Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Reproducibility Assessment of Fusion Gene Candidates Identified by Long-Read Transcriptome Sequencing

Aye Chan Pyae
The University of Tokyo
Sayaka Arimatsu
The University of Tokyo
Hiroyuki Kato
The University of Tokyo
Keisuke Yamamoto
The University of Tokyo
Ko Ikemoto
The University of Tokyo
Motoko Unoki
The University of Tokyo
Keisuke Tateishi
The University of Tokyo
Akihiro Fujimoto
afujimoto@m.u-tokyo.ac.jp

The University of Tokyo

Research Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11440207/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Reproducibility Assessment of Fusion Gene Candidates Identified by Long-Read Transcriptome Sequencing

Aye Chan Pyae[1], Sayaka Arimatsu[1], Hiroyuki Kato[2], Keisuke Yamamoto[2], Ko Ikemoto[1], Motoko Unoki[1], Keisuke Tateishi[2,3] and Akihiro Fujimoto[1]

1; Department of Human Genetics, School of International Health, Graduate School of Medicine, The University of Tokyo, Tokyo, 113-0033, Japan

2; Department of Gastroenterology, The University of Tokyo, Tokyo, 113-0033, Japan

3; Department of Gastroenterology, St Marianna University School of Medicine, Kanagawa, 216-8511, Japan

*Correspondence: afujimoto@m.u-tokyo.ac.jp

Abstract

Fusion genes are important therapeutic targets and biomarkers in precision oncology; however, their accurate detection from transcriptome sequencing data remains challenging. Long-read transcriptome sequencing, including Nanopore sequencing, has the potential to improve fusion detection by resolving transcript structures, but the reliability of computationally predicted fusion events has not been systematically evaluated. Here, we performed long-read transcriptome sequencing of 17 pancreatic organoid samples and identified fusion candidates using four fusion gene detection tools. A large number of candidates (3,643–14,698 per sample) were identified, raising concerns about false-positive predictions. To assess the reproducibility of these candidates, we experimentally validated 45 selected fusions by PCR using cDNAs independently synthesized from the same RNA samples by three different reverse-transcription methods. Only five candidates were reproducibly detected in all three independently synthesized cDNAs, whereas the remaining candidates showed inconsistent detection. Sequence-level analysis revealed that short homologous sequences, proximity to canonical splice sites, and repeat elements were associated with non-reproducible fusion predictions. These findings indicate that computational fusion detection from long-read transcriptome data can generate a substantial number of non-reproducible candidates and that validation using a single cDNA preparation may be insufficient. We propose independent cDNA synthesis as an important component of experimental fusion validation and provide sequence features that may help prioritize candidates for validation.

Introduction

Fusion genes are caused by genomic alterations in cancer that can contribute to tumor development and serve as therapeutic targets and biomarkers in precision oncology (1–6). Well-characterized examples include *BCR-ABL1*, *EML4-ALK*, and *NTRK* fusions, for which targeted therapies have demonstrated substantial clinical benefit (7–10).

To detect fusion genes, transcriptome sequencing methods is widely used. Transcriptome sequencing can directly detect expressed fusion transcripts resulting from genomic rearrangements (11,12). RNA sequencing (RNA-seq) using short-read platforms is a commonly used approach, and many fusion genes have been identified with this method (13–15). However, short-read sequencing has notable limitations because of its short read length. In particular, fusion transcripts may be missed when breakpoints occur in repetitive or homologous regions or when the fusion structure cannot be resolved by short reads (16–18).

Long-read sequencing technologies have the potential to overcome these limitations. Because long reads can capture full-length transcripts, they provide greater structural information and have the potential to improve fusion transcript detection (19–21). Several studies have demonstrated that long-read transcriptome sequencing can identify fusion transcripts that are difficult to resolve using short-read sequencing, including virus–human fusion transcripts and novel gene fusions (19,22–24). These studies highlight the potential of long-read sequencing for comprehensive fusion transcript discovery. Several fusion gene detection tools, including FusionSeeker, LongGF, JAFFAL, and SPLICE, have been developed to identify fusion genes from long-read RNA-seq data (19,25–27).

However, the accuracy and reproducibility of fusion gene candidates identified from long-read RNA-seq data remain insufficiently characterized, and standardized strategies for their experimental validation have not been established. A recent study highlighted limitations of current computational approaches by comparing three fusion gene detection tools using previously published long-read sequencing datasets from two cell lines (28). In that benchmark, false-positive detections outnumbered true-positive detections based on known fusion genes. These findings suggest that computational prediction alone may not be sufficient to reliably identify fusion transcripts and underscore the need for systematic experimental validation.

In this study, we performed long-read transcriptome sequencing using Oxford Nanopore Technologies (ONT) to detect fusion gene transcripts in three-dimensional pancreatic cancer organoid samples. We systematically evaluated the reproducibility of

computationally predicted fusion gene candidates by PCR using cDNAs independently synthesized from the same RNA samples by multiple reverse-transcription methods. We then examined sequence and mapping characteristics associated with non-reproducible fusion gene candidates. Our findings highlight the importance for assessing fusion candidates identified by long-read transcriptome sequencing.

Materials and Methods

Samples

Total RNA was extracted from organoids established from 11 pancreatic ductal adenocarcinomas (PDAC), three intraductal papillary mucinous neoplasms (IPMNs), two IPMN-derived carcinomas and one normal pancreas (n = 17) (29) (Table S1). RNA quality was assessed using RNA Integrity Number (RIN) values measured with Tapestation (Agilent), and all samples had RIN values greater than 8.0 (Table S1).

Library preparation and cDNA sequencing

For cDNA synthesis, 250 ng of RNA was used for each sample (Table S2). First-strand cDNA synthesis was performed using a SMARTer cDNA Synthesis Kit (Takara Bio) according to the manufacturer's instruction. Eighteen cycles of PCR were performed using an Advantage 2 PCR Kit (Takara Bio) to amplify double-stranded cDNA. The cDNA was then purified using Agencourt AMPure XP magnetic beads (Beckman Coulter). Libraries were prepared using an SQK-LSK110 Kit (ONT) and sequenced on a MinION Flow Cell (R9.4.1) (ONT) using a MinION sequencer (ONT). Basecalling was performed using Guppy software (version 6.3.7). Adapter trimming was carried out using Porechop (version 0.2.4) (30).

Fusion gene detection

Fusion genes were identified using four fusion gene detection tools: SPLICE (19), FusionSeeker version 1.0.1 (25), LongGF version 0.1.2 (26) and JAFFAL version 2.3 (27). The following commands were executed:

(i) SPLICE:

Preparation:

```
python ref_exonnum.py \
 -i <path to reference transcriptome (UCSC table browser output file)> \
 -g <path to reference genome sequence (FASTA)> \
```

Annotation to reference transcriptome:

```
python annot.py \
 -i <path to sample1(FASTQ)> \
 -g <path to reference genome sequence (FASTA)> \
 -r <path to output file of ref_exonum.py> \
 -p <prefix sample1> --tmp <path to temporary output file of sample>
```

(ii) FusionSeeker:

```
fusionseeker --bam <path to input file of sample (sorted bam file)> \
--outpath <path to output file of sample> \
--datatype nanopore \
--ref <path to reference file (hg38.fa)> \
--gtf  <path to gtf file (Homo_sapiens.GRCh38.104.chrname.gtf) > \
--min-supp 2 \
--max-distance 40 \
--keepfile \
--thread 8
```

(iii) LongGF:

```
LongGF --bam <path to input file of sample (bam file) > \
--gtf <path to gtf file (Homo_sapiens.GRCh38.104.chrname.gtf)> \
--min-overlap-len  100 \
--bin-size 50 \
--min-map-len 200 \
--pseudogene 0 <default = 0> \
--secondary-alignment 0 <default = 0> \
--support-read 2 <default = 2> \
--out <path to output file>
```

(iv) JAFFAL:

```
apptainer run \
 -B <path to reference file (hg38.fa)> \
 -B <path to input file> \
```

docker://davidsongroup/jaffa:latest \

JAFFA/JAFFAL.groovy <path to input file (*.fastq.gz)

For fusion genes identified by SPLICE, we applied the filtering criteria described in our previous study to remove candidates potentially arising from insufficient read support or read-through transcription (19): (1) candidates supported by fewer than two reads, and (2) candidates in which both partner genes were located on the same chromosome and within 200 kb of each other. For fusion genes identified by the other detection tools, candidates supported by at least two reads were selected for subsequent analyses.

Selection of candidates for fusion gene detection

Among the fusion gene candidates identified by the four fusion gene detection tools, we selected 45 candidates based on their potential biological relevance and/or read support. Specifically, candidates were prioritized if either the 5′ or 3′ partner gene was a known PDAC driver gene or a cancer-associated gene reported in previous studies (31–33), or if the candidate was supported by a relatively high number of reads in our dataset. A minimum read support threshold of two reads was applied to all selected candidates.

cDNA synthesis for validation experiment

Fusion gene candidates (n = 45) were amplified by RT-PCR followed by Sanger sequencing. For each sample, independent cDNA preparations were generated from the same RNA sample using three reverse-transcription approaches: (1) the SMARTer cDNA Synthesis Kit, which was also used for long-read sequencing library preparation; (2) the PrimeScript cDNA Synthesis Kit using the manufacturer's recommended primer mixture; and (3) the PrimeScript cDNA Synthesis Kit using a gene-specific primer targeting the 3′ partner gene (3'GSP; 3'gene-specific primer) (Fig. S1, Table S2). In the third method, a 3'GSP was used for reverse transcription instead of the primer mixture containing oligo (dT) and random hexamers. This approach was designed to initiate reverse transcription from the 3' partner sequence, thereby reducing nonspecific cDNA synthesis relative to conventional priming strategies.

Experimental validation of fusion genes

PCR primers were designed using Primer3 software (version 0.4.0) (34) to generate amplicons of approximately 250 bp spanning the predicted fusion junctions. PCR was performed using KOD -Multi & Epi-™ (TOYOBO). Primer sequences are listed in Table S3. PCR products were verified by 1.5% agarose gel electrophoresis and subjected to Sanger sequencing. A candidate was considered reproducibly validated when an amplicon of the expected size was detected in PCR using all three independently synthesized cDNA preparations, and the predicted fusion junction was confirmed by Sanger sequencing in both the SMARTer cDNA and the 3′GSP PrimeScript cDNA preparations.

Statistical analysis

The number of detected fusion genes and sequencing data yield (Gbp) were accessed for each sample and fusion gene detection tool (SPLICE, FusionSeeker, LongGF, and JAFFAL). The correlation between sequencing data yield and number of detected fusion genes was evaluated using Spearman's rank correlation analysis in R.

We analyzed the characteristics of reproducible and non-reproducible fusion gene candidates, including transposable element (TE) overlap, repetitive sequences, low-complexity sequences, short homologous sequences (SHSs) at fusion junctions, and mapping uniqueness. TE sequences were identified using RepeatMasker (version 4.1.8) (35). Low-complexity and repetitive sequences were identified using WindowMasker + SDust in the UCSC genome browser (36). The presence of SHSs at fusion junctions was assessed using the EMBOSS Waterman algorithm (37). We defined a SHS as a sequence of at least five nucleotides shared between the sequences flanking 5' and 3' breakpoints, allowing one nucleotide mismatch in the sequence. To evaluate mapping uniqueness, we assessed the mapping quality (MAPQ) score of supporting reads, which reflects the confidence of the read alignment. A MAPQ score of 60 represents the highest mapping quality (38).

Results

Long-read sequencing and identification of fusion gene candidates

We performed long-read sequencing using cDNAs synthesized from organoid RNA samples derived from 11 PDAC, three IPMN, two IPMN-derived carcinomas and one normal pancreas

(n = 17). The average data yield was 11.6 Gbp (2.1 - 28.2 Gbp), and the average read length was 830 bp (610 - 1,140 bp) (Table S1).

Fusion gene candidates were identified using four computational approaches: SPLICE (19), FusionSeeker version 1.0.1 (25), LongGF version 0.1.2 (26) and JAFFAL version 2.3 (27). Across the 17 samples, SPLICE detected 7,539 candidates, with support reads ranging from 2 to 77 (Fig. 1A). FusionSeeker, LongGF and JAFFAL identified 6,642, 3,643 and 14,698 candidates, respectively (support reads ≥ 2) (Table S4). The number of predicted fusion gene candidates by each fusion gene detection tool was positively correlated with the sequencing data yield (Fig. 1B). Among the four tools, JAFFAL detected the largest number of candidates, whereas LongGF detected the fewest.

We next examined overlap among the four fusion-calling approaches. A total of 1,950 candidates were detected by all four tools. An additional 593 candidates were detected by three tools (range 32 - 1,209, mean = 593), and 280 candidates by two tools (range 39 - 955, mean = 280). Candidates identified exclusively by a single tool numbered 9,023 for JAFFAL, 1,571 for SPLICE, 1,292 for FusionSeeker, and 177 for LongGF (Fig. 1C). These results indicate substantial differences in the candidate sets generated by the four fusion gene detection tools.

Experimental validation of fusion genes

Given the large number of computationally predicted candidates, we experimentally assessed the reproducibility of the selected candidates using independently synthesized cDNAs from the same RNA samples. We selected 45 candidates from the combined set of candidates identified by the four fusion gene detection tools (Fig. 2A, Table 1, and Table S5).

We first evaluated the 45 candidates using cDNA synthesized with the SMARTer cDNA Synthesis Kit, which was also used for the long-read sequencing library preparation. Forty-three of the 45 candidates (95.6%) were amplified by PCR, and the expected fusion junctions were confirmed by Sanger sequencing (Table 1).

Because most candidates were detectable using SMARTer-derived cDNA, we next evaluated the same candidates using cDNA independently synthesized with the PrimeScript RT Reagent Kit with gDNA Eraser. Using the same PCR primer pairs as in the initial validation, 9 of the 43 candidates detected in SMARTer-derived cDNA (20.9%) were detected in

PrimeScript-derived cDNA (Table 1). Thus, detection of the majority of candidates was not reproduced following an independent cDNA synthesis procedure.

To further assess the reproducibility of these candidates, we used a third cDNA synthesis approach based on 3′gene-specific primer (3′GSP). Because 3′GSP could not be designed for one candidate (*KRAS-VHL*), only the remaining eight candidates were included in this experiment. Among them, *GON4L-KLHDC9* (support reads = 48), *RABGAP1L-COP1* (support reads = 27), *MET-COPA* (support reads = 5), *MAP3K13-BRAF* (support reads = 2) and *BFAR-SMAD4* (support reads = 2) were successfully amplified by PCR and confirmed by Sanger sequencing (Fig. S2). The remaining three candidates (*GTF3C4-TP53, FRY-FOXP1* and *RUNX1-GNAS*), which were supported by two, five, and two support reads respectively, were not identified in PCR using 3'GSP-derived cDNA.

Overall, among the 45 candidates, five fusion genes were reproducibly detected using all three independent cDNA synthesis approaches (Table 1).

Comparison of fusion candidate detection across computational approaches

We next examined the overlap among computational approaches for the 45 experimentally evaluated candidates. Of the 40 candidates that were not reproducibly detected across all three cDNA synthesis approaches, 15 were detected by all four fusion gene detection tools, 11 by at least three tools, five only by SPLICE and FusionSeeker, six only by SPLICE and LongGF, one only by SPLICE and JAFFAL, and two exclusively by SPLICE (Fig. 2B, Table S6). Among the five reproducibly detected candidates, four were detected by all tools, whereas *MAP3K13-BRAF* fusion gene was detected exclusively by SPLICE (Fig. 2B).

Characteristics of fusion gene candidates

We next examined sequence and alignment characteristics that might be associated with the lack of reproducibility of fusion gene candidates.

(1) *SHS at fusion breakpoints*

Reverse transcriptase can undergo template switching during cDNA synthesis, in which the nascent cDNA dissociates from its original RNA template and reassociates with another RNA

molecule containing a homologous sequence (Fig. 3A) (39,40). We therefore examined whether SHSs were present at the predicted fusion junctions.

Fifteen of the 45 candidates (33.3%) contained SHSs in the regions flanking the predicted fusion junctions (Fig. 3B, C, Table 1). SHSs were identified at the breakpoint in one of the five reproducible candidates (20.0%) and in 14 of the 40 non-reproducible candidates (35.0%). Thus, SHSs were more frequently observed in non-reproducible candidates than in reproducible candidates.

(2) *MAPQ scores in fusion genes*

We examined MAPQ scores of the support reads for each fusion gene candidate. Among the 45 candidates, four showed MAPQ scores lower than 60 in either the 5' or 3' partner gene. Of these, three were non-reproducible (7.5%), and one was a reproducible fusion gene candidate (20.0%).

(3) *Breakpoint position relative to canonical splice sites*

Previous studies have reported that fusion breakpoints frequently occur at canonical exon-exon junctions and that non-canonical breakpoint locations may be associated with chimeric artifacts generated during sequencing library preparation (27,41,42). We therefore examined the positions of predicted fusion breakpoints.

Among the 45 candidates, 30 (66.7%) did not occur at canonical splicing sites (Fig. 3C, Table 1). Breakpoints in two of the five reproducible candidates (40.0%) were located at non-canonical exon–exon junctions, compared with 28 of the 40 non-reproducible candidates (70.0%). Thus, non-canonical breakpoint locations were more frequently observed in non-reproducible candidates than in reproducible candidates.

(4) *Repeat sequences at fusion breakpoints*

We next examined whether repetitive sequences were present at the predicted fusion breakpoints. Using RepeatMasker and WindowMasker, we identified repeat sequences (TE and low-complexity sequences) overlapping the breakpoint regions of the fusion gene candidates. Twenty-four of the 45 candidates (53.3%) had repeat sequences at the breakpoint region of either the 5' or 3' partner gene (Fig. 3C, Table 1). Repeat sequences were identified at the breakpoint in two of the five reproducible candidates (40.0%) and in 22 of the 40 non-

reproducible candidates (55.0%). Thus, repeat sequences were more frequently observed in non-reproducible candidates than in reproducible candidates.

Differences between reproducible and non-reproducible candidates

Taken together, these findings suggest that several features including SHSs, repetitive sequences at predicted breakpoints, and non-canonical breakpoint positions help identify fusion candidates that warrant additional experimental validation (Fig. 3C). The proportion of fusion gene candidates exhibiting at least one of these features at a breakpoint was significantly higher among non-reproducible candidates than among reproducible candidates ($P = 0.03$; Fig. 3C).

Non-reproducible KRAS fusion genes

In our analysis, seven recurrent *KRAS* fusion gene candidates were identified in the initial candidate set. However, all *KRAS* fusion gene candidates detected in the pancreatic samples were non-reproducible. Repeat element analysis of these *KRAS* fusion gene candidates revealed the presence of repeat elements, specifically short interspersed nuclear elements (SINEs), at the breakpoint regions of both the 5′ and 3′ partner genes (Fig. S3). The presence of these repeat elements may have contributed to the lack of reproducibility of the *KRAS* fusion gene candidates.

Discussion

Fusion transcripts are important molecular alterations in cancer and can serve as therapeutic targets and biomarkers for precision oncology (31). However, accurate identification of fusion transcripts from sequencing data remains challenging, particularly when technical artifacts generated during library preparation or read alignment resemble genuine fusion events. In this study, we evaluated the reproducibility of computationally predicted fusion transcript candidates identified from long-read transcriptome sequencing of pancreatic organoid samples. We assessed whether their detection could be replicated using independently synthesized cDNAs generated by different reverse-transcription approaches. Of the 45 selected candidates, 43 were detected using the SMARTer-based cDNA preparation used for long-read sequencing,

whereas only 9 of these candidates were detected using independently synthesized PrimeScript-derived cDNA. Of the eight candidates that could subsequently be evaluated using 3′GSP-derived cDNA, five were detected consistently across all three cDNA synthesis approaches. These findings demonstrate that experimental reproducibility can differ substantially among independently synthesized cDNA.

To our knowledge, this is the first study to apply 3′GSP-based cDNA synthesis for the experimental validation of fusion transcript candidates identified by long-read sequencing. This approach initiates reverse transcription from a sequence within the 3′ partner gene, rather than relying on oligo(dT) and random hexamers, and may therefore reduce nonspecific cDNA synthesis and the generation or enrichment of chimeric cDNAs. Although the use of 3′GSP requires additional primer design and experimental procedures for each candidate and is therefore more laborious than conventional cDNA synthesis, our results suggest that it can serve as a useful complementary approach for evaluating the reproducibility of fusion transcript candidates.

An important observation was that agreement among computational fusion detection tools did not necessarily predict experimental reproducibility. Among the 40 candidates that were not reproducibly detected across all three cDNA synthesis approaches, 11 had been identified by at least three of the four fusion gene detection tools. Conversely, the reproducibly detected *MAP3K13-BRAF* candidate was identified only by SPLICE. These findings indicate that concordance among fusion gene detection tools alone may not be sufficient to establish the reproducibility of a fusion transcript candidate. Differences in the assumptions and filtering criteria of individual algorithms may contribute to this discrepancy. In the case of *MAP3K13-BRAF*, the *BRAF* portion of the fusion corresponded to a novel transcript, while the *MAP3K13* fragment was only 87 bp long. Thus, the candidate was not identified by the other fusion gene detection tools because of algorithm-specific constraints on novel transcripts or short gene fragments.

We examined sequence characteristics that may help explain the lack of reproducibility. The proportion of candidates with SHSs were higher in non-reproducible than reproducible candidates (Fig. 3C). This enrichment of SHSs among non-reproducible candidates is consistent with previous reports showing that reverse transcriptase can undergo template switching between RNA molecules containing homologous sequences (43,44). In such a process, a nascent cDNA molecule may switch from one RNA template to another RNA

molecule containing a homologous sequence, potentially producing a chimeric cDNA molecule. Our data do not directly demonstrate that template switching generated the individual candidates examined in this study. Nevertheless, the enrichment of SHSs among non-reproducible candidates suggests that SHSs may be useful sequence features for prioritizing fusion candidates for experimental validation.

We also examined the presence of repeat elements at predicted fusion breakpoints. Repeat elements were more frequently observed at these breakpoints in non-reproducible candidates (Fig. 3C). Because many repetitive elements share substantial sequence similarity, they may contribute to the generation of false-positive fusion candidates. Although the present study cannot distinguish between these possible mechanisms, the presence of repeat elements at breakpoint regions may provide useful information for identifying false-positive fusion genes.

Breakpoint positions relative to canonical splice sites provided another potentially informative feature. The proportion of non-canonical exon-exon junctions were higher in non-reproducible than reproducible candidates (Fig. 3C). This observation is consistent with previous studies showing that information on canonical splice junctions can improve fusion transcript detection and help distinguish plausible transcript structures from some chimeric products generated during library preparation (27,41,42). Nevertheless, two non-canonical candidates in our study were reproducibly detected across all three cDNA synthesis approaches. Therefore, non-canonical breakpoint locations should not be considered evidence that a fusion transcript is artificial. Rather, they may useful when prioritizing candidates for additional validation.

Taken together, these findings suggest that several features including SHSs, repetitive sequences at predicted breakpoints, and non-canonical breakpoint positions help identify fusion candidates that warrant additional experimental validation (Fig. 3C). However, these features should not be interpreted as definitive criteria for excluding fusion candidates. Our results instead support a prioritization strategy in which candidates with such features receive particular attention during validation.

Our findings also have implications for the design of experimental validation strategies. When fusion candidates are validated using the same cDNA preparation used for long-read sequencing, technical features introduced during cDNA synthesis may potentially be shared between discovery and validation experiments. Independent cDNA synthesis using a different

reverse-transcription strategy can provide an additional level of evidence for reproducible detection. This approach may be particularly useful when candidates are supported by limited numbers of reads, have ambiguous mapping characteristics, contain repetitive or homologous sequences, or have non-canonical breakpoint structures. For candidates intended for functional characterization or potential clinical application, independent experimental confirmation should therefore be considered before downstream interpretation.

In this study, the number of fusion transcript candidates detected was substantially higher than that reported in previous studies (19,22,45). One possible explanation is the high expression of repetitive elements associated with global DNA hypomethylation in pancreatic cancer (46,47). Highly expressed repetitive sequences may increase the opportunity for SHS-mediated template switching during cDNA synthesis, potentially generating chimeric cDNAs that are subsequently identified as fusion transcript candidates. Therefore, the large number of candidates and the relatively low reproducibility observed in this study may, at least in part, reflect characteristics of the pancreatic cancer samples examined here. The frequency of SHS-associated non-reproducible fusion transcript candidates may differ in other cancer types.

This study has several limitations. First, experimental validation was performed for 45 selected candidates rather than for the complete set of computationally predicted candidates. Second, the number of samples and experimentally evaluated candidates was relatively small, limiting the statistical power to evaluate associations between individual sequence features and reproducibility. Larger studies with systematic candidate selection will be required to determine the predictive value of SHSs, repeat elements, breakpoint position, MAPQ, and other features. Finally, the present analysis focused on pancreatic organoid samples, and the abundance and composition of repetitive transcripts may differ among cancer types and tissue contexts. Validation in additional cancer types will therefore be necessary to determine the generalizability of these observations.

In conclusion, long-read transcriptome sequencing provides substantial advantages for characterizing transcript structures and identifying fusion transcript candidates because individual reads can span extended regions of transcripts and fusion junctions. However, computationally predicted candidates may not be reproducibly detected when independently synthesized cDNAs are examined. In our dataset, agreement among multiple fusion detection tools did not necessarily ensure experimental reproducibility, whereas sequence features such as SHSs, repeat elements, and non-canonical breakpoint locations were frequently observed

among non-reproducible candidates. These findings support the use of independent cDNA synthesis as an additional validation step for fusion transcript candidates identified by long-read sequencing, particularly when candidates are considered for functional or clinical applications.

Declarations

Abbreviations

ONT	Oxford Nanopore Technologies
PDAC	Pancreatic Ductal Adenocarcinoma
IPMN	Intraductal Papillary Mucinous Neoplasms
RIN	RNA Integrity Number
TE	Transposable Element
3'GSP	3'Gene-Specific Primer
SHS	Short Homologous Sequence
MAPQ	Mapping Quality
MMLV RT	Moloney Murine Leukemia Virus Reverse Transcriptase

Ethics approval and consent to participate

The research plan for this work was approved by the Institutional Review Board (IRB) at the University of Tokyo (approval no.G10096 and 3591). Written informed consent was obtained from all participants. All experiments involving human samples were performed in accordance with the Declaration of Helsinki and relevant national guidelines and regulations.

Consent for publication

Not applicable

Availability of data and materials

Sequencing data have been deposited in the JGA database (https://humandbs.dbcls.jp/en) in Japan under accession numbers JGAS000957 (https://humandbs.dbcls.jp/en/dataset/JGAD001103).

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by AMED (25ck0106873h0004, 24fk0310522h0003 to AF), Grant-in-Aid for Scientific Research (B) from the Japanese Society for the Promotion of Science (23H02743 to A.F.) and the Takeda Science Foundation (AF).

Authors' contributions

AF and KT designed the study. HK, KY and KT prepared the samples. ACP performed the computational analyses. ACP and SA performed the experiments. ACP, SA, KT, IK, MU and AF interpreted the results. ACP and AF wrote the manuscript. All authors approved the final manuscript.

Acknowledgements

The super-computing resource was provided by the Human Genome Centre, Institute of Medical Science, the University of Tokyo.

Figure legends

Figure 1. Fusion genes detected in 17 pancreatic organoid samples.

(A) Distribution of fusion genes' support reads across 17 samples.

(B) Scatter plot showing correlation between the number of fusion genes identified by each fusion gene detection tool and sequencing data yield (Gbp) for each sample. Each dot represents a fusion gene detection tool: SPLICE (orange), Fusion Seeker (green), LongGF (blue) and JAFFAL (purple). $r = 0.96$, $P = 3.94 \times 10^{-10}$ (SPLICE); $r = 0.96$, $P = 1.35 \times 10^{-6}$ (Fusion Seeker); $r = 0.96$, $P = 3.47 \times 10^{-10}$ (LongGF); and $r = 0.97$, $P = 5.41 \times 10^{-6}$ (JAFFAL). *r,* Spearman's correlation coefficient.

(C) UpSet plot showing the intersection sizes and detection patterns of fusion genes across the four fusion gene detection tools. LGF, LongGF; JFL, JAFFAL; and FS, FusionSeeker.

Figure 2. Fusion gene candidates selected for PCR validation and identification by fusion gene detection tools.

(A) Fusion gene candidates selected for PCR validation (n = 45) and distribution of support reads.

(B) Bar graph showing number of reproducible and non-reproducible fusion gene candidates identified by all four tools, any three tools, any two tools, or exclusively one tool.

Figure 3. Comprehensive analysis of reproducible and non-reproducible fusion gene candidates.

(A) Schematic illustration of artificial fusion transcript formation induced by homology-dependent template switching by Moloney murine leukaemia virus reverse transcriptase (MMLV RT) during cDNA synthesis.

(B) Example of fusion gene candidates showing short homologous sequences (SHSs) at fusion breakpoints.

(C) Proportion of fusion gene candidates in reproducible and non-reproducible groups. The proportion of candidates exhibiting at least one of the following features (SHS, repetitive sequences at breakpoints, non-canonical splice sites) was significantly higher among non-reproducible candidates than among reproducible candidates.

References

1. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nature Reviews Cancer. 2007. p. 233–45. doi:10.1038/nrc2091 PubMed PMID: 17361217.
2. Roukos V, Misteli T. The biogenesis of chromosome translocations. 2014.
3. Mitelman F, Johansson B, Mertens F. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. Nature Genetics. 2004. p. 331–4. doi:10.1038/ng1335 PubMed PMID: 15054488.
4. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012 Jun 21;486(7403):405–9. doi:10.1038/nature11154 PubMed PMID: 22722202.
5. Koo KM, Wee EJH, Trau M. Colorimetric TMPRSS2-ERG gene fusion detection in prostate cancer urinary samples via recombinase polymerase amplification. Theranostics. 2016;6(9):1415–24. doi:10.7150/thno.15250 PubMed PMID: 27375789.
6. Choi Y, Kwon CH, Lee SJ, Park J, Shin JY, Park DY. Integrative analysis of oncogenic fusion genes and their functional impact in colorectal cancer. Br J Cancer. 2018 Jul 17;119(2):230–40. doi:10.1038/s41416-018-0153-3 PubMed PMID: 29955133.
7. Nussenzweig A, Nussenzweig MC. Origin of Chromosomal Translocations in Lymphoid Cancer. Cell. Elsevier B.V.; 2010. p. 27–38. doi:10.1016/j.cell.2010.03.016 PubMed PMID: 20371343.
8. Cui JJ, Tran-Dubé M, Shen H, Nambu M, Kung PP, Pairish M, et al. Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal-epithelial transition factor (c-MET) kinase and anaplastic lymphoma kinase (ALK). J Med Chem. 2011 Sep 22;54(18):6342–63. doi:10.1021/jm2007613 PubMed PMID: 21812414.
9. Laé M, Fréneaux P, Sastre-Garau X, Chouchane O, Sigal-Zafrani B, Vincent-Salomon A. Secretory breast carcinomas with ETV6-NTRK3 fusion gene belong to the basal-

like carcinoma spectrum. Modern Pathology. 2009;22(2):291–8. doi:10.1038/modpathol.2008.184 PubMed PMID: 19011601.

10. Laetsch TW, DuBois SG, Mascarenhas L, Turpin B, Federman N, Albert CM, et al. Larotrectinib for paediatric solid tumours harbouring NTRK gene fusions: phase 1 results from a multicentre, open-label, phase 1/2 study. Lancet Oncol. 2018 May 1;19(5):705–14. doi:10.1016/S1470-2045(18)30119-0 PubMed PMID: 29606586.

11. 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 Oct 21;20(1). doi:10.1186/s13059-019-1842-9 PubMed PMID: 31639029.

12. Neckles C, Sundara Rajan S, Caplen NJ. Fusion transcripts: Unexploited vulnerabilities in cancer? Wiley Interdiscip Rev RNA. 2020 Jan 1;11(1). doi:10.1002/wrna.1562 PubMed PMID: 31407506.

13. Shah N, Lankerovich M, Lee H, Yoon JG, Schroeder B, Foltz G. Exploration of the gene fusion landscape of glioblastoma using transcriptome sequencing and copy number data [Internet]. 2013. Available from: http://www.biomedcentral.com/1471-2164/14/818

14. Bae K, Kim JH, Jung H, Kong SY, Kim YH, Kim S, et al. A fusion of CD63–BCAR4 identified in lung adenocarcinoma promotes tumorigenicity and metastasis. Br J Cancer. 2021 Jan 5;124(1):290–8. doi:10.1038/s41416-020-01146-3 PubMed PMID: 33204025.

15. Kerbs P, Vosberg S, Krebs S, Graf A, Blum H, Swoboda A, et al. Fusion gene detection by RNA-sequencing complements diagnostics of acute myeloid leukemia and identifies recurring NRIP1-MIR99AHG rearrangements. Haematologica. 2022 Jan 1;107(1):100–11. doi:10.3324/haematol.2021.278436 PubMed PMID: 34134471.

16. Mandelker D, Schmidt RJ, Ankala A, McDonald Gibson K, Bowser M, Sharma H, et al. Navigating highly homologous genes in a molecular diagnostic setting: A resource for clinical next-generation sequencing. Genetics in Medicine. 2016 Dec 1;18(12):1282–9. doi:10.1038/gim.2016.58 PubMed PMID: 27228465.

17. Chen X, Harting J, Farrow E, Thiffault I, Kasperaviciute D, Hoischen A, et al. Comprehensive SMN1 and SMN2 profiling for spinal muscular atrophy analysis using long-read PacBio HiFi sequencing. Am J Hum Genet. 2023 Feb 2;110(2):240–50. doi:10.1016/j.ajhg.2023.01.001 PubMed PMID: 36669496.

18. Olivucci G, Iovino E, Innella G, Turchetti D, Pippucci T, Magini P. Long read sequencing on its way to the routine diagnostics of genetic diseases. Frontiers in Genetics. Frontiers Media SA; 2024. doi:10.3389/fgene.2024.1374860

19. Kiyose H, Nakagawa H, Ono A, Aikata H, Ueno M, Hayami S, et al. Comprehensive analysis of full-length transcripts reveals novel splicing abnormalities and oncogenic

transcripts in liver cancer. PLoS Genet. 2022 Aug 4;18(8). doi:10.1371/journal.pgen.1010342 PubMed PMID: 35926060.

20. T Veiga DF, Nesta A, Zhao Y, Deslattes Mays A, Huynh R, Rossi R, et al. A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. Sci. Adv [Internet]. 2022. Available from: https://www.science.org

21. Sun Q, Han Y, He J, Wang J, Ma X, Ning Q, et al. Long-read sequencing reveals the landscape of aberrant alternative splicing and novel therapeutic target in colorectal cancer. Genome Med. 2023 Dec 1;15(1). doi:10.1186/s13073-023-01226-y PubMed PMID: 37735421.

22. Liu M, Han Z, Zhi Y, Ruan Y, Cao G, Wang G, et al. Long-read sequencing reveals oncogenic mechanism of HPV-human fusion transcripts in cervical cancer. Translational Research. 2023 Mar 1;253:80–94. doi:10.1016/j.trsl.2022.09.004 PubMed PMID: 36223881.

23. Vollger MR, Korlach J, Eldred KC, Swanson E, Underwood JG, Bohaczuk SC, et al. Synchronized long-read genome, methylome, epigenome and transcriptome profiling resolve a Mendelian condition. Nat Genet. 2025 Feb 1;57(2):469–79. doi:10.1038/s41588-024-02067-0 PubMed PMID: 39880924.

24. Rybacki K, Xu F, Deutsch HM, Ahsan MU, Chan J, Liang Z, et al. Combining panel-based and whole-transcriptome-based gene fusion detection by long-read sequencing. Cell Reports Methods. 2025 Aug 18;5(8). doi:10.1016/j.crmeth.2025.101111 PubMed PMID: 40695274.

25. Chen Y, Wang Y, Chen W, Tan Z, Song Y, Chen H, et al. Gene Fusion Detection and Characterization in Long-Read Cancer Transcriptome Sequencing Data with FusionSeeker. Cancer Res. 2023 Jan 1;83(1):28–33. doi:10.1158/0008-5472.CAN-22-1628 PubMed PMID: 36318117.

26. Liu Q, Hu Y, Stucky A, Fang L, Zhong JF, Wang K. LongGF: computational algorithm and software tool for fast and accurate detection of gene fusions by long-read transcriptome sequencing. BMC Genomics. 2020 Dec 1;21. doi:10.1186/s12864-020-07207-4 PubMed PMID: 33372596.

27. Davidson NM, Chen Y, Sadras T, Ryland GL, Blombery P, Ekert PG, et al. JAFFAL: detecting fusion genes with long-read transcriptome sequencing. Genome Biol. 2022 Dec 1;23(1). doi:10.1186/s13059-021-02588-5 PubMed PMID: 34991664.

28. Masuda K, Sota Y, Matsuda H. Detecting Fusion Genes in Long-Read Transcriptome Sequencing Data with FUGAREC. IPSJ Transactions on Bioinformatics. 2024;17:1–9. doi:10.2197/ipsjtbio.17.1

29. Kato H, Tateishi K, Fujiwara H, Nakatsuka T, Yamamoto K, Kudo Y, et al. MNX1-HNF1B Axis Is Indispensable for Intraductal Papillary Mucinous Neoplasm Lineages.

Gastroenterology. 2022 Apr 1;162(4):1272-1287.e16. doi:10.1053/j.gastro.2021.12.254 PubMed PMID: 34953915.

30. https://github.com/rrwick/Porechop/issues/86.

31. Gkountakos A, Singhi AD, Westphalen CB, Scarpa A, Luchini C. Fusion genes in pancreatic tumors. Trends in Cancer. Cell Press; 2024. p. 430–43. doi:10.1016/j.trecan.2024.01.009 PubMed PMID: 38378317.

32. Reshkin SJ, Cardone RA, Koltai T. Genetic Signature of Human Pancreatic Cancer and Personalized Targeting. Cells. Multidisciplinary Digital Publishing Institute (MDPI); 2024. doi:10.3390/cells13070602 PubMed PMID: 38607041.

33. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018 Apr 5;173(2):371-385.e18. doi:10.1016/j.cell.2018.02.060 PubMed PMID: 29625053.

34. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-new capabilities and interfaces. Nucleic Acids Res. 2012 Aug;40(15). doi:10.1093/nar/gks596 PubMed PMID: 22730293.

35. Smit, A., Hubley, R. & Green, P. RepeatMasker. http://www.repeatmasker.org (2013).

36. Morgulis A, Gertz EM, Schäffer AA, Agarwala R. WindowMasker: Window-based masker for sequenced genomes. Bioinformatics. 2006 Jan 15;22(2):134–41. doi:10.1093/bioinformatics/bti774 PubMed PMID: 16287941.

37. https://www.ebi.ac.uk/jdispatcher/psa/emboss_water.

38. Li H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 2018 Sep 15;34(18):3094–100. doi:10.1093/bioinformatics/bty191 PubMed PMID: 29750242.

39. Verwilt J, Mestdagh P, Vandesompele JO. Artifacts and biases of the reverse transcription reaction in RNA sequencing [Internet]. 2023. doi:10.1261/rna

40. Cocquet J, Chong A, Zhang G, Veitia RA. Reverse transcriptase template switching and false alternative transcripts. Genomics. 2006 Jul;88(1):127–31. doi:10.1016/j.ygeno.2005.12.013 PubMed PMID: 16457984.

41. Li W, Wan R, Guo L, Chang G, Jiang D, Meng L, et al. Reliability analysis of exonic-breakpoint fusions identified by DNA sequencing for predicting the efficacy of targeted therapy in non-small cell lung cancer. BMC Med. 2022 Dec 1;20(1). doi:10.1186/s12916-022-02362-9 PubMed PMID: 35534835.

42. Lu B, Jiang R, Xie B, Wu W, Zhao Y. Fusion genes in gynecologic tumors: the occurrence, molecular mechanism and prospect for therapy. Cell Death and Disease. Springer Nature; 2021. doi:10.1038/s41419-021-04065-0 PubMed PMID: 34381020.

43. Li X, Zhao L, Jiang H, Wang W. Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. J Mol Evol. 2009 Jan;68(1):56–65. doi:10.1007/s00239-008-9187-0 PubMed PMID: 19089307.

44. Luo G, Taylor J. Template Switching by Reverse Transcriptase during DNA Synthesis. JOURNAL OF VIROLOGY [Internet]. 1990. Available from: https://journals.asm.org/journal/jvi

45. Mock A, Braun M, Scholl C, Fröhling S, Erkut C. Transcriptome profiling for precision cancer medicine using shallow nanopore cDNA sequencing. Sci Rep. 2023 Dec 1;13(1). doi:10.1038/s41598-023-29550-8 PubMed PMID: 36759549.

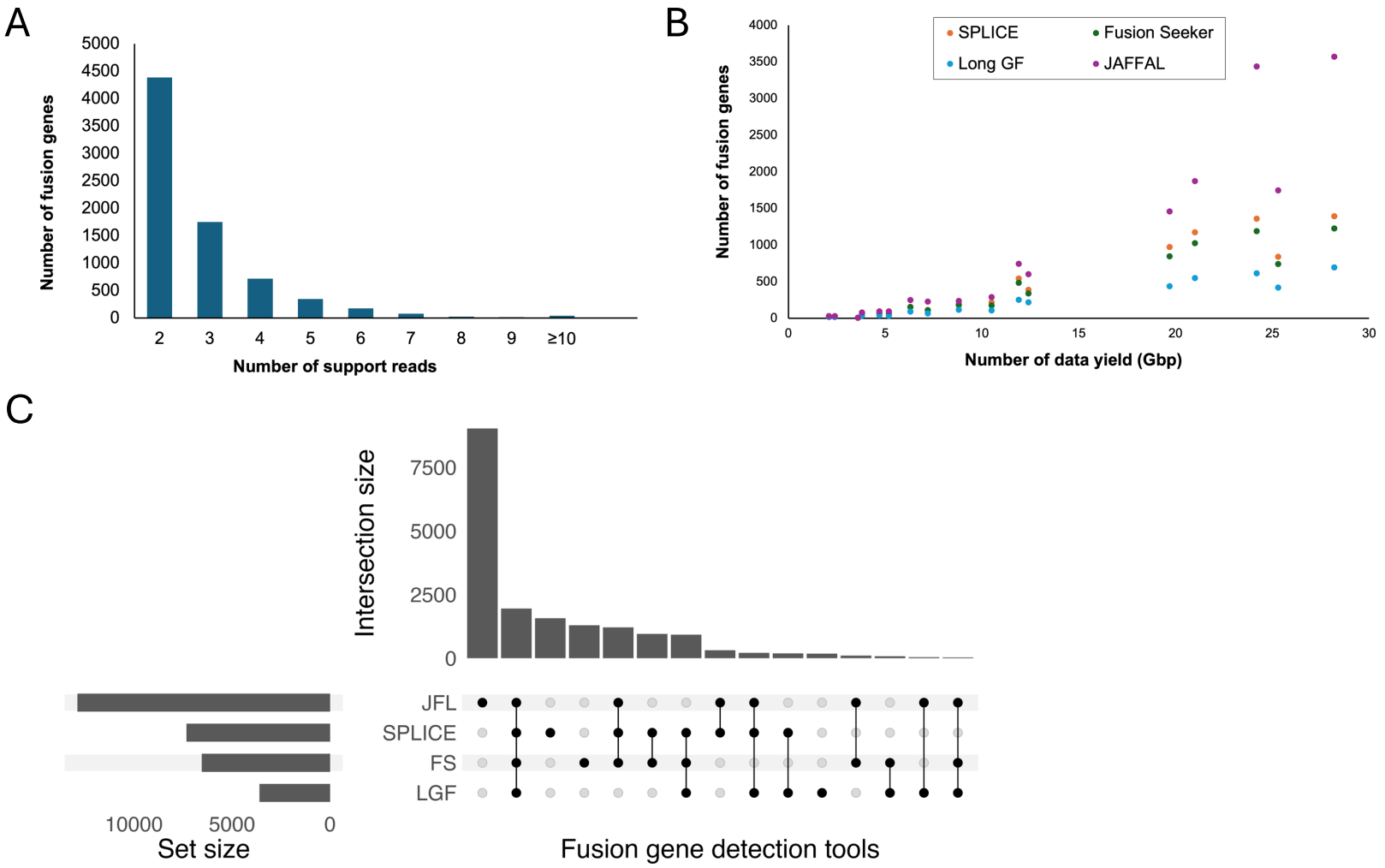
46. You E, Danaher P, Lu C, Sun S, Zou L, Phillips IE, et al. Disruption of cellular plasticity by repeat RNAs in human pancreatic cancer. Cell. 2024 Dec 12;187(25):7232-7247.e23. doi:10.1016/j.cell.2024.09.024 PubMed PMID: 39383862.

47. Espinet E, Gu Z, Imbusch CD, Giese NA, Büscher M, Safavi M, et al. Aggressive pdacs show hypomethylation of repetitive elements and the execution of an intrinsic ifn program linked to a ductal cell of origin. Cancer Discov. 2021 Mar 1;11(3):638–59. doi:10.1158/2159-8290.CD-20-1202 PubMed PMID: 33060108.

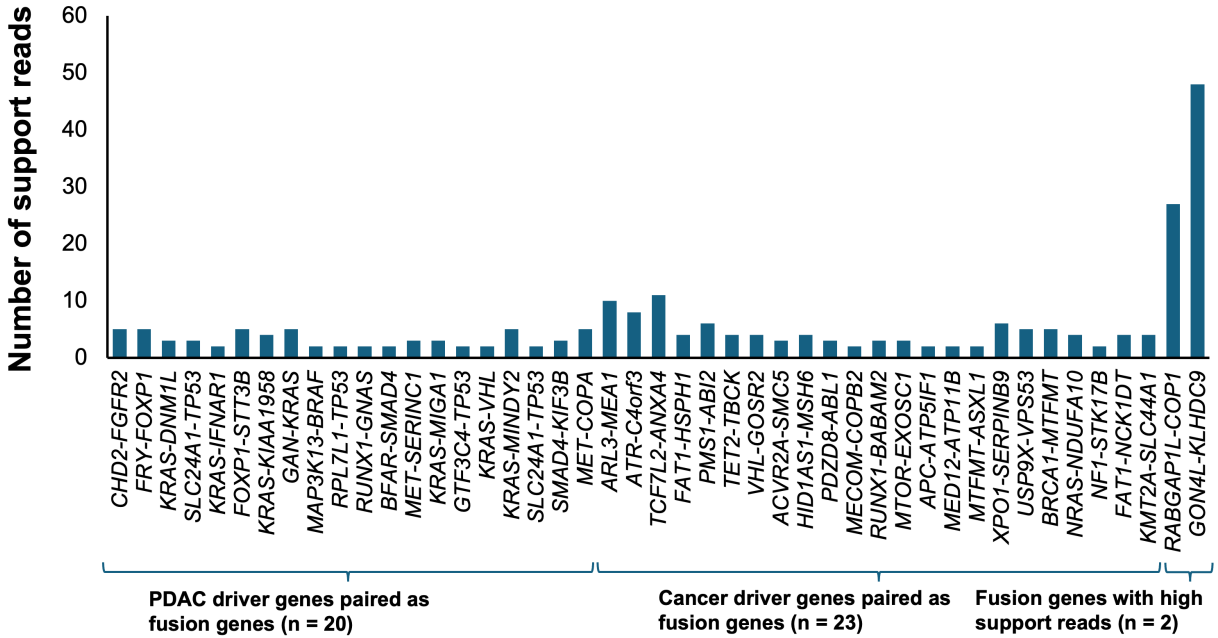
Table 1. PCR validation result of fusion gene candidates.

No	Sample name	Fusion gene	Number of support reads	SMARTer cDNA	PrimeScript cDNA	3′ GSP cDNA	MAPQ [a]	SHS at breakpoint	Breakpoint at canonical splice site
1	Pancreas 11	*GON4L-KLHDC9*	48	Detected	Detected	Detected	60/60	Not present	Yes
2	Pancreas 2	*RABGAP1L-COP1*	27						
3	Pancreas 17	*MET-COPA*	5						No
4	Pancreas 9	*BFAR-SMAD4*	2					Present	
5	Pancreas 3	*MAP3K13-BRAF*	2				60/42	Not present	Yes
6	Pancreas 2	*FRY-FOXP1*	5			Not Detected	60/60	Present	
7	Pancreas 9	*RUNX1-GNAS*	2						
8	Pancreas 10	*GTF3C4-TP53*	2						No
9	Pancreas 10	*KRAS-VHL*	2			Not tested			
10	Pancreas 4	*RPL7L1-TP53*	2		Not Detected		15/60		
11	Pancreas 14	*XPO1-SERPINB9*	6				52/60	Not present	
12	Pancreas 17	*KMT2A-SLC44A1*	4				36/60		
13	Pancreas 2	*ATR-C4orf3*	8				60/60	Present	
14	Pancreas 14	*USP9X-VPS53*	5						
15	Pancreas 3	*KRAS-KIAA1958*	4						
16	Pancreas 14	*NRAS-NDUFA10*	4						
17	Pancreas 10	*MET-SERINC1*	3						
18	Pancreas 2	*KRAS-IFNAR1*	2						
19	Pancreas 2	*CHD2-FGFR2*	5						Yes
20	Pancreas 14	*BRCA1-MTFMT*	5						No
21	Pancreas 10	*RUNX1-BABAM2*	3						Yes
22	Pancreas 2	*ARL3-MEA1*	10					Not present	No
23	Pancreas 3	*VHL-GOSR2*	4						
24	Pancreas 3	*HID1AS1-MSH6*	4						
25	Pancreas 2	*KRAS-DNM1L*	3						
26	Pancreas 2	*SLC24A1-TP53*	3						
27	Pancreas 10	*KRAS-MIGA1*	3						
28	Pancreas 17	*SMAD4-KIF3B*	3						
29	Pancreas 3	*PDZD8-ABL1*	3						
30	Pancreas 16	*SLC24A1-TP53*	2						
31	Pancreas 10	*MED12-ATP11B*	2						
32	Pancreas 11	*MTFMT-ASXL1*	2						
33	Pancreas 14	*NF1-STK17B*	2						
34	Pancreas 2	*TCF7L2-ANXA4*	11						Yes
35	Pancreas 2	*PMS1-ABI2*	6						
36	Pancreas 3	*FOXP1-STT3B*	5						
37	Pancreas 2	*FAT1-HSPH1*	4						
38	Pancreas 2	*TET2-TBCK*	4						
39	Pancreas 17	*FAT1-NCK1DT*	4						
40	Pancreas 3	*ACVR2A-SMC5*	3						
41	Pancreas 10	*MTOR-EXOSC1*	3						
42	Pancreas 9	*MECOM-COPB2*	2						
43	Pancreas 10	*APC-ATP5IF1*	2						No
44	Pancreas 3	*GAN-KRAS*	5	Not Detected					
45	Pancreas 14	*KRAS-MINDY2*	5						

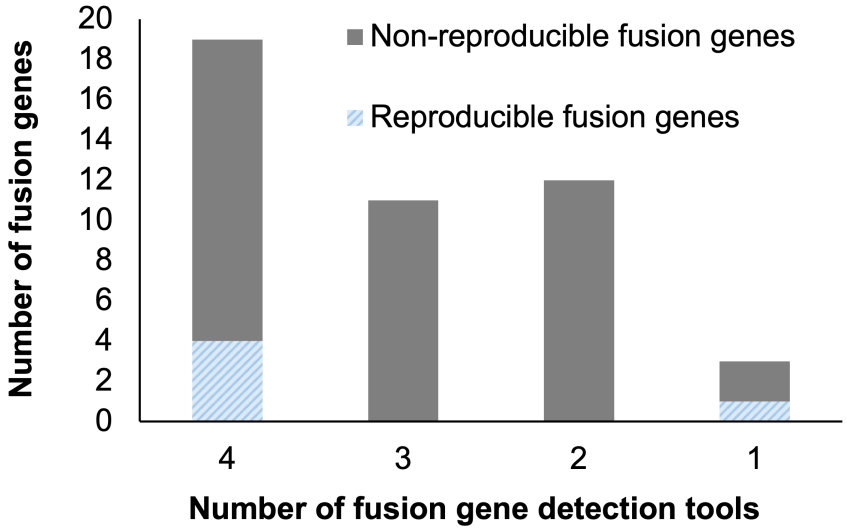
[a] MAPQ of 5′ and 3′ genes


A
Number of fusion genes
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
2
3
4
5
6
7
8
9
≥10
Number of support reads
B
Number of fusion genes
4000
3500
3000
2500
2000
1500
1000
500
0
SPLICE
Fusion Seeker
Long GF
JAFFAL
0
5
10
15
20
25
30
Number of data yield (Gbp)
C
Intersection size
7500
5000
2500
0
JFL
SPLICE
FS
LGF
10000
5000
0
Set size
Fusion gene detection tools

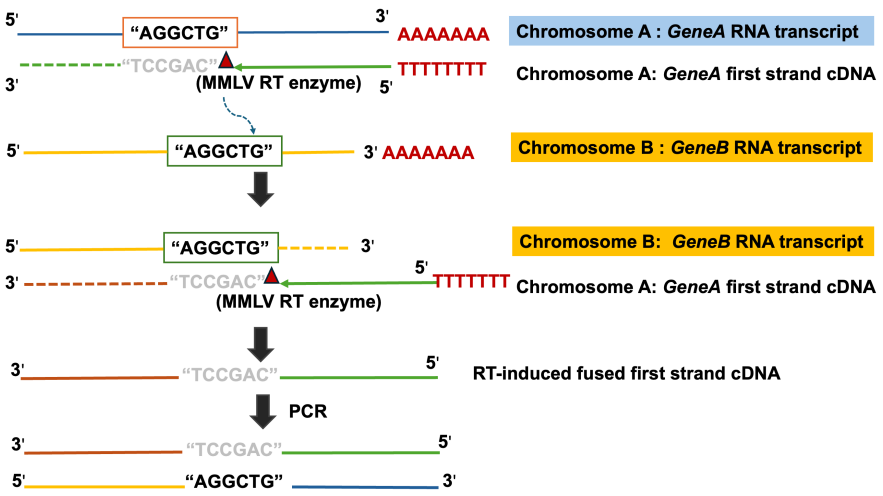
A


Number of support reads
60
50
40
30
20
10
0
CHD2-FGFR2
FRY-FOXP1
KRAS-DNM1L
SLC24A1-TP53
KRAS-IFNAR1
FOXP1-STT3B
KRAS-KIAA1958
GAN-KRAS
MAP3K13-BRAF
RPL7L1-TP53
RUNX1-GNAS
BFAR-SMAD4
MET-SERINC1
KRAS-MIGA1
GTF3C4-TP53
KRAS-VHL
KRAS-MINDY2
SLC24A1-TP53
SMAD4-KIF3B
MET-COPA
ARL3-MEA1
ATR-C4orf3
TCF7L2-ANXA4
FAT1-HSPH1
PMS1-ABI2
TET2-TBCK
VHL-GOSR2
ACVR2A-SMC5
HID1AS1-MSH6
PDZD8-ABL1
MECOM-COPB2
RUNX1-BABAM2
MTOR-EXOSC1
APC-ATP5IF1
MED12-ATP11B
MTFMT-ASXL1
XPO1-SERPINB9
USP9X-VPS53
BRCA1-MTFMT
NRAS-NDUFA10
NF1-STK17B
FAT1-NCK1DT
KMT2A-SLC44A1
RABGAP1L-COP1
GON4L-KLHDC9
PDAC driver genes paired as fusion genes (n = 20)
Cancer driver genes paired as fusion genes (n = 23)
Fusion genes with high support reads (n = 2)

B


Number of fusion genes
20
18
16
14
12
10
8
6
4
2
0
Non-reproducible fusion genes
Reproducible fusion genes
4
3
2
1
Number of fusion gene detection tools

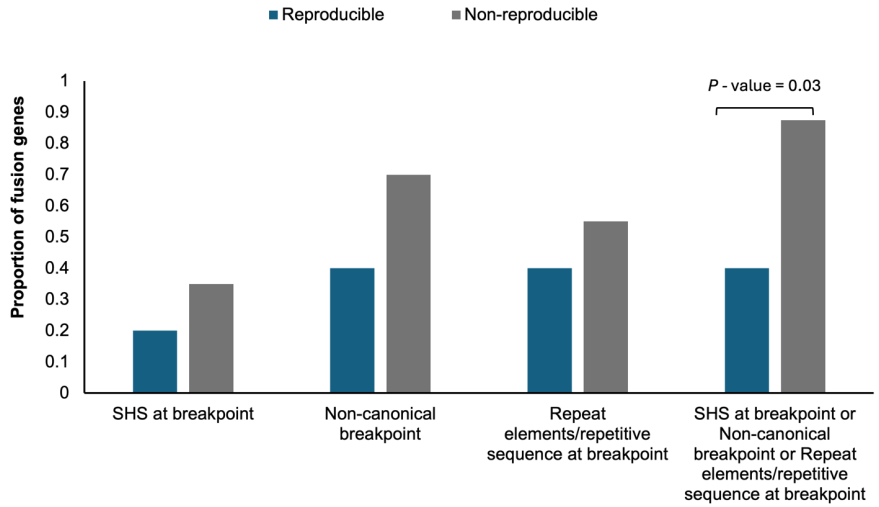
A



B

MET-SERINC1	*MET*	---------GTTGCCAGGGCTGTAG**TGCAGTG|TGCAGTG**AGCCAAGGTCACGTCA--------	*SERINC1*
BRCA1-MTFMT	*BRCA1*	---------ACCCCAGATCCCCCAC**AGCCAC|AGCCAC**CGTGCCTGGCCTGGAA---------	*MTFMT*
KRAS-IFNAR1	*KRAS*	--------- CTCGCTCTTGTCACCC**AGGCTG|AGGCTG**AGGCAGGAGAATCGCT ---------	*IFNAR1*
KRAS-KIAA1958	*KRAS*	--------- CACTGCAACCTCCATC**TCCCAG|TCCCAG**CTACTTGAGAGTCTGA ---------	*KIAA1958*
RPL7L1-TP53	*RPL7L1*	--------- TAACTAAAAACTGTAT**TGCTGG|TGCTGG**GATTACAATTGTGAGC ---------	*TP53*
CHD2-FGFR2	*CHD2*	--------- ACACAGCAATGCATCG**AGGTA|AGGTA**CATGATGATGAGGGAC ---------	*FGFR2*
RUNX1-GNAS	*RUNX1*	---------CACAGTGCTTCATGAG**AGGTG|AGGTG**CTGGAGAATCTGGTAA ---------	*GNAS*

C



Supplementary Files

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

- Suppplementaryinformation.docx