, this speed, which is 450 times faster than that of its nearest competitor, TopHat2 [110], comes with high memory demands. Analyzing a single human genome sample requires approximately 30 GB of memory, which increases to approximately 40 GB for gene fusion

Table 4 RNA-seq-based fusion transcript predictors evaluated [94]

Method	Brief overview of methodology	Sequencing type	Reference
STAR-Fusion	Uses chimeric read alignments reported by STAR in its Chimeric.out.junction file to identify candidate fusions followed by extensive filtering of likely artifacts	RNA-seq	[106]
FusionCatcher	Leverages a collection of alignment utilities including Bowtie [107], Bowtie2 [108], BLAT [109], and STAR [110]with a collection of customized target databases to identify and characterize fusion candidates. Rigorous filtering of fusion predictions according to gene and fusion annotations is employed	RNA-seq	[70, 111]
Arriba	Arriba extracts gene fusions from the chimeric alignments reported by STAR [110] by applying a collection of filters which recognize frequent types of artifacts found in RNA-Seq data	RNA-seq	[112]

detection, necessitating substantial computational resources. Although exceptionally efficient, STAR-Fusion has limitations, particularly in identifying breakpoints between genes and within introns, where tools such as Arriba may offer more precise results.

FusionCatcher: detecting somatic gene fusions in RNA-seq data

FusionCatcher is a versatile tool adept at identifying both known and previously unknown gene fusion events. Moreover, this approach is highly sensitive, particularly for detecting low-frequency gene fusions, making it suitable for a broad spectrum of gene fusion testing applications in cancer research and other disease contexts [111]. The tool incorporates an integrated quality control mechanism that automatically filters out artifacts and low-quality data, thereby enhancing the accuracy of its findings. FusionCatcher primarily conducts its analysis at the RNA level, with a specific focus on meticulously examining the transcriptome. The tool employs the Ensembl genome annotation [116] and Bowtie aligner [107] for processing sequencing reads, treating them as single reads within the transcriptome context. Demonstrating robust performance, FusionCatcher has proven effective in detecting fusion events and validating them via real-time PCR in RNA-seq data from tumor cells. Notably, this technique has achieved a remarkable success rate, identifying the DNAJB1-PRKACA fusion in 100% of fibrolamellar hepatocellular carcinoma patients, underscoring its efficacy and reliability [117].

Software tools such as STAR-Fusion and FusionCatcher lack the capacity to recognize breakpoints that occur between genes and inside introns. Additionally, duplications of exons and gene fusions that are inverted cannot be detected. Current gene fusion detection tools exclusively utilize RNA data to identify gene fusions, hence restricting their ability to detect gene fusions that exist in DNA but are not expressed in RNA.

Arriba: rapid and accurate gene fusion detection in RNA-Seq data

Arriba is a specialized software package for the detection of gene fusions within RNA-Seq data. It stands out due to its efficient algorithms and parallel computing capabilities, allowing it to process swiftly large datasets. This approach increases the overall efficiency and fault tolerance of the analysis, enabling it to handle data from diverse samples and experimental conditions, thereby enhancing its versatility in various research scenarios [118]. Unlike many other fusion detection pipelines, Arriba can utilize existing STAR alignments, eliminating the need for separate read alignments designed specifically for gene fusion analysis. Arriba excels in detecting intragene inversions, duplications, and translocations within both intron and intergene regions, demonstrating its remarkable performance, particularly when analyzing a small quantity of fusion transcripts [117]. Additionally, Arriba includes visualization tools, which greatly aid in the interpretation and understanding of gene fusion events.

Arriba excels in sensitivity across four benchmark dataset types, with remarkable specificity even at lower fusion transcript concentrations, approaching optimal true-positive enrichment. Like FusionCatcher, Arriba accommodates known fusions. Its workflow, similar to that of STAR-Fusion but more efficient, reduces the fusion candidate filtering time. By running Arriba sequentially, peak memory usage can be minimized, though at a slight runtime extension [118]. In summary, Arriba is a high-performance gene fusion detection tool that offers exceptional sensitivity, specificity, and multitype fusion detection capabilities. It also provides intuitive visualization tools for rapid and accurate fusion event detection.

Advancing gene fusion research through artificial intelligence (AI)

Artificial Intelligence (AI) enables computers to perform tasks that typically require human intelligence, such as voice recognition, natural language understanding, image recognition, problem-solving, and learning. AI aims to create intelligent systems that can adapt to new situations by mimicking human thought processes and learning capabilities. AI technologies are categorized into rule-based systems, machine learning (ML) [119], deep learning (DL) [120], natural language processing (NLP) [121], and computer vision [122], among others. Rule-based systems rely on logical rules, while ML and DL learn patterns through data. NLP allows machines to understand human language, and computer vision enables them to process information from images.

As mentioned earlier, despite its potential for gene fusion detection, ECL faces hurdles such as the need for specific probes, complex sample preparation, reliance on bioinformatics, and competition with established methods like NGS and RT-PCR. Therefore, the significant potential of ECL in detecting gene fusions could be fully realized by leveraging AI to address the challenges of complex probe design and bioinformatics analysis. The advantages of AI, including efficient automated design, optimization of specificity and sensitivity, rapid processing of large-scale data, and precise prediction of gene functions and disease associations, can overcome existing bottlenecks in ECL for gene fusion detection.

The complexity and diversity of gene fusions, coupled with the vast amount of genetic data, necessitate advanced computational methods for their identification, characterization, and functional analysis. DL, particularly convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variant, long short-term memory (LSTM) networks, have shown immense potential in the field of gene fusion research [123]. These sophisticated AI models, with their exceptional ability to recognize complex patterns in data, have been applied to identify gene fusions from genomic sequences and transcriptomic data. By learning from the inherent features of the data rather than relying on explicitly programmed genetic characteristics, these models aim to capture the linear structure and spatial dependencies within DNA and RNA sequences, providing powerful tools for the detection and analysis of gene fusions [124]. Furthermore, DL not only plays a crucial role in detecting gene fusions, it significantly advances our understanding of the functions of fusion proteins. By analyzing the genetic sequences of fusion proteins, AI models can predict their oncogenic potential, interactions within cellular pathways, and possible impacts on tumor behavior, thereby revealing new therapeutic targets and offering personalized and precise approaches for cancer treatment.

Conducted by Marta Lovino [125], the study employs CNNs for annotating chimeric transcripts using raw fusion sequence data, aiming to bypass potential biases from protein domain analysis. By feeding the actual amino acid composition of fused proteins into CNNs, the model learns classification features directly, eliminating the need for handcrafted descriptors. The network generates a 0–1 score reflecting the chimeric transcript's potential oncogenic involvement, which classifies gene fusions into oncogenic or not, with associated confidence levels. This approach showcases the versatility of CNNs in pattern recognition and classification, highlighting their significant potential in bioinformatics without relying on additional protein domain data.

Pora et al. [126] have developed a deep residual neural network model named FusionAI, designed to predict potential gene fusion breakpoints based on RNA-seq data. FusionAI operates by solely utilizing raw DNA sequences as input, attempting to decipher the characteristics enriched in genomic breakpoint regions. This model leverages gene fusion data from The Cancer Genome Atlas (TCGA) to distinguish between fusion-positive and fusion-negative breakpoints. Through the analysis of approximately 26 K fusion breakpoints and application to multiple external validation datasets, FusionAI has demonstrated its potential to enhance the specificity of gene fusion prediction. Additionally, the model explores genomic features associated with fusion events, including specific DNA sequence motifs enriched in fusion-positive sequences. Overall, FusionAI represents an example of the application of DL in studying human genomic breakages and their related genomic regions, showcasing the powerful potential of DL in complex genomic data analysis.

Vipulkumar et al. [127] developed a DL-based algorithm to identify ERG rearrangement status in prostatic adenocarcinoma using only digitized slides of H&E morphology. The study utilized whole slide images from 392 cases, annotated with QuPath, and exported image patches of 224 × 224 pixels at $10 \times$, $20 \times$, and $40 \times$ magnifications for input into a MobileNetV2 convolutional neural network model. Separate models were trained for each magnification using 261 cases for training and 131 for testing. The model output predicted ERG-positive (ERG rearranged) or ERG-negative (ERG not rearranged) status for each input patch. The models showed similar ROC curves with AUC results ranging from 0.82 to 0.85, and the 20× model demonstrated sensitivity and specificity of 75.0% and 83.1%, respectively. This DL model successfully predicted ERG rearrangement status in the majority of prostatic adenocarcinomas, potentially eliminating the need for ancillary studies to assess ERG gene rearrangement.

While AI offers numerous advantages in gene fusion detection and analysis, it also has notable limitations. A significant challenge is the high false negative rate, as AI models may miss certain gene fusions. The need for specific probes and complex sample preparation can reduce the efficiency and practicality of AI applications. Additionally, the reliance on bioinformatics for data processing and analysis add to complexity, requiring specialized knowledge and resources. AI models are also vulnerable to biases in training data, which can affect the fairness and accuracy of their predictions. Addressing these limitations will be crucial for fully realizing AI's potential in gene fusion detection and analysis.

Future prospects of AI in enhancing gene fusion research and personalized medicine

The future of AI in gene fusion research is promising, with advancements anticipated in several key areas. As DL technology continues to evolve, AI models are expected to become more accurate and sensitive, capable of detecting subtle fusion events within complex genomic data, even in early-stage cancers or low-abundance samples. AI will play a significant role in predicting the functions and disease relevance of gene fusion products, shedding light on their involvement in cellular signaling pathways, metabolic routes, and association with specific disease states. This insight will propel personalized medicine forward, utilizing individual genomic information, including gene fusion characteristics, to tailor treatment plans, match patients with the most effective drugs and therapies, and predict responses to specific treatments. Furthermore, AI will increasingly integrate data from various biological levels-genomics, transcriptomics, proteomics, and metabolomics-offering a more comprehensive biological perspective and deepening our understanding of complex biological processes. Advances in wearable devices and biomarker detection technology may also enable AI to play a role in real-time monitoring of gene fusion expression and activity, aiding in early disease diagnosis, treatment monitoring, and disease progression. Overall, the future trends in AI applications for gene fusion research and applications are set to focus more on precision medicine, multi-omics data integration, and real-time disease monitoring, driving new advancements in biomedical research and clinical treatments.

Summary and outlook

Various techniques are available for assessing gene fusion, which can be categorized as single-gene detection or multiplex detection, traditional or next-generation methods, and DNA-, RNA-, or protein-based methods. Single-detection methods include FISH, IHC, and RT-PCR [74]. With the rapid development of NGS technology, this approach has been applied to gene fusion detection. NGS applications encompass WGS, WES, and RNA-Seq, each offering different levels of research possibilities [127]. Table 5 presents a summary of commonly used gene fusion detection methods, comparing specificity, sensitivity, and cost based on recent reviews or articles in the field for each marker across various technology applications.

The advent of molecular diagnostic technology in tumor molecular pathology studies has been revolutionary, broadening the range of pathological research. It enables the examination of tumor source, structure, and biological characteristics at the molecular level. Many molecular diagnostic techniques have become increasingly mature.

Accurate gene fusion detection is essential for assisting clinicians in cancer diagnosis, classification, and treatment planning. Current gene fusion diagnostic methods, such as FISH, IHC, and RT-PCR, although sensitive, often lack the resolution and throughput required for comprehensive analysis. These traditional methods heavily rely on tester expertise and may not identify novel fusion partners or address complex structural rearrangements. They typically detect a single gene fusion, resulting in time-consuming, costly, and potentially prone to false-negative results.

By utilizing fluorescence, hybridization, and other detection techniques, the reliance on subjective assessment by technicians is reduced, hence improving the credibility and accuracy of the detection process. Nevertheless, despite its convenience and speed, the NGS-based anchoring amplicon method is prone to false positives. To improve the precision of these methods, it is crucial to incorporate bioinformatics methodologies to exclude gene fusions with low confidence levels.

Due to its cost-effectiveness, heightened sensitivity, broad applicability, and capacity to identify previously unidentified gene fusions, the DNA-targeted sequencing technique has gained extensive use. However, the identification of gene fusions using only DNA-targeted sequencing also has specific limitations: the sensitivity of the NGS technique is directly proportional to the coverage. If all intron length fragments need to be covered, the cost of detection will significantly increase. Furthermore, certain intron regions present challenges for probe design. Therefore, if a fusion breakpoint involves untargeted intronic regions that cannot be covered by the probe, the sensitivity of detection may be compromised.

In contrast to DNA-seq, RNA-targeted sequencing technology offers solutions to several of the aforementioned challenges in gene fusion detection. First, the design of probes for transcripts is not hindered by intron regions, thereby significantly enhancing sensitivity and cost-effectiveness. Second, RNA-Seq provides the additional benefit of assessing the function and expression of gene fusions while simultaneously revealing them, allowing for the secondary validation of gene fusions.

In the gene fusion analysis process, software tools are commonly employed to aid in the analysis. It is important to note that different analysis software may yield varying false positive rates. As a result, several studies have utilized multiple fusion analysis software packages simultaneously to enhance specificity [78]. The selection of a particular technology is not inherently a matter of superiority or inferiority; it is primarily driven by the specific requirements of the detection process. For certain clinical needs, such as the swift identification of known fusions or in situations that involve low-throughput and frequent testing, quantitative reverse transcription PCR (RT-qPCR) has become particularly relevant, especially within the realm of blood diagnostics. Furthermore, the use of FISH technology, especially when integrated with imaging data, is recognized as a convenient and effective method for gene fusion detection.

In contrast, NGS technology is suitable for high-throughput and unknown fusion detection; however, it has relatively lower sensitivity, higher costs, and longer processing cycles. When selecting a technology, it is advisable to balance and choose based on specific needs.

The prospects for AI in the realm of gene fusion research are exceptionally promising, with advancements in DL set to refine the precision of detecting nuanced fusion events amidst the complexity of genomic data. AI is rapidly extending its influence, promising to accurately predict the functionalities and clinical implications of gene fusion products. This progression is ushering in a new era of personalized medicine, where treatments are customized based on the unique genomic signatures of individuals. Moreover, the capacity of AI to amalgamate data across genomics, transcriptomics, proteomics, and metabolomics is offering a more nuanced and comprehensive understanding of

Table 5 Co	ommonly used gei	ne fusion detection methods					
Method	Markers	The fusion type can be detected	Test the sample	Operation time	Fee (RMB)	Strengths	Limitations
HSH	DNA	All fusion types but not distin- guishable	FFPE cytology samples	1–2 d	1500-3000	High sensitivity, excellent stabil- ity, low cost, target position, low requirements for tumor purity, gold standard	The detection throughput is small, the experimental operation is cumbersome, the technical requirements are high, and the fusion partner cannot be identi- fied
IHC	Protein	All fusion types but not distin- guishable	FFPE cell wax block	2–3 d	500-2000	Simpler, economical, fast and sensitive	There is a certain subjectivity in the interpretation, and the judg- ment of weak positive results needs to be further verified by FISH and other techniques, and the specificity is limited
RT-PCR	RNA	Common fusion patterns distin- guish between major subtypes	FFPE cell wax block	2–3 h	500-40000	Short cycle, high sensitivity, strong operability, high speci- ficity	Inability to detect unknown fusion partners and subtypes, high sample quality requirements, and greater influence of RNA quality may result in some false-negative results
NGS	DNA	All fusion types are subclassially distinguishable	FFPE cytology sam- ples, Blood samples	3–10 d	2000–30000	Multiple genes are detected in one experiment, with high throughput and sensitivity, and novel fusion partners can be detected	The experimental data analysis time is long, and the require- ments for biocredit analysis are high, and the cost is high
NGS	RNA	All fusion types are subclassially distinguishable	FFPE cytology samples	3–10 d	2000-30000	Multiple genes are detected in one experiment, with high throughput and sensitivity, and novel fusion partners can be detected	The sample quality requirements are high
ECL	Luminescent reporter mol- ecules	All fusion types, especially those expressed in RNA	Any biological sample where RNA can be extracted	1–2 d	2000-5000	High sensitivity; suitable for a wide range of biomarkers; no need for radioactivity; quantitative	Technical complexity; signal inter- ference; throughput limitations

🖄 Springer

biological mechanisms. The advent of innovative wearable technologies and advancements in biomarker detection are poised to enable real-time tracking of gene fusion activities, significantly improving early diagnosis, treatment monitoring, and the management of disease progression. Fundamentally, AI is on the brink of catalyzing breakthroughs in precision medicine, intricate data integration, and continuous health assessment, transforming both the research landscape of gene fusions and their clinical application.

Continuous advancements in technology are significantly enhancing our ability to perform gene fusion detection and related analyses. With the development of technology, we have gained increasingly comprehensive insight into the genesis, functions, and connections of gene fusions with various diseases. This progress not only boosts the accuracy of gene fusion detection but also amplifies its clinical significance. Consequently, ongoing technological innovations are opening up new vistas for a deeper understanding of gene fusions and broadening their potential clinical applications.

Authors' contributions X.S. wrote the entire article, organized the data and tables, and created the figures. Q.Z. conducted overall revisions to the content of the article and consolidated the figures. D.H. made modifications to the article. P.S. conceptualized the writing approach for the entire article and supervised the manuscript. Each author contributed to the editing and composition of the final version of the manuscript.

Funding This work was supported by the National Key R&D Program of China (No. 2021YFF1200300), the National Natural Science Foundation of China (No.22174094. No.22225402), the Fundamental Research Funds for the Central Universities (YG2023QNA33). Young Leading Scientists Cultivation Plan supported by Shanghai Municipal Education Commission (ZXWH1082101).

Declarations

Competing interests The authors declare no confict of interest.

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

- Wu H, Li X, Li H. Gene fusions and chimeric RNAs, and their implications in cancer. Genes & Diseases. 2019. https://doi.org/ 10.1016/j.gendis.2019.08.002.
- Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. Nat Rev Cancer. 2004. https://doi.org/10.1038/nrc1299.
- Nowell, P. C., Hungerford, D. A. & Nowell, P. C. A minute chromosome in human chronic granulocytic leukemia. *Science*, 1960.
- Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature. 1973. https://doi.org/ 10.1038/243290a0.
- Shtivelman E, Lifshitz B, Gale RP, Canaani E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature. 1985. https://doi.org/10.1038/315550a0.
- Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. Nat Rev Cancer. 2015. https://doi.org/10.1038/nrc3947.
- Dai XF, Theobard R, Cheng HY, Xing MT, Zhang JY. Fusion genes: A promising tool combating against cancer. Biochim Biophys Acta-Rev Cancer. 2018. https://doi.org/10.1016/j.bbcan. 2017.12.003.
- Yun JW, Yang LX, Park HY, Lee CW, Cha HG, Shin HT, Noh KW, Choi YL, Park WY, Park PJ. Dysregulation of cancer genes by recurrent intergenic fusions. Genome Biol. 2020. https://doi. org/10.1186/s13059-020-02076-2.
- Gao, Q. S., Liang, W. W., Foltz, S. M., Mutharasu, G., Jayasinghe, R. G., Cao, S., Liao, W. W., Reynolds, S. M., Wyczalkowski, M. A., Yao, L. J., Yu, L. H., Sun, S. Q., Chen, K., Lazar, A. J., Fields, R. C., Wendl, M. C., Van Tine, B. A., Vij, R., Chen, F., Nykter, M., Shmulevich, I., Ding, L., Fusion Anal Working, G. & Canc Genome Atlas Res, N. Driver Fusions and Their Implications in the Development and Treatment of Human Cancers. *Cell Reports*, 2018. https://doi.org/10.1016/j.celrep.2018.03.050
- Picco G, Chen ED, Alonso LG, Behan FM, Goncalves E, Bignell G, Matchan A, Fu BY, Banerjee R, Andersonl E, Butler A, Benes CH, McDermott U, Dow D, Iorio F, Stronach E, Yang FT, Yusa K, Saez-Rodriguez J, Garnett MJ. Functional linkage of gene fusions to cancer cell fitness assessed by pharmacological and CRISPR-Cas9 screening. Nat Commun. 2019. https://doi.org/10. 1038/s41467-019-09940-1.
- Kumar-Sinha C, Kalyana-Sundaram S, Chinnaiyan AM. Landscape of gene fusions in epithelial cancers: seq and ye shall find. Genome Med. 2015. https://doi.org/10.1186/s13073-015-0252-1.
- 12. Wang L, Yamaguchi S, Burstein MD, Terashima K, Chang K, Ng HK, Nakamura H, He Z, Doddapaneni H, Lewis L, Wang M, Suzuki T, Nishikawa R, Natsume A, Terasaka S, Dauser R, Whitehead W, Adekunle A, Sun J, Qiao Y, Marth G, Muzny DM, Gibbs RA, Leal SM, Wheeler DA, Lau CC. Novel somatic and germline mutations in intracranial germ cell tumours. Nature. 2014. https://doi.org/10.1038/nature13296.
- Lamant, L., Dastugue, N., Pulford, K., Delsol, G. & Mariamé, B. A new fusion gene TPM3-ALK in anaplastic large cell lymphoma created by a (1;2)(q25;p23) translocation. *Blood*, 1999.
- 14. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007. https://doi.org/ 10.1038/nature05945.
- Sánchez-Herrero, E., Provencio, M. & Romero, A. Clinical utility of liquid biopsy for the diagnosis and monitoring of EML4-ALK

NSCLC patients. Advances in Laboratory Medicine / Avances en Medicina de Laboratorio, 2020. https://doi.org/10.1515/ almed-2019-0019

- Mano H. Non-solid oncogenes in solid tumors: EML4-ALK fusion genes in lung cancer. Cancer Sci. 2008. https://doi.org/ 10.1111/j.1349-7006.2008.00972.x.
- 17. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, Hu Y, Tan ZP, Stokes M, Sullivan L, Mitchell J, Wetzel R, MacNeill J, Ren JM, Yuan J, Bakalarski CE, Villen J, Kornhauser JM, Smith B, Li D, Zhou X, Gygi SP, Gu TL, Polakiewicz RD, Rush J, Comb MJ. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell. 2007. https://doi.org/10.1016/j.cell.2007.11.025.
- Takeuchi K, Choi YL, Togashi Y, Soda M, Hatano S, Inamura K, Takada S, Ueno T, Yamashita Y, Satoh Y, Okumura S, Nakagawa K, Ishikawa Y, Mano H. KIF5B-ALK, a Novel Fusion Oncokinase Identified by an Immunohistochemistry-based Diagnostic System for ALK-positive Lung Cancer. Clin Cancer Res. 2009. https://doi.org/10.1158/1078-0432.Ccr-08-3248.
- Togashi Y, Soda M, Sakata S, Sugawara E, Hatano S, Asaka R, Nakajima T, Mano H, Takeuchi K. KLC1-ALK: A Novel Fusion in Lung Cancer Identified Using a Formalin-Fixed Paraffin-Embedded Tissue Only. PLoS One. 2012. https://doi.org/ 10.1371/journal.pone.0031323.
- Jung Y, Kim P, Jung Y, Keum J, Kim SN, Choi YS, Do IG, Lee J, Choi SJ, Kim S, Lee JE, Kim J, Lee S, Kim J. Discovery of ALK-PTPN3 gene fusion from human non-small cell lung carcinoma cell line using next generation RNA sequencing. Gene Chromosomes Cancer. 2012. https://doi.org/10.1002/gcc.21945.
- 21. Tan DSW, Thomas M, Kim DW, Szpakowski S, Urban P, Mehra R, Chow LQM, Sharma S, Solomon BJ, Felip E, Camidge DR, Vansteenkiste J, Petruzzelli L, Pantano S, Shaw AT. Genetic landscape of patients with ALK-rearranged non-small-cell lung cancer (NSCLC) and response to ceritinib in ASCEND-1 study. Lung Cancer. 2022. https://doi.org/10.1016/j.lungcan.2021.11.007.
- Choi YL, Lira ME, Hong M, Kim RN, Choi SJ, Song JY, Pandy K, Mann DL, Stahl JA, Peckham HE, Zheng Z, Han J, Mao M, Kim J. A novel fusion of TPR and ALK in lung adenocarcinoma. J Thorac Oncol. 2014. https://doi.org/10.1097/jto.000000000 000093.
- Li WH, Guo L, Liu YT, Dong L, Yang L, Chen L, Liu KH, Shao Y, Ying JM. Potential Unreliability of Uncommon ALK, ROS1, and RET Genomic Breakpoints in Predicting the Efficacy of Targeted Therapy in NSCLC. J Thorac Oncol. 2021. https://doi.org/ 10.1016/j.jtho.2020.10.156.
- 24. Thunnissen E, Bubendorf L, Dietel M, Elmberger G, Kerr K, Lopez-Rios F, Moch H, Olszewski W, Pauwels P, Penault-Llorca F, Rossi G. EML4-ALK testing in non-small cell carcinomas of the lung: a review with recommendations. Virchows Arch. 2012. https://doi.org/10.1007/s00428-012-1281-4.
- 25. Choi YL, Takeuchi K, Soda M, Inamura K, Togashi Y, Hatano S, Enomoto M, Hamada T, Haruta H, Watanabe H, Kurashina K, Hatanaka H, Ueno T, Takada S, Yamashita Y, Sugiyama Y, Ishikawa Y, Mano H. Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. Cancer Res. 2008. https://doi.org/10.1158/0008-5472.Can-07-6158.
- Koivunen JP, Mermel C, Zejnullahu K, Murphy C, Lifshits E, Holmes AJ, Choi HG, Kim J, Chiang D, Thomas R, Lee J, Richards WG, Sugarbaker DJ, Ducko C, Lindeman N, Marcoux JP, Engelman JA, Gray NS, Lee C, Meyerson M, Jänne PA. EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. Clin Cancer Res. 2008. https://doi.org/10.1158/ 1078-0432.Ccr-08-0168.
- Hasegawa N, Kohsaka S, Kurokawa K, Shinno Y, Nakamura IT, Ueno T, Kojima S, Kawazu M, Suehara Y, Ishijima M, Goto Y,

Kojima Y, Yonemori K, Hayashi T, Saito T, Shukuya T, Takahashi F, Takahashi K, Mano H. Highly sensitive fusion detection using plasma cell-free RNA in non-small-cell lung cancers. Cancer Sci. 2021. https://doi.org/10.1111/cas.15084.

- Westover D, Zugazagoitia J, Cho BC, Lovly CM, Paz-Ares L. Mechanisms of acquired resistance to first- and second-generation EGFR tyrosine kinase inhibitors. Ann Oncol. 2018. https:// doi.org/10.1093/annonc/mdx703.
- 29. Soria JC, Tan DSW, Chiari R, Wu YL, Paz-Ares L, Wolf J, Geater SL, Orlov S, Cortinovis D, Yu CJ, Hochmair M, Cortot AB, Tsai CM, Moro-Sibilot D, Campelo RG, McCulloch T, Sen P, Dugan M, Pantano S, Branle F, Massacesi C, de Castro G. 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. https://doi.org/10.1016/s0140-6736(17)30123-x.
- 30. Hida T, Nokihara H, Kondo M, Kim YH, Azuma K, Seto T, Takiguchi Y, Nishio M, Yoshioka H, Imamura F, Hotta K, Watanabe S, Goto K, Satouchi M, Kozuki T, Shukuya T, Nakagawa K, Mitsudomi T, Yamamoto N, Asakawa T, Asabe R, Tanaka T, Tamura T. Alectinib versus crizotinib in patients with ALK-positive non-small-cell lung cancer (J-ALEX): an open-label, randomised phase 3 trial. Lancet. 2017. https://doi. org/10.1016/s0140-6736(17)30565-2.
- Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim DW, Ou SHI, Perol M, Dziadziuszko R, Rosell R, Zeaiter A, Mitry E, Golding S, Balas B, Noe J, Morcos PN, Mok T, Investigators AT. Alectinib versus Crizotinib in Untreated ALK-Positive Non-Small-Cell Lung Cancer. N Engl J Med. 2017. https://doi. org/10.1056/NEJMoa1704795.
- 32. Shaw AT, Bauer TM, de Marinis F, Felip E, Goto Y, Liu G, Mazieres J, Kim DW, Mok T, Polli A, Thurm H, Calella AM, Peltz G, Solomon BJ, Investigators CT. First-Line Lorlatinib or Crizotinib in Advanced ALK-Positive Lung Cancer. N Engl J Med. 2020. https://doi.org/10.1056/NEJMoa2027187.
- 33. Camidge DR, Kim DW, Tiseo M, Langer CJ, Ahn MJ, Shaw AT, Huber RM, Hochmair MJ, Lee DH, Bazhenova LA, Gold KA, Ou SHI, West HL, Reichmann W, Haney J, Clackson T, Kerstein D, Gettinger SN. Exploratory Analysis of Brigatinib Activity in Patients With Anaplastic Lymphoma Kinase-Positive Non-Small-Cell Lung Cancer and Brain Metastases in Two Clinical Trials. J Clin Oncol. 2018. https://doi.org/10.1200/jco. 2017.77.5841.
- 34. Gainor JF, Dardaei L, Yoda S, Friboulet L, Leshchiner I, Katayama R, Dagogo-Jack I, Gadgeel S, Schultz K, Singh M, Chin E, Parks M, Lee D, DiCecca RH, Lockerman E, Huynh T, Logan J, Ritterhouse LL, Le LP, Muniappan A, Digumarthy S, Channick C, Keyes C, Getz G, Dias-Santagata D, Heist RS, Lennerz J, Sequist LV, Benes CH, Iafrate AJ, Mino-Kenudson M, Engelman JA, Shaw AT. Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer. Cancer Discov. 2016. https://doi.org/ 10.1158/2159-8290.Cd-16-0596.
- 35. Gautschi O, Milia J, Filleron T, Wolf J, Carbone DP, Owen D, Camidge R, Narayanan V, Doebele RC, Besse B, Remon-Masip J, Janne PA, Awad MM, Peled N, Byoung CC, Karp DD, Van Den Heuvel M, Wakelee HA, Neal JW, Mok TSK, Yang JCH, Ou SHI, Pall G, Froesch P, Zalcman G, Gandara DR, Riess J, Velcheti V, Zeidler K, Diebold J, Fruh M, Michels S, Monnet I, Popat S, Rosell R, Karachaliou N, Rothschild SI, Shih JY, Warth A, Muley T, Cabillic F, Mazieres J, Drilon A. Targeting RET in Patients With RET-Rearranged Lung Cancers: Results From the Global. Multicenter RET Registry J Clin Oncol. 2017. https://doi.org/10.1200/jco.2016.70.9352.
- Zhang S, Anjum R, Squillace R, Nadworny S, Zhou T, Keats J, Ning YY, Wardwell SD, Miller D, Song Y, Eichinger L, Moran

L, Huang WS, Liu SY, Zou D, Wang Y, Mohemmad Q, Jang HG, Ye E, Narasimhan N, Wang F, Miret J, Zhu XT, Clackson T, Dalgarno D, Shakespeare WC, Rivera VM. The Potent ALK Inhibitor Brigatinib (AP26113) Overcomes Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in Preclinical Models. Clin Cancer Res. 2016. https://doi.org/ 10.1158/1078-0432.Ccr-16-0569.

- 37. Menichincheri, M., Ardini, E., Magnaghi, P., Avanzi, N., Banfi, P., Bossi, R., Buffa, L., Canevari, G., Ceriani, L., Colombo, M., Corti, L., Donati, D., Fasolini, M., Felder, E., Fiorelli, C., Fiorentini, F., Galvani, A., Isacchi, A., Borgia, A. L., Marchionni, C., Nesi, M., Orrenius, C., Panzeri, A., Pesenti, E., Rusconi, L., Saccardo, M. B., Vanotti, E., Perrone, E. & Orsini, P. Discovery of Entrectinib: A New 3-Aminoindazole as a Potent Anaplastic Lymphoma Kinase (ALK), c-ros Oncogene 1 Kinase (ROS1), and Pan-Tropomyosin Receptor Kinases (Pan-TRKs) inhibitor (vol 59, pg 3392, 2016). J. Med. Chem., 2019. https://doi.org/10.1021/acs.jmedchem.9b01259
- Ardini E, Menichincheri M, Banfi P, Bosotti R, De Ponti C, Pulci R, Ballinari D, Ciomei M, Texido G, Degrassi A, Avanzi N, Amboldi N, Saccardo MB, Casero D, Orsini P, Bandiera T, Mologni L, Anderson D, Wei G, Harris J, Vernier JM, Li G, Felder E, Donati D, Isacchi A, Pesenti E, Magnaghi P, Galvani A. Entrectinib, a Pan-TRK, ROS1, and ALK Inhibitor with Activity in Multiple Molecularly Defined Cancer Indications. Mol Cancer Ther. 2016. https://doi.org/10.1158/1535-7163. Mct-15-0758.
- Lovly CM, Heuckmann JM, de Stanchina E, Chen H, Thomas RK, Liang C, Pao W. Insights into ALK-Driven Cancers Revealed through Development of Novel ALK Tyrosine Kinase Inhibitors. Cancer Res. 2011. https://doi.org/10.1158/0008-5472. Can-10-3879.
- 40. Lovly CM, Infante JR, Blumenschein GR, Reckamp K, Wakelee H, Carter CA, Vvagar SN, Neal J, Gockerman JP, Dukart G, Harrow K, Liang C, Gibbons JJ, Horn L. Phase I/II trial of X-396, a novel anaplastic lymphoma kinase (ALK) inhibitor, in patients with ALK plus non-small cell lung cancer (NSCLC). Cancer Res. 2016. https://doi.org/10.1158/1538-7445.Am2016-ct088.
- 41. Wu YM, Su FY, Kalyana-Sundaram S, Khazanov N, Ateeq B, Cao XH, Lonigro RJ, Vats P, Wang R, Lin SF, Cheng AJ, Kunju LP, Siddiqui J, Tomlins SA, Wyngaard P, Sadis S, Roychowdhury S, Hussain MH, Feng FY, Zalupski MM, Talpaz M, Pienta KJ, Rhodes DR, Robinson DR, Chinnaiyan AM. Identification of Targetable FGFR Gene Fusions in Diverse Cancers. Cancer Discov. 2013. https://doi.org/10.1158/2159-8290.Cd-13-0050.
- 42. Ross JS, Wang K, Chmielecki J, Gay L, Johnson A, Chudnovsky J, Yelensky R, Lipson D, Ali SM, Elvin JA, Vergilio JA, Roels S, Miller VA, Nakamura BN, Gray A, Wong MK, Stephens PJ. The distribution of BRAF gene fusions in solid tumors and response to targeted therapy. Int J Cancer. 2016. https://doi.org/10.1002/ ijc.29825.
- 43. Drilon A, Rekhtman N, Arcila M, Wang L, Ni A, Albano M, Van Voorthuysen M, Somwar R, Smith RS, Montecalvo J, Plodkowski A, Ginsberg MS, Riely GJ, Rudin CM, Ladanyi M, Kris MG. Cabozantinib in patients with advanced RET-rearranged non-small-cell lung cancer: an open-label, single-centre, phase 2, single-arm trial. Lancet Oncol. 2016. https://doi.org/10.1016/ s1470-2045(16)30562-9.
- Arnaoutakis K. Crizotinib in ROS1-Rearranged Non-Small-Cell Lung Cancer. N Engl J Med. 2015. https://doi.org/10.1056/ NEJMc1415359.
- 45. Arai Y, Totoki Y, Hosoda F, Shirota T, Hama N, Nakamura H, Ojima H, Furuta K, Shimada K, Okusaka T, Kosuge T, Shibata T. Fibroblast Growth Factor Receptor 2 Tyrosine Kinase Fusions Define a Unique Molecular Subtype of Cholangiocarcinoma. Hepatology. 2014. https://doi.org/10.1002/hep.26890.

- 46. Ross JS, Wang K, Gay L, Al-Rohil R, Rand JV, Jones DM, Lee HJ, Sheehan CE, Otto GA, Palmer G, Yelensky R, Lipson D, Morosini D, Hawryluk M, Catenacci DVT, Miller VA, Churi C, Ali S, Stephens PJ. New Routes to Targeted Therapy of Intrahepatic Cholangiocarcinomas Revealed by Next-Generation Sequencing. Oncologist. 2014. https://doi.org/10.1634/theon cologist.2013-0352.
- Farshidfar, F., Zheng, S. Y., Gingras, M. C., Newton, Y., Shih, J., Robertson, A. G., Hinoue, T., Hoadley, K. A., Gibb, E. A., Roszik, J., Covington, K. R., Wu, C. C., Shinbrot, E., Stransky, N., Hegde, A., Yang, J. D., Reznik, E., Sadeghi, S., Pedamallu, C. S., Ojesina, A. I., Hess, J. M., Auman, J. T., Rhie, S. K., Bowlby, R., Borad, M. J., Zhu, A. X., Stuart, J. M., Sander, C., Akbani, R., Cherniack, A. D., Deshpande, V., Mounajjed, T., Foo, W. C., Torbenson, M. S., Kleiner, D. E., Laird, P. W., Wheeler, D. A., McRee, A. J., Bathe, O. F., Andersen, J. B., Bardeesy, N., Roberts, L. R., Kwong, L. N. & Canc Genome Atlas, N. Integrative Genomic Analysis of Cholangiocarcinoma Identifies Distinct IDH-Mutant Molecular Profiles (vol 18, pg 2780, 2017). *Cell Reports*, 2017. https://doi.org/10.1016/j.celrep.2017.06.008
- Jain A, Borad MJ, Kelley RK, Wang Y, Abdel-Wahab R, Meric-Bernstam F, Baggerly KA, Kaseb AO, Al-Shamsi HO, Ahn DH, DeLeon T, Bocobo AG, Bekaii-Saab T, Shroff RT, Javle M. Cholangiocarcinoma With FGFR Genetic Aberrations: A Unique Clinical Phenotype. JCO Precis Oncol. 2018. https://doi.org/10. 1200/po.17.00080.
- 49. Abou-Alfa GK, Sahai V, Hollebecque A, Vaccaro G, Melisi D, Al-Rajabi R, Paulson AS, Borad MJ, Gallinson D, Murphy AG, Oh DY, Dotan E, Catenacci DV, Van Cutsem E, Ji T, Lihou CF, Zhen HL, Feliz L, Vogel A. Pemigatinib for previously treated, locally advanced or metastatic cholangiocarcinoma: a multicentre, open-label, phase 2 study. Lancet Oncol. 2020. https://doi. org/10.1016/s1470-2045(20)30109-1.
- 50. Sia D, Losic B, Moeini A, Cabellos L, Hao K, Revill K, Bonal D, Miltiadous O, Zhang ZY, Hoshida Y, Cornella H, Castillo-Martin M, Pinyol R, Kasai Y, Roayaie S, Thung SN, Fuster J, Schwartz ME, Waxman S, Cordon-Cardo C, Schadt E, Mazzaferro V, Llovet JM. Massive parallel sequencing uncovers actionable FGFR2-PPHLN1 fusion and ARAF mutations in intrahepatic cholangiocarcinoma. Nat Commun. 2015. https://doi.org/10. 1038/ncomms7087.
- Hoy S. M. Pemigatinib: First Approval. Drugs. 2020. https://doi. org/10.1007/s40265-020-01330-y.
- Kang C. Infigratinib: First Approval. Drugs. 2021. https://doi. org/10.1007/s40265-021-01567-1.
- 53. Casak SJ, Pradhan S, Fashoyin-Aje LA, Ren Y, Shen Y-L, Xu Y, Chow ECY, Xiong Y, Zirklelbach JF, Liu J, Charlab R, Pierce WF, Fesenko N, Beaver JA, Pazdur R, Kluetz PG, Lemery SJ. FDA Approval Summary: Ivosidenib for the Treatment of Patients with Advanced Unresectable or Metastatic, Chemotherapy Refractory Cholangiocarcinoma with an IDH1 Mutation. Clin Cancer Res. 2022. https://doi.org/10.1158/1078-0432. Ccr-21-4462.
- 54. Gandhy SU, Casak SJ, Mushti SL, Cheng J, Subramaniam S, Zhao H, Zhao M, Bi Y, Liu G, Fan J, Adeniyi O, Charlab R, Kufrin D, Thompson MD, Jarrell K, Auth D, Lemery SJ, Pazdur R, Kluetz PG, Fashoyin-Aje LA. FDA Approval Summary: Futibatinib for Unresectable Advanced or Metastatic, Chemotherapy Refractory Intrahepatic Cholangiocarcinoma with FGFR2 Fusions or Other Rearrangements. Clin Cancer Res. 2023. https://doi.org/10.1158/1078-0432.Ccr-23-1042.
- 55. Abou-Alfa GK, Sahai V, Hollebecque A, Vaccaro GM, Melisi D, Al-Rajabi RMT, Paulson AS, Borad MJ, Gallinson DH, Murphy AG, Oh DY, Dotan E, Catenacci D, Van Cutsem E, Lihou CF, Zhen HL, Feliz L, Vogel A. Pemigatinib for previously treated locally advanced/metastatic cholangiocarcinoma (CCA): Update

of FIGHT-202. J Clin Oncol. 2021. https://doi.org/10.1200/JCO. 2021.39.15_suppl.4086.

- 56. Javle M, Lowery M, Shroff RT, Weiss KH, Springfeld C, Borad MJ, Ramanathan RK, Goyal L, Sadeghi S, Macarulla T, El-Khoueiry A, Kelley RK, Borbath I, Choo SP, Oh DY, Philip PA, Chen LT, Reungwetwattana T, Van Cutsem E, Yeh KH, Ciombor K, Finn RS, Patel A, Sen S, Porter D, Isaacs R, Zhu AX, Abou-Alfa GK, Bekaii-Saab T. Phase II Study of BGJ398 in Patients With FGFR-Altered Advanced Cholangiocarcinoma. J Clin Oncol. 2018. https://doi.org/10.1200/jco.2017.75.5009.
- Leslie M. Ivosidenib Boosts OS in Cholangiocarcinoma. Cancer Discov. 2021. https://doi.org/10.1158/2159-8290. Cd-nb2021-0389.
- Meric-Bernstam F, Bahleda R, Hierro C, Sanson M, Bridgewater J, Arkenau HT, Tran B, Kelley RK, Park JO, Javle M, He Y, Benhadji KA, Goyal L. Futibatinib, an Irreversible FGFR1-4 Inhibitor, in Patients with Advanced Solid Tumors Harboring FGF/ FGFR Aberrations: A Phase I Dose-Expansion Study. Cancer Discov. 2022. https://doi.org/10.1158/2159-8290.Cd-21-0697.
- 59. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, Menon A, Jing X, Cao Q, Han B, Yu J, Wang L, Montie JE, Rubin MA, Pienta KJ, Roulston D, Shah RB, Varambally S, Mehra R, Chinnaiyan AM. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature. 2007. https://doi.org/10.1038/nature06024.
- Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. Recurrent gene fusions in prostate cancer. Nat Rev Cancer. 2008. https://doi.org/ 10.1038/nrc2402.
- Kalvala A, Gao L, Aguila B, Dotts K, Rahman M, Nana-Sinkam SP, Zhou X, Wang QE, Amann J, Otterson GA, Villalona-Calero MA, Duan W. Rad51C-ATXN7 fusion gene expression in colorectal tumors. Mol Cancer. 2016. https://doi.org/10.1186/ s12943-016-0527-1.
- Annala MJ, Parker BC, Zhang W, Nykter M. Fusion genes and their discovery using high throughput sequencing. Cancer Lett. 2013. https://doi.org/10.1016/j.canlet.2013.01.011.
- 63. Chung CB, Lee J, Barritault M, Bringuier PP, Xu ZL, Huang WY, Beharry A, Castillo J, Christiansen J, Lin JC, Sheffield BS. Evaluating Targeted Next-Generation Sequencing Assays and Reference Materials for NTRK Fusion Detection. J Mol Diagn. 2022. https://doi.org/10.1016/j.jmoldx.2021.09.008.
- 64. Jin Z, Huang W, Shen N, Li J, Wang X, Dong J, Park PJ, Xi R. Single-cell gene fusion detection by scFusion. Nat Commun. 2022. https://doi.org/10.1038/s41467-022-28661-6.
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009. https://doi.org/10. 1038/nrg2484.
- Bruno R, Fontanini G. Next Generation Sequencing for Gene Fusion Analysis in Lung Cancer: A Literature Review. Diagnostics. 2020. https://doi.org/10.3390/diagnostics10080521.
- 67. Marino FZ, Pagliuca F, Ronchi A, Cozzolino I, Montella M, Berretta M, Errico ME, Donofrio V, Bianco R, Franco R. NTRK Fusions, from the Diagnostic Algorithm to Innovative Treatment in the Era of Precision Medicine. Int J Mol Sci. 2020. https://doi. org/10.3390/ijms21103718.
- Nookaew I, Papini M, Pornputtapong N, Scalcinati G, Fagerberg L, Uhlen M, Nielsen J. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. Nucleic Acids Res. 2012. https:// doi.org/10.1093/nar/gks804.
- 69. Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, Lee J, Jung YJ, Kim JO, Shin JY, Yu SB, Kim J, Lee ER, Kang CH, Park IK, Rhee H, Lee SH, Kim JI, Kang JH, Kim YT. The transcriptional

landscape and mutational profile of lung adenocarcinoma. Genome Res. 2012. https://doi.org/10.1101/gr.145144.112.

- Beaubier N, Bontrager M, Huether R, Igartua C, Lau D, Tell R, Bobe AM, Bush S, Chang AL, Hoskinson DC, Khan AA, Kudalkar E, Leibowitz BD, Lozachmeur A, Michuda J, Parsons J, Perera JF, Salahudeen A, Shah KP, Taxter T, Zhu W, White KP. Integrated genomic profiling expands clinical options for patients with cancer. Nat Biotechnol. 2019. https://doi.org/10. 1038/s41587-019-0259-z.
- Sheikine Y, Kuo FC, Lindeman NI. Clinical and Technical Aspects of Genomic Diagnostics for Precision Oncology. J Clin Oncol. 2017. https://doi.org/10.1200/jco.2016.70.7539.
- Davies KD, Aisner DL. Wake Up and Smell the Fusions: Single-Modality Molecular Testing Misses Drivers. Clin Cancer Res. 2019. https://doi.org/10.1158/1078-0432.Ccr-19-1361.
- Teixido C, Gimenez-Capitan A, Molina-Vila MA, Peg V, Karachaliou N, Rodriguez-Capote A, Castellvi J, Rosell R. RNA Analysis as a Tool to Determine Clinically Relevant Gene Fusions and Splice Variants. Arch Pathol Lab Med. 2018. https:// doi.org/10.5858/arpa.2017-0134-RA.
- Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet. 2019. https://doi.org/10.1038/ s41576-019-0150-2.
- Cheng, W., Lin, Z., Zhao, L., Fan, N., Bai, H., Cheng, W., Zhao, M. & Ding, S. CeO2/MXene heterojunction-based ultrasensitive electrochemiluminescence biosensing for BCR-ABL fusion gene detection combined with dual-toehold strand displacement reaction for signal amplification. *Biosensors and Bioelectronics*, 2022. https://doi.org/10.1016/j.bios.2022.114287
- 76. Wu, Y., Guo, J., Li, W., Xiu, X., Thirunavukarasu, D., Wang, Y., Wang, K., Chen, W., Zhang, D. y., Yang, X., Fan, C. & Song, P. Enhanced Detection of Novel Low-Frequency Gene Fusions via High-Yield Ligation and Multiplexed Enrichment Sequencing. *Angewandte Chemie International Edition*, https://doi.org/10. 1002/anie.202316484
- Heyer EE, Deveson IW, Wooi D, Selinger CI, Lyons RJ, Hayes VM, O'Toole SA, Ballinger ML, Gill D, Thomas DM, Mercer TR, Blackburn J. Diagnosis of fusion genes using targeted RNA sequencing. Nat Commun. 2019. https://doi.org/10.1038/ s41467-019-09374-9.
- Harrison CJ, Secker-Walker LM. The importance of cytogenetics and associated molecular techniques in the management of patients with leukaemia. Clin Oncol (R Coll Radiol). 1998. https://doi.org/10.1016/s0936-6555(98)80013-7.
- 79. Zito Marino F, Buono S, Montella M, Giannatiempo R, Messina F, Casaretta G, Arpino G, Vita G, Fiorentino F, Insabato L, Sgambato A, Orditura M, Franco R, Accardo M. NTRK gene aberrations in triple-negative breast cancer: detection challenges using IHC, FISH, RT-PCR, and NGS. J Pathol Clin Res. 2023. https://doi.org/10.1002/cjp2.324.
- 80. van Dongen, J. J., Macintyre, E. A., Gabert, J. A., Delabesse, E., Rossi, V., Saglio, G., Gottardi, E., Rambaldi, A., Dotti, G., Griesinger, F., Parreira, A., Gameiro, P., Diáz, M. G., Malec, M., Langerak, A. W., San Miguel, J. F. & Biondi, A. Standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease. Report of the BIOMED-1 Concerted Action: investigation of minimal residual disease in acute leukemia. *Leukemia*, 1999. https://doi.org/10.1038/sj.leu.2401592
- Pallisgaard, N., Hokland, P., Riishøj, D. C., Pedersen, B. & Jørgensen, P. Multiplex reverse transcription-polymerase chain reaction for simultaneous screening of 29 translocations and chromosomal aberrations in acute leukemia. *Blood*, 1998.
- Clavé S, Jackson JB, Salido M, Kames J, Gerding KMR, Verner EL, Kong EF, Weingartner E, Gibert J, Hardy-Werbin M, Rocha P, Riera X, Torres E, Hernandez J, Cerqueira G, Nichol D,

Simmons J, Taus Á, Pijuan L, Bellosillo B, Arriola E. Comprehensive NGS profiling to enable detection of ALK gene rearrangements and MET amplifications in non-small cell lung cancer. Front Oncol. 2023. https://doi.org/10.3389/fonc.2023.12256 46.

- 83. Kalemkerian GP, Narula N, Kennedy EB, Biermann WA, Donington J, Leighl NB, Lew M, Pantelas J, Ramalingam SS, Reck M, Saqi A, Simoff M, Singh N, Sundaram B. Molecular Testing Guideline for the Selection of Patients With Lung Cancer for Treatment With Targeted Tyrosine Kinase Inhibitors: American Society of Clinical Oncology Endorsement of the College of American Pathologists/International Association for the Study of Lung Cancer/Association for Molecular Pathology Clinical Practice Guideline Update. J Clin Oncol. 2018. https://doi.org/ 10.1200/jco.2017.76.7293.
- Kozuma Y, Toyokawa G, Seto T. ALK testing methods: is there a winner or loser? Expert Rev Anticancer Ther. 2019. https://doi. org/10.1080/14737140.2019.1562343.
- Gocke CD, Mason J, Brusca L, Laosinchai-Wolf W, Higgs C, Newell H, Masters A, Friar L, Karp J, Griffiths M, Wei Q, Labourier E. Risk-based classification of leukemia by cytogenetic and multiplex molecular methods: results from a multicenter validation study. Blood Cancer J. 2012. https://doi.org/10.1038/bcj. 2012.24.
- Qu X, Yeung C, Coleman I, Nelson PS, Fang M. Comparison of four next generation sequencing platforms for fusion detection: Oncomine by ThermoFisher, AmpliSeq by illumina, FusionPlex by ArcherDX, and QIAseq by QIAGEN. Cancer Genet. 2020. https://doi.org/10.1016/j.cancergen.2020.02.007.
- Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA, Mattick JS, Rinn JL. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat Biotechnol. 2012. https://doi.org/10.1038/nbt.2024.
- Wada Y, Matsuura M, Sugawara M, Ushijima M, Miyata S, Nagasaki K, Noda T, Miki Y. Development of detection method for novel fusion gene using GeneChip exon array. Journal of clinical bioinformatics. 2014. https://doi.org/10.1186/2043-9113-4-3.
- Heyer EE, Blackburn J. Sequencing Strategies for Fusion Gene Detection. Bioessays. 2020. https://doi.org/10.1002/bies.20200 0016.
- Peng Q, Xu C, Kim D, Lewis M, DiCarlo J, Wang YX. Targeted Single Primer Enrichment Sequencing with Single End Duplex-UMI. Sci Rep. 2019. https://doi.org/10.1038/ s41598-019-41215-z.
- Mercer TR, Clark MB, Crawford J, Brunck ME, Gerhardt DJ, Taft RJ, Nielsen LK, Dinger ME, Mattick JS. Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat Protoc. 2014. https://doi.org/10.1038/nprot.2014.058.
- 92. Zheng ZL, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, Chen JX, Robinson HE, Shim HS, Chmielecki J, Pao W, Engelman JA, Iafrate AJ, Le LP. Anchored multiplex FOR for targeted next-generation sequencing. Nat Med. 2014. https://doi.org/10.1038/nm.3729.
- Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biol. 2019. https://doi.org/10.1186/s13059-019-1842-9.
- 94. Griffioen, M., Arindrarto, W., Borras, D., van Diessen, S., van der Meijden, E. D., Honders, M. W., Alloul, M., Jedema, I., Kroes, W. G. M., Valk, P. J. M., Janssen, B., van Bergen, C. A. M., Kielbasa, S. M., Laros, J. F. J., t Hoen, P. A. C. & Veelken, H. Whole Transcriptome Sequencing (RNAseq) As a Comprehensive, Cost-Efficient Diagnostic Tool for Acute Myeloid Leukemia. *Blood*, 2016. https://doi.org/10.1182/blood.V128.22.1701. 1701

- Wei L, Zhang Y, Eziz N, Yang Y, Li G, Guan M. An ultrasensitive electrochemiluminescence immunosensor for alpha-fetoprotein based on a poly(aniline-luminol)/graphene oxide nanocomposite. Anal Bioanal Chem. 2019. https://doi.org/10.1007/ s00216-019-01897-w.
- 96. Lv H, Chen A, Cheng W, Kong L, Zhao M, Ding S, Ju H, Cheng W. Efficient DNA Walker Guided with Well-Regulated Interfacial Tracks for Ultrasensitive Electrochemiluminescence Biosensing. Anal Chem. 2020. https://doi.org/10.1021/acs. analchem.0c03893.
- 97. Wei, W., Lin, H., Hao, T., Su, X., Jiang, X., Wang, S., Hu, Y. & Guo, Z. Dual-mode ECL/SERS immunoassay for ultrasensitive determination of Vibrio vulnificus based on multifunctional MXene. Sensors and Actuators B: Chemical, 2021. https://doi. org/10.1016/j.snb.2021.129525
- Engstrom, P. G., Steijger, T., Sipos, B., Grant, G. R., Kahles, A., Ratsch, G., Goldman, N., Hubbard, T. J., Harrow, J., Guigo, R., Bertone, P. & Consortium, R. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods. 2013. https://doi.org/10.1038/nmeth.2722.
- 99. Steijger, T., Abril, J. F., Engstrom, P. G., Kokocinski, F., Hubbard, T. J., Guigo, R., Harrow, J., Bertone, P. & Consortium, R. Assessment of transcript reconstruction methods for RNA-seq. Nat Methods. 2013. https://doi.org/10.1038/nmeth.2714.
- 100. Liu S, Tsai WH, Ding Y, Chen R, Fang Z, Huo ZG, Kim S, Ma TZ, Chang TY, Priedigkeit NM, Lee AV, Luo JH, Wang HW, Chung IF, Tseng GC. Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. Nucleic Acids Res. 2016. https://doi.org/10.1093/nar/gkv1234.
- 101. Rapaport F, Khanin R, Liang YP, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND, Betel D. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data (vol 14, R95, 2013). Genome Biol. 2015. https://doi.org/ 10.1186/s13059-015-0813-z.
- 102. Su ZQ, Labaj PP, Li S, Thierry-Mieg J, Thierry-Mieg D, Shi W, Wang C, Schroth GP, Setterquist RA, Thompson JF, Jones WD, Xiao WH, Xu WH, Jensen RV, Kelly R, Xu J, Conesa A, Furlanello C, Gao HL, Hong HX, Jafari N, Letovsky S, Liao Y, Lu F, Oakeley EJ, Peng ZY, Praul CA, Santoyo-Lopez J, Scherer A, Shi T, Smyth GK, Staedtler F, Sykacek P, Tan XX, Thompson EA, Vandesompele J, Wang MD, Wang J, Wolfinger RD, Zavadil J, Auerbach SS, Bao WJ, Binder H, Blomquist T, Brilliant MH, Bushel PR, Cain WM, Catalano JG, Chang CW, Chen T, Chen G, Chen R, Chierici M, Chu TM, Clevert DA, Deng YP, Derti A, Devanarayan V, Dong ZR, Dopazo J, Du TT, Fang H, Fang YX, Fasold M, Fernandez A, Fischer M, Furio-Tari P, Fuscoe JC, Caiment F, Gaj S, Gandara J, Gao H, Ge WG, Gondo Y, Gong BS, Gong MH, Gong ZL, Green B, Guo C, Guo L, Guo LW, Hadfield J, Hellemans J, Hochreiter S, Jia MW, Jian M, Johnson CD, Kay S, Kleinjans J, Lababidi S, Levy S, Li QZ, Li L, Li L, Li P, Li Y, Li HQ, Li JY, Li SY, Lin SM, Lopez FJ, Lu X, Luo H, Ma XW, Meehan J, Megherbi DB, Mei N, Mu B, Ning BT, Pandey A, Perez-Florido J, Perkins RG, Peters R, Phan JH, Pirooznia M, Qian F, Qing T, Rainbow L, Rocca-Serra P, Sambourg L, Sansone SA, Schwartz S, Shah R, Shen J, Smith TM, Stegle O, Stralis-Pavese N, Stupka E, Suzuki Y, Szkotnicki LT, Tinning M, Tu BM, van Deft J, Vela-Boza A, Venturini E, Walker SJ, Wan LQ, Wang W, Wang JH, Wang J, Wieben ED, Willey JC, Wu PY, Xuan J, Yang Y, Ye Z, Yin Y, Yu Y, Yuan YC, Zhang J, Zhang KK, Zhang WQ, Zhang WW, Zhang YY, Zhao C, Zheng YT, Zhou YM, Zumbo P, Tong WD, Kreil DP, Mason CE, Shi LM. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol. 2014. https://doi.org/10.1038/nbt.2957.

- 103. Hayer KE, Pizarro A, Lahens NF, Hogenesch JB, Grant GR. Benchmark analysis of algorithms for determining and quantifying full-length mRNA splice forms from RNA-seq data. Bioinformatics. 2015. https://doi.org/10.1093/bioinformatics/ btv488.
- Fonseca NA, Marioni J, Brazma A. RNA-Seq gene profiling–a systematic empirical comparison. PLoS One. 2014. https://doi. org/10.1371/journal.pone.0107026.
- 105. Haas, B. J., Dobin, A., Stransky, N., Li, B., Yang, X., Tickle, T., Bankapur, A., Ganote, C., Doak, T. G., Pochet, N., Sun, J., Wu, C. J., Gingeras, T. R. & Regev, A. STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv*, 2017. https://doi.org/10.1101/120295
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009. https://doi.org/10.1186/ gb-2009-10-3-r25.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012. https://doi.org/10.1038/nmeth. 1923.
- 108. Kent WJ. BLAT-the BLAST-like alignment tool. Genome Res. 2002. https://doi.org/10.1101/gr.229202.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013. https://doi.org/10.1093/ bioinformatics/bts635.
- 110. Nicorici, D., Şatalan, M., Edgren, H., Kangaspeska, S., Murumägi, A., Kallioniemi, O., Virtanen, S. & Kilkku, O. FusionCatcher – a tool for finding somatic fusion genes in pairedend RNA-sequencing data. *bioRxiv*, 2014. https://doi.org/10. 1101/011650
- 111. Uhrig S, Ellermann J, Walther T, Burkhardt P, Froehlich M, Hutter B, Toprak UH, Neumann O, Stenzinger A, Scholl C, Froehling S, Brors B. Accurate and efficient detection of gene fusions from RNA sequencing data. Genome Res. 2021. https://doi.org/ 10.1101/gr.257246.119.
- 112. Vicente-Garcés C, Maynou J, Fernández G, Esperanza-Cebollada E, Torrebadell M, Català A, Rives S, Camós M, Vega-García N. Fusion InPipe, an integrative pipeline for gene fusion detection from RNA-seq data in acute pediatric leukemia. Front Mol Biosci. 2023. https://doi.org/10.3389/fmolb.2023.1141310.
- 113. Vu TN, Deng WJ, Trac QT, Calza S, Hwang W, Pawitan Y. A fast detection of fusion genes from paired-end RNA-seq data. BMC Genomics. 2018. https://doi.org/10.1186/s12864-018-5156-1.
- Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. Nat Commun. 2014. https:// doi.org/10.1038/ncomms5846.
- 115. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, Gil L, García-Girón C, Gordon L, Hourlier T, Hunt S, Juettemann T, Kähäri AK, Keenan S, Komorowska M, Kulesha E, Longden I, Maurel T, McLaren WM, Muffato M, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E, Riat HS, Ritchie GR, Ruffier M, Schuster M, Sheppard D, Sobral D, Taylor K, Thormann A, Trevanion S, White S, Wilder SP, Aken BL, Birney E, Cunningham F, Dunham I, Harrow J, Herrero J, Hubbard TJ,

Johnson N, Kinsella R, Parker A, Spudich G, Yates A, Zadissa A, Searle SM. Ensembl 2013. Nucleic Acids Res. 2013. https://doi.org/10.1093/nar/gks1236.

- 116. Honeyman JN, Simon EP, Robine N, Chiaroni-Clarke R, Darcy DG, Lim IIP, Gleason CE, Murphy JM, Rosenberg BR, Teegan L, Takacs CN, Botero S, Belote R, Germer S, Emde AK, Vacic V, Bhanot U, LaQuaglia MP, Simon SM. Detection of a Recurrent DNAJB1-PRKACA Chimeric Transcript in Fibrolamellar Hepatocellular Carcinoma. Science. 2014. https://doi.org/10. 1126/science.1249484.
- 117. Uhrig S, Ellermann J, Walther T, Burkhardt P, Fröhlich M, Hutter B, Toprak UH, Neumann O, Stenzinger A, Scholl C, Fröhling S, Brors B. Accurate and efficient detection of gene fusions from RNA sequencing data. Genome Res. 2021. https://doi.org/10. 1101/gr.257246.119.
- Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015. https://doi.org/10.1126/scien ce.aaa8415.
- 119. Bengio Y, Lecun Y, Hinton G. Deep learning for AI. Commun ACM. 2021. https://doi.org/10.1145/3448250.
- Patten T, Jacobs P. Natural-language processing. IEEE Expert. 1994. https://doi.org/10.1109/64.295134.
- Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep Learning for Computer Vision: A Brief Review. Comput Intell Neurosci. 2018. https://doi.org/10.1155/2018/7068349.
- 122. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S. & Lew, M. S. Deep learning for visual understanding: A review. *Neurocomputing*, 2016. https://doi.org/10.1016/j.neucom.2015.09.116
- 123. Zhang JX, Yordanov B, Gaunt A, Wang MX, Dai P, Chen Y-J, Zhang K, Fang JZ, Dalchau N, Li J, Phillips A, Zhang DY. A deep learning model for predicting next-generation sequencing depth from DNA sequence. Nat Commun. 2021. https://doi.org/ 10.1038/s41467-021-24497-8.
- Lovino M, Urgese G, Macii E, Di Cataldo S, Ficarra E. A Deep Learning Approach to the Screening of Oncogenic Gene Fusions in Humans. Int J Mol Sci. 2019. https://doi.org/10.3390/ijms2 0071645.
- 125. Kim, P., Tan, H., Liu, J., Yang, M. & Zhou, X. FusionAI: Predicting fusion breakpoint from DNA sequence with deep learning. *iScience*, 2021. https://doi.org/10.1016/j.isci.2021.103164
- 126. Dadhania V, Gonzalez D, Yousif M, Cheng J, Morgan TM, Spratt DE, Reichert ZR, Mannan R, Wang X, Chinnaiyan A, Cao X, Dhanasekaran SM, Chinnaiyan AM, Pantanowitz L, Mehra R. Leveraging artificial intelligence to predict ERG gene fusion status in prostate cancer. BMC Cancer. 2022. https://doi.org/10. 1186/s12885-022-09559-4.
- 127. Wang QG, Xia JF, Jia PL, Pao W, Zhao ZM. Application of next generation sequencing to human gene fusion detection: computational tools, features and perspectives. Brief Bioinform. 2013. https://doi.org/10.1093/bib/bbs044.

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