Gene Fusion Identification Using Anchor-Based Multiplex PCR and Next-Generation Sequencing

Yu-Wei Cheng,[a,*] Anders Meyer,[b,c] Maureen A. Jakubowski,[a] Sean O. Keenan,[a,d] Jay E. Brock,[a] Elizabeth M. Azzato,[a] Michael Weindel,[a] Daniel H. Farkas,[a] and Brian P. Rubin[c]

Background: Methods for identifying gene fusion events, such as fluorescence in situ hybridization (FISH), immunohistochemistry (IHC), and transcriptome analysis, are either single gene approaches or require bioinformatics expertise not generally available in clinical laboratories. We analytically validated a customized next-generation sequencing (NGS) panel targeting fusion events in 34 genes involving soft-tissue sarcomas.

Methods: Specimens included 87 formalin-fixed paraffin-embedded (FFPE) tissues with known gene fusion status. Isolated total nucleic acid was used to identify fusion events at the RNA level. The potential fusions were targeted by gene-specific primers, followed by primer extension and nested PCR to enrich for fusion candidates with subsequent bioinformatics analysis.

Results: The study generated results using the following quality metrics for fusion detection: (*a*) ≥100 ng total nucleic acid, (*b*) RNA average unique start sites per gene-specific primer control ≥10, (*c*) quantitative PCR assessing input RNA quality had a crossing point <30, (*d*) total RNA percentage ≥30%, and (*e*) total sequencing fragments ≥500 000.

Conclusions: The test validation study demonstrated analytical sensitivity of 98.7% and analytical specificity of 90.0%. The NGS-based panel generated highly concordant results compared to alternative testing methods.

INTRODUCTION

A fusion gene is a hybrid gene formed from 2 previously independent genes and mostly occurs due to chromosomal translocations or inversions. Fusion can juxtapose functional domains of 2 distant genes resulting in new gene function and driving tumorigenesis. The *BCR–ABL1* gene fusion, the first identified and most well-known gene fusion event, led to the development of Imatinib (Gleevec), a tyrosine kinase inhibitor that inhibits progression of chronic myelogenous leukemia (1–3). Fusion genes were subsequently discovered in solid tumors. Examples include *ALK*- and *NTRK*-related fusions that are found in patients with nonsmall cell lung cancer, e.g., *EML4–ALK* fusion, and in adult and pediatric tumors (4, 5). Kinase inhibitors such as Crizotinib target *ALK* fusions and transforming tyrosine kinase (TRK) inhibitors targeting neurotrophic receptor

[a]Department of Laboratory Medicine, Robert J. Tomsich Pathology and Laboratory Medicine Institute, Cleveland Clinic, Cleveland, OH; [b]Department of Pathology and Laboratory Medicine, The University of Kansas, Kansas City, KS; [c]Department of Pathology, Robert J. Tomsich Pathology and Laboratory Medicine Institute, Cleveland Clinic, Cleveland, OH; [d]CellNetix Pathology & Laboratories, Seattle, WA..
*Address correspondence to this author at: Department of Laboratory Medicine, LL2–139, Robert J. Tomsich Pathology & Laboratory Medicine Institute, Cleveland Clinic, 9500 Euclid Ave., Cleveland, OH 44195. Fax (216) 445-0681; e-mail chengy@ccf.org; yxc22@case.edu.
Received August 30, 2020; accepted October 30, 2020.
DOI: 10.1093/jalm/jfaa230

IMPACT STATEMENT

Gene fusions continue to be identified, at the molecular scale, to guide cancer diagnosis and personalized targeted therapy. Many clinical molecular laboratories are looking for guidance to establish a next-generation sequencing (NGS) -based method for gene fusion detection. Therefore, in this article, we described our approach of validating a customized, candidate gene-based NGS method to determining gene fusions in solid tumors.

tyrosine kinase gene (*NTRK1*, *NTRK2*, and *NTRK3*) have been FDA approved (6–8). The TRK inhibitors Larotrectinib and Entrectinib induce high response rates (>75%) in tumors harboring *NTRK* fusions (9, 10). In the context of soft-tissue tumors, the detection of fusion events aids in the diagnosis and management of patients with these (11–14), and the identification of gene fusions has become increasingly important in the diagnosis and management of patients with cancer.

The traditional methods used to detect gene rearrangements in the clinical laboratory include chromosome analysis, fluorescence in situ hybridization (FISH), reverse-transcriptase polymerase chain reaction (RT-PCR), and nucleic acid sequencing (15–17). While traditional methodologies can offer excellent sensitivity and specificity, as well as relatively straightforward interpretation, these methods lack scalability to include multiple targets and often require a sequential approach to clinical testing. This can be difficult in cases with limited tissue and in clinical scenarios requiring quick turn-around times. In addition, while FISH can be agnostic to translocation partner, RT-PCR approaches require foreknowledge of both partners and anticipated fusion breakpoints. With the quickly evolving clinical molecular testing landscape, next-generation sequencing (NGS) techniques provide the ability to do large scale parallel sequencing. NGS approaches, such as whole-transcriptome sequencing (RNA-Seq) and whole genome sequencing (WGS), enable the most comprehensive fusion gene testing (18). However, these approaches can be challenging with data analysis and interpretation, analytical validation, and difficulties detecting low-expression fusions by RNA-Seq alone. As such, although RNA-Seq and WGS have been shown to have clinical efficacy, they are currently not suitable for most clinical laboratories.

To determine fusion events without bioinformatic resources like those needed for genome-scale sequencing and data analysis, NGS-based candidate gene approaches have recently become available to interrogate multiple targets at one time (19–21). The Anchored Multiplex PCR (AMP™) technology is one of the methods targeting RNA transcript-derived cDNAs by gene-specific primers (GSP), followed by unidirectional extension to enrich potential fusion products, then amplification and sequencing (21). This approach is scalable, operates within a defined gene list for data analysis, and also captures both known and previously unreported fusion gene partners. Adapters used for sequencing library construction contain unique molecular barcodes and sample indices that enable the lessening of duplicated reads when samples are multiplex sequenced. In this article, we share our experience analytically validating a 34-gene panel that targets fusion events primarily in soft-tissue sarcomas, as well as

Table 1. TNA stability.

Specimen	Extraction date	Library prep date	Storage month	Fusion detected	Unique start sites ≥3	Unique reads ≥5	Percentage of reads supporting fusion ≥10	Average unique RNA start sites per gsp2 control ≥10	Percentage of RNA
28	8/15/2018	9/14/2018	1	*SS18–SSX1*	218	2257	50	150	62
28	8/15/2018	10/12/2018	2	*SS18–SSX1*	215	2855	54	146	64
67	8/9/2018	8/10/2018	0	*NAB2–STAT6*	173	5596	39	129	61
67	8/9/2018	10/8/2018	2	*NAB2–STAT6*	267	7478	57	121	60
45	8/20/2018	9/14/2018	1	*FUS–DDIT3*	11	31	78	55	57
45	8/20/2018	10/12/2018	2	*FUS–DDIT3*	25	68	15	59	57
45	8/20/2018	11/13/2018	3	*FUS–DDIT3*	41	109	15	57	74
3	8/10/2018	8/15/2018	0	*EWSR1–WT1*	110	614	39	29	43
3	8/10/2018	11/28/2018	3	*EWSR1–WT1*	106	409	29	56	48

other tumors that may exhibit *NTRK* fusions. The observation of fusions, or lack thereof, may aid in diagnosis, prognosis, patient management, and therapy.

MATERIALS AND METHODS

Panel Design

Candidate genes and specific exons were chosen targeting those fusions that occur most commonly in soft-tissue sarcomas and via extensive literature review (3, 22). The GSPs were designed using Archer designer software based on the known fusion break points and the gene product orientations. Thirty-four genes, including 182 exons, were included in the customized fusion gene panel (Table 1 in the online Data Supplement).

Sample Selection

Clinical specimens of FFPE resections with known fusion gene status were selected from the Cleveland Clinic Pathology Department archive system. Criteria for choosing a diverse pool of validation samples were based on (*a*) known fusion events with both host and partner fusion genes identified, (*b*) >20% tumor cellularity, and (*c*) for samples with previously determined FISH or IHC results, only true positive cases were selected. Samples with ambiguous or borderline FISH or IHC results were excluded (except one, see Discussion). The IHC positive cases were chosen based on well-documented histology and antibody staining characterization of high sensitivity and specificity described in the literature. Control FFPE reference materials HD796 (fusion positive control) and HD783 (fusion negative control) were purchased from Horizon Discovery. The HD796 possesses 4 fusion gene products that are targeted in our NGS Panel. The concentrations of these fusions *EML4–ALK*, *SLC34A2–ROS1*, *TPM3–NTRK1*, and *ETV6–NTRK3* are at 3.9, 4.2, 14.7, and 12.0 copies/ng total extracted RNA, respectively, based on droplet digital PCR measurement.

Nucleic Acid Extraction and Quality Control

Total nucleic acid (TNA) was extracted using the Maxwell RSC RNA FFPE kit (Promega). Extracted TNA was stored at −70 °C or less. The ReliaPrep™ FFPE Total RNA Miniprep System (Promega) was validated as a backup, manual extraction method.

Quality determination of extracted TNA was assessed using iTAQ™ Universal SYBR® Green Supermix (BIO-RAD Laboratories). It was determined that reliable fusion NGS results could be obtained when the LightCycler PreSeq crossing point values <30; optimally 200 ng TNA was employed in library preparation.

NGS Library Preparation and Sequencing

Library preparation was performed according to manufacturer-recommended protocols. Briefly, cDNA was made from the extracted TNA, followed by end-repair, barcoded adapter ligation, and 2 rounds of nested-PCR amplification using GSPs and reagents obtained from ArcherDX. All libraries were purified, qualified, pooled, and then sequenced on Illumina MiSeq instruments. Sequencing run quality indicators of MiSeq v.3 chemistry were tracked using an Illumina Sequencing Analysis Viewer. To increase sequencing diversity, 5% denatured 10 pM PhiX was added to each final library pool. The sequencing reads of >Q30 score ranged from 83.9 to 90.6%. The cluster passing filter range was 80.1–94.9%. To ensure no cross contamination during library preparation, a "no template control" (NTC) of molecular grade water was included with every batch of library preparation. Any NTC with a final library concentration >3.0 ng/μL was sequenced. To be considered in range, sequenced NTC needed to demonstrate quality metrics well below the clinical sample library quality metrics with no fusion detected. The fusion gene scoring criteria are as follows: (*a*) number of unique fusion gene reads ≥5; (*b*) unique fusion gene start sites ≥3; and (*c*) percentage of fusion gene reads ≥10%.

NGS Data Analysis

Bcl2fastq program (Illumina) was used to extract raw sequencing reads and generate fastq files for each specimen. Read deduplication using the unique molecular barcode, alignment to the hg19 reference genome, and identification of fusion events were performed using Archer Analysis software v.5.1.3. The resultant analytical data were processed and displayed using in-house developed bioinformatics protocols for quality control. In general, the following quality metrics were used to determine sequencing quality of each sample: (*a*) total read fragments ≥500 000; (*b*) total RNA read percentage ≥30%; and (*c*) RNA average unique start sites per GSPS2 control ≥10. Identified fusion events were expected to meet the following criteria: (*a*) ≥10% fusion read percentage; (*b*) ≥3 RNA unique start site; and (*c*) ≥5 unique RNA reads.

RESULTS

Total Nucleic Acid (TNA) Stability and Input

TNA stability was assessed by storing 4 specimens for extended amounts of time, with up to 3 (*FUS–DDIT3*) freeze–thaw cycles, to determine the effect of thawing frozen TNA after varying storage lengths on test performance. The goal was to have at least 2 TNA samples tested at each of the storage times (i.e., 0, 1, 2, and 3 months). The fusion genes were successfully detected when TNA had been stored up to 3.5 months and thawed up to 3 times (see Table 1). In some samples, the observation of increased unique start sites and reads may be attributed to variations in library preparation (e.g., pipetting techniques) or freeze–thaw caused slight degradation of TNA. Nevertheless, these NGS quality metrics for the defined fusion genes were of the same orders of magnitude comparing the fresh and thawed specimens, which indicate test performance was satisfactory.

We defined the suboptimal amount of TNA input for NGS library preparation. In Table 2, samples were ordered in decreasing TNA input amount. Concordant results were observed between the current test and orthogonal methods in all but 3 cases (e.g., 36, 61, and 86), where known

Table 2. Suboptimal TNA input.

Sample no. and input amount	Fusion detected	Unique start sites ≥3	Unique reads ≥5	Percentage of reads supporting fusion ≥10	Average unique RNA start sites per GSP2 control ≥10	Percentage RNA	PreSeq Ct <30	Previously reported result	Result concordance
40_194 ng	*EWSR1–ZNFF444*	14	28	26.2	11	27	Yes	*EWSR1–ZNFF444* (NGS)	Yes
11_176 ng	*NFD*	na	na	na	42	37	Yes	*NFD* (NGS)	Yes
54_134 ng	*EWSR1–DDIT3*	6	21	17.8	29	29	Yes	*DDIT3* (break-apart FISH)	Yes
58_117 ng	*EWSR1–NR4A3*	11	30	66.7	4	26	No	*EWSR1* (break-apart FISH)	Yes
26_114 ng	*EWSR1–CREB1*	52	178	37.9	60	34	Yes	*EWSR1* (break-apart FISH)	Yes
12_108 ng	*NFD*	na	na	na	45	51	Yes	*NFD* (NGS)	Yes
86_100 ng	*FUS–DDIT3*	6	7	13.2	13.9	60	Yes	*DDIT3* (break-apart FISH)	Yes
100_100 ng	*LMNA–NTRK1*	261	4254	89.4	138	68	Yes	*LMNA–NTRK1* (NGS)	Yes
101_100 ng	*CIC–DUX4*	126	182	63.2	144	69	Yes	*CIC–DUX4* (NGS)	Yes
3_80 ng	*EWSR1–WT1*	106	409	29.4	56	48	Yes	*EWSR1–WT1* (NGS)	Yes
36_80 ng	*NFD*	na	na	na	8	48	No	*EML4–ALK* (NGS)	No
61_79 ng	*NFD*	na	na	na	1.6	26	No	*DDIT3* (break-apart FISH)	No
80_79 ng	*FUS–DDIT3*	25	145	31.9	16	46	Yes	*DDIT3* (break-apart FISH)	Yes
74_69 ng	*FUS–DDIT3*	13	17	22.7	41	39	Yes	*DDIT3* (break-apart FISH)	Yes
55_68 ng	*EWSR1–FLI1*	9	13	31.7	6	49	Yes	*EWSR1* (break-apart FISH)	Yes
74_69 ng	*FUS–DDIT3*	13	17	22.7	41	39	Yes	*DDIT3* (break-apart FISH)	Yes
86_50 ng	*NFD*	na	na	an	6.1	60	No	*DDIT3* (break-apart FISH)	No
101_50 ng	*CIC–DUX4*	7	8	4.8	171	68	Yes	*CIC–DUX4* (NGS)	Yes

NFD: no fusion detected.

gene fusions or rearrangements were no longer identifiable using the anchor-based NGS approach. These 3 discordant cases all have 80 ng or less TNA input, contrary to other samples of 100 ng or higher TNA input that show 100% concordant test results. Particularly, samples 86 and 101 were performed in 2 different TNA inputs (i.e., 50 and 100 ng) to determine the variation of sequencing quality metrics due to TNA inputs. Our data show at 50 ng comparing to 100 ng inputs for sample 101, a drastic decrease of the detectable gene fusion sequencing quality such as unique start sites, unique reads, and percentage reads. Furthermore, for sample 86, no fusion detected at the 50 ng TNA input while a *FUS–DDIT3* fusion was identified at the 100-ng input. This observation indicates that reduced TNA input may contribute to false-negative test results, and at least 100 ng of TNA input should be considered for the presented fusion detecting method.

Accuracy

Accuracy studies were undertaken with newly extracted TNA from 87 FFPE specimens that had been previously tested in clinical laboratory settings using various fusion detection methods. Five (14, 18, 34, 36, 61) of the 87 samples were not included in the accuracy assessment because their quality metrics (Materials and Methods section) were unsatisfactory. Library preparation and sequencing were performed on the remaining 82 samples, where gene fusion status was previously determined via molecular or antibody staining methods: 29 by NGS, 43 by interphase FISH, 9 by IHC, and one by RT-PCR. Three discordant results (samples 39, 62, and 88) between the fusion NGS Panel and previously reported findings were observed (see Table 3 and Supplemental Table 2).

The specimen, sample 39, initially demonstrated a *SS18–SSX4* fusion in this validation exercise that is different from the previously reported *SS18–SSX2* fusion; it was thus considered a discordant case. Because the *SSX1*, *SSX2*, *and SSX4* genes may present a challenge in sequencing read alignment, an updated bioinformatics pipeline Archer Analysis software v.6.0.3.2 was applied to reanalyze all the validation samples, including sample 39, to more accurately align *SSX* related genes. Indeed, the updated pipeline correctly identified *SS18–SSX2* fusion in sample 39, while the fusion gene status of the remaining validation samples showed no changes. Thus, the discordant result was resolved. Another specimen, sample 62, was previously reported positive for *EWSR1* translocation by break-apart FISH testing, but no fusion was detected using the NGS Panel. Upon further review, the original FISH report noted that 82% of the cells in the sample displayed monosomy 22 with concomitant loss of the *EWSR1* gene locus, and <20% of the cells harbored the translocation event. Because different sections of the FFPE tissue block were cut for FISH and the fusion NGS test, the tumor heterogeneity and the test sensitivity may explain the disparate results between 2 different testing platforms (see Analytical Sensitivity). There was one false positive specimen identified (sample 88; see Discussion). Thus, the resolved accuracy observed during this validation exercise was 97.6% (80/82). The analytical performance of sensitivity and specificity were calculated at 98.7 and 90.0%, respectively, by comparing fusion variants detected by the customized fusion NGS Panel to previously reported results. The calculation of analytical sensitivity and specificity is shown in the Supplemental Table 3.

Precision

TNA extracted from 4 FFPE specimens that contained different fusion genes were tested for intra- and interrun reproducibility, i.e., precision. Samples with *SS18–SSX1*, *EWSR1–FLI1*, and *TFE3–DVL2* fusions were tested in triplicate in the same run (intrarun) with 3 different library preparations and different barcodes. The identical fusions were

Table 3. Accuracy.

Concordance	Sample number	Fusion detected by NGS	Fusion detected by orthogonal methods
Y	1	*SS18–SSX4*	*SS18–SSX* (RT-PCR)
Y	3	*EWSR1–WT1*	*EWSR1–WT1* (NGS)
Y	4	*ASPSCR1–TFE3*	*ASPSCR1–TFE3* (NGS)
Y	5	*WWTR1–CAMTA1*	*WWTR1–CAMTA1* (NGS)
Y	6	*EWSR1–ATF1*	*EWSR1–ATF1* (NGS)
Y	7	*PAX3–FOXO1*	*PAX3–FOXO1* (NGS)
Y	8	*YAP1–TFE3*	*TFE3* (break-apart FISH)
Y	9	*COL1A1–PDGFB*	*COL1A1–PDGFB* (NGS)
Y	10	*NAB2–STAT6*	*STAT6* (IHC)
Y	11	*NFD*	*NFD* (NGS)
Y	12	*NFD*	*NFD* (NGS)
Y	13	*EWSR1–FLI1*	*EWSR1–FLI1* (NGS)
Y	15	*EWSR1–ATF1*	*EWSR1–ATF1* (NGS)
Y	16	*EWSR1–ATF1*	*EWSR1–ATF1* (NGS)
Y	17	*PAX3–FOXO1*	*PAX3–FOXO1* (NGS)
Y	19	*SS18–SSX1*	*SS18–SSX1* (NGS)
Y	20	*EWSR1–ATF1*	*EWSR1–ATF1* (NGS)
Y	21	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)
Y	23	*ACTB–FOSB*	*ACTB–FOSB* (NGS)
Y	24	*SS18–SSX4*	*SS18* (break-apart FISH)
Y	25	*EWSR1–ERG*	*EWINGS* (break-apart FISH)
Y	26	*EWSR1–CREB1*	*EWSR1* (break-apart FISH)
Y	27	*YAP1–TFE3*	*TFE3* (break-apart FISH)
Y	28	*SS18–SSX1*	*SS18* (break-apart FISH)
Y	29	*ACTB–GLI1*	*ACTB–GLI1* (NGS)
Y	30	*WWTR1–CAMTA1*	*WWTR1–CAMTA1* (NGS)
Y	32	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)
Y	33	*NFD*	*NFD* (NGS)
Y	35	*ETV6–NTRK3*	*ETV6–NTRK3* (NGS)
Y	37	*NFD*	*NFD* (NGS)
Y	38	*EWSR1–WT1*	*EWSR1–WT1* (NGS)
N	39	*SS18–SSX4*	*SS18–SSX2* (NGS)
Y	40	*EWSR1–ZNFFF*	*EWSR1–ZNFFF* (NGS)
Y	41	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)
Y	43	*SS18–SSX4*	*SS18* (break-apart FISH)
Y	44	*EWSR1–ATF1*	*EWSR1* (break-apart FISH)
Y	45	*FUS–DDIT3*	*DDIT3* (break-apart FISH)
Y	46	*EWSR1–ATF1*	*EWSR1* (break-apart FISH)
Y	47	*FUS–DDIT3*	*DDIT3* (break-apart FISH)

Continued

Table 3. (continued)

Concordance	Sample number	Fusion detected by NGS	Fusion detected by orthogonal methods
Y	48	*EWRS1–NR4A3*	*NR4A3* (break-apart FISH)
Y	49	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)
Y	50	*CIC–DUX4*	*CIC–DUX4* (NGS)
Y	51	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)
Y	52	*EWSR1–ERG*	*EWSR1* (break-apart FISH)
Y	53	*EWSR1–NR4A3*	*NR4A3* (break-apart FISH)
Y	54	*EWSR1–DDIT3*	*DDIT3* (break-apart FISH)
Y	55	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)
Y	56	*FUS–DDIT3*	*DDIT3* (break-apart FISH)
Y	57	*EWSR1–CREB3L1*	*MUC4* (IHC)
Y	58	*EWSR1–NR4A3*	*EWSR1* (break-apart FISH)
Y	59	*NAB2–STAT6*	*STAT6* (IHC)
Y	60	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)
N	62	*NFD*	*EWSR1* (break-apart FISH)
Y	63	*EWSR1–ATF1*	*EWSR1* (break-apart FISH)
Y	64	*EWSR1-ERG*	*EWSR1* (break-apart FISH)
Y	65	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)
Y	66	*PAX3–FOXO1*	*FOXO1* (break-apart FISH)
Y	67	*NAB2–STAT6 Intronic*	*STAT6* (IHC)
Y	68	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)
Y	70	*PAX–FOXO1*	*FOXO1* (break-apart FISH)
Y	71	*SQSRM1–ALK*	*ALK* (IHC)
Y	72	*NAB2–STAT6*	*NAB2–STAT6* (NGS)
Y	73	*EML4–ALK*	*ALK* (IHC)
Y	74	*FUS–DDIT3*	*DDIT3* (break-apart FISH)
Y	75	*NAB2–STAT6*	*STAT6* (IHC)
Y	76	*SS18–SSX1*	*SS18* (break-apart FISH)
Y	78	*SS18–SSX1*	*SS18* (break-apart FISH)
Y	79	*SS18–SSX1*	*SS18* (break-apart FISH)
Y	80	*FUS–DDIT3*	*DDIT3* (break-apart FISH)
Y	81	*FUS–DDIT3*	*DDIT3* (break-apart FISH)
Y	82	*NFD*	*NFD* (break-apart FISH)
Y	83	*DVL2–TFE3*	*TFE3* (break-apart FISH)
Y	84	*TFE3–PRCC*	*TFE3* (break-apart FISH)
Y	86	*FUS–DDIT3*	*DDIT3* (break-apart FISH)
Y	87	*EWSR1–FLI1*	*EWSR1–FLI1* (NGS)
Y	88	*EWSR1–CREB3L2*	*MUC-4* (IHC)
False positive		*WWTR1–CAMTA1*	
Y	89	*WWRT1–CAMTA1*	*CAMTA1* (IHC)
Y	90	*PAX3–FOXO1*	*FOXO1* (break-apart FISH)

Continued

Table 3. (continued)

Concordance	Sample number	Fusion detected by NGS	Fusion detected by orthogonal methods
Y	97	*EML4–ALK*	*EML4–ALK* (NGS)
Y	98	*EWSR1–NR4A3*	*EWSR1* (break-apart FISH)
Y	100	*LMNA–NTRK1*	*LMNA–NTRK1* (NGS)
Y	101	*CIC–DUX4*	*CIC–DUX4* (NGS)

NFD: no fusion detected.

detected each time, i.e., 100% reproducibility and concordance (9/9; see Table 4 and Supplemental Table 4). Samples with *EWSR1–FLI1*, *SS18–SSX1*, and *FUS–DDIT3* fusions were tested in 3 separate runs (interrun) on different days using 3 separate libraries prepared by 3 different technologists. Interrun assessment was 100% reproducible and concordant (9/9; see Table 5 and Supplemental Table 5). Similar quality metrics and read statistics were observed in both intra- and interruns precision measurements.

Additionally, a commercially available FFPE reference standard, HD796, known to harbor several fusion genes was tested in 18 independent runs. This reference material contains 4 known fusion genes (*TPM3–NTRK1*, *SLC34A2–ROS1*, *ETV6–NTRK3*, and *EML4–ALK*) that were designed into our NGS Panel. All 4 fusions were accurately identified (72/72) during each test run with similar quality metrics and read statistics observed (Supplemental Table 6). The consistent identification of these fusion genes over a 3-month period indicates that test performance is highly reproducible.

Limit of Detection

We further diluted the HD796 fusion positive control with its counterpart HD783 fusion negative reference material at 20, 10, 5, and 2.5% composition (weight/weight) to determine the lowest amount of detectable fusion transcripts using this panel. In this limit of detection (LoD) study, all 4 fusion genes (*TPM3–NTRK1*, *SLC34A2–ROS1*, *ETV6–NTRK3*, and *EML4–ALK*) were identified at 10% dilution in various copies of transcripts (Table 6). At 5 and 2.5% dilution levels, the *EML4–ALK* fusions were not detected even when visual inspection was performed on the sequencing pile-up data. The identical LoD study results have been obtained in several replicates (data not shown). Note that at the 10% dilution, the *EML4–ALK* fusion was detected (unique fusion start sites = 5, unique fusion reads = 5) near the fusion gene scoring criteria cut-off (see Materials and Methods) despite the fact that it was calculated at 78 copies. When the sample was further diluted to the 5% level, *EML4–ALK* fusion is no longer identifiable at 39 copies while *SLC34A2–ROS1* fusion is still detected (at 42 copies). Overall, our data have shown a specific fusion product is detectable when its fusion transcripts is at or >60 copies in a reaction, while it may not be detected if the targeted fusion transcripts are <40 copies. Although these varied detection sensitivities may be reasoned by technologist pipetting techniques, they are most likely explained by the differences in fusion gene sequence contexts that affect the gene-specific primer annealing and the subsequent amplification efficiency. For a fusion product where either the partner gene or the corresponding tumor type has not been reported previously, along with its quality scoring being near the established cut-off, orthogonal testing is recommended for confirmation.

Table 4. Intrarun MiSeq.

Sample number	Fusion detected by NGS	Fusion detected by orthogonal methods	Library made	Library tested
28-2	*SS18–SSX1*	*SS18* (break-apart FISH)	9/11/2018	9/14/2018
28-3	*SS18–SSX1*	*SS18* (break-apart FISH)	9/11/2018	9/14/2018
28-4	*SS18–SSX1*	*SS18* (break-apart FISH)	9/11/2018	9/14/2018
49-2	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)	11/13/2018	11/26/2018
49-3	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)	11/13/2018	11/26/2018
49-4	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)	11/13/2018	11/26/2018
83-2	*TFE3–DVL2*	*TFE3* (break-apart FISH)	11/13/2018	11/26/2018
83-3	*TFE3–DVL2*	*TFE3* (break-apart FISH)	11/13/2018	11/26/2018
83-4	*TFE3–DVL2*	*TFE3* (break-apart FISH)	11/13/2018	11/26/2018

Table 5. Interrun MiSeq.

Sample number	Fusion detected by NGS	Fusion detected by orthogonal methods	Library made	Library tested
32-1	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)	8/22/2018	8/27/2018
32-2	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)	10/18/2018	11/5/2018
32-3	*EWSR1–FLI1*	*EWSR1* (break-apart FISH)	11/13/2018	11/26/2018
28-1	*SS18–SSX1*	*SS18* (break-apart FISH)	8/22/2018	8/27/2018
28-2	*SS18–SSX1*	*SS18* (break-apart FISH)	10/9/2018	10/12/2018
28-3	*SS18–SSX1*	*SS18* (break-apart FISH)	10/18/2018	11/5/2018
45-1	*FUS–DDIT3*	*DDIT3* (break-apart FISH)	9/11/2018	9/14/2018
45-2	*FUS–DDIT3*	*DDIT3* (break-apart FISH)	10/9/2018	10/12/2018
45-3	*FUS–DDIT3*	*DDIT3* (break-apart FISH)	11/13/2018	11/26/2018

DISCUSSION

Our study has demonstrated that the anchor-based NGS method is a robust test to determine gene fusions in soft-tissue sarcoma. We have shown the extracted total nucleic acids that are used as starting materials for this test have a stability of up to 3.5 months storage at −20 °C with up to 3 freeze and thaw cycles. This knowledge suits routine clinical laboratory operation well, where samples may be collected and batch-processed within a few weeks of collection. It also provides the opportunity to use stored TNAs for future test validation in the event that additional fusion candidates are required for the NGS Panel update. We have also shown that the optimal total TNA input amount should not be <100 ng for a given specimen. Less than the defined TNA input may lead to no or even false-negative test results. For example, in our laboratory, samples with a TNA amount <100 ng are considered suboptimal for proceeding with the assay.

We have demonstrated an overall LoD for this fusion NGS test; however, the exact LoD for each fusion event may be challenging to determine accurately because of variability in RNA expression.

Table 6. Limit of detection.

Specimen	Fusion detected	Calculated transcript copies	Unique start sites ≥3	Unique reads ≥5	Percentage of reads supporting fusion ≥10	Average unique RNA start sites per GSP2 control ≥10
HD796	TPM3(6)-NTRK1(9)	2940	109	290	85	85
	ETV6(5)-NTRK3(14)	2400	77	189	86	85.9
	SLC34A2(4)-ROS1(32)	840	79	186	94	93.5
	EML4(12)-ALK(20)	780	29	60	92	92.3
HD783	No Fusion Detected	0	na	na	na	249
20% Dilution	TPM3(6)-NTRK1(9)	588	52	88	56	283
	ETV6(5)-NTRK3(14)	480	43	61	79	283
	SLC34A2(4)-ROS1(32)	168	36	53	23	283
	EML4(12)-ALK(20)	156	16	24	80	283
10% Dilution	TPM3(6)-NTRK1(9)	294	25	28	32	249
	ETV6(5)-NTRK3(14)	240	24	31	91	249
	SLC34A2(4)-ROS1(32)	84	21	24	96	249
	EML4(12)-ALK(20)	78	5	5	71	249
5% Dilution	TPM3(6)-NTRK1(9)	147	15	19	26	241
	ETV6(5)-NTRK3(14)	120	11	11	65	241
	SLC34A2(4)-ROS1(32)	42	14	17	100	241
	EML4(12)-ALK(20)	39	not detectable	not detectable	not detectable	241
2.5% Dilution	TPM3(6)-NTRK1(9)	74	6	7	13	241
	ETV6(5)-NTRK3(14)	60	5	6	75	241
	SLC34A2(4)-ROS1(32)	21	5	5	100	241
	EML4(12)-ALK(20)	19	not detectable	not detectable	not detectable	241

Basal level transcription typically varies widely among different genes, as well as for the same gene in different tissues. Therefore, unlike a diploid human genome that contains 2 copies of each gene, there is no defined baseline for each targeted gene's transcription level unless matched normal tissues and quantitative or digital PCR methods are employed. One could extrapolate the sensitivity study results to estimate the lowest percentage of tumor cellularity needed for the LoD, by assuming HD796 (fusion positive) control of 100% tumor cellularity and diluent HD783 (fusion negative) of 0% tumor cells. Because the assumption of HD796 of 100% tumor cellularity is very likely an over-estimation, a 10% dilution of HD796, in fact, means that <10% of tumor cells are actually present in the testing condition. Since all 4 fusion genes (*TPM3–NTRK1*, *SLC34A2–ROS1*, *ETV6–NTRK3*, and *EML4–ALK*) were identified at a 10% dilution point (Table 6), this finding indicates that the presented method may identify fusion products in <10% of tumor cells in a given FFPE tissue biopsy. Based on these assumptions, our data support the notion that fusion gene LoD can conservatively be achieved at 20% tumor content. In our laboratory, most soft-tissue tumors have

adequate tissue and tumor cellularity (usually >20%) to extract sufficient TNA for library preparation and sequencing.

One discrepant result was observed in the accuracy study, with both *EWSR1–CREB3L2* and *WWTR1–CAMTA1* fusions detected in specimen sample 88 (Table 3). This specimen is derived from a low-grade fibromyxoid sarcoma (LGFMS) and was documented as *MUC4* positive by IHC testing, which is a highly sensitive (100%) and specific diagnostic marker for LGFMS (23). Because the most common (>90%) gene fusion seen in LGFMS involves the *CREB3L2* gene, the identification of abundant *EWSR1–CREB3L2* fusion (146 start sites, 843 unique reads, and 39% of reads supporting the fusion) corroborates the LGFMS classification (24). The *WWTR1–CAMTA1* fusion, however, was present at a minimal level with 5 start sites and 5 unique reads. Since a single driver fusion is typically present in LGFMS, the possibility of a second, low-level fusion present in sample 88 warrants further investigation using *ERG* and *CAMTA* antibody staining. Interestingly, the IHC testing demonstrated negative results, indicating (*a*) the *WWTR1–CAMTA1* fusion was likely a false positive signal or (*b*) the extremely low-level fusion products were undetectable, perhaps due to tumor heterogeneity, in the FFPE tissue resections used for IHC (25).

It is worth noting that various genomic rearrangements may be discordant between fusion NGS and FISH results. For example, using the EWSR1 FISH test in identifying EWSR1 rearrangement has been well documented (25). It has been shown in Ewing sarcoma that *EWSR1–FLI1* represents the primary fusion event (~80%), and that *EWSR1–ATF1* and *EWSR1–ERG* fusions occur secondarily (~10% each) (26, 27). Most *EWSR1* gene rearrangement events generally involve a balanced translocation between *EWSR1* and the partner gene on a different chromosome. Thus, the break-apart *EWSR1* FISH probe can readily identify these types of fusion events as Ewing cases (Table 3). However, for atypical rearrangement cases, i.e., the so-called "cryptic" *EWSR1–ERG* fusions, this small *EWSR1* genomic rearrangement does not usually interfere with the ~100 kb FISH probe hybridization. Therefore, the *EWSR1* FISH break-apart probe does not separate or break apart and may fail to identify this fusion because the cryptic locus is a small region in the *EWSR1* gene. Nevertheless, because the fusion NGS test is designed to detect fused RNA transcripts that resulted from gene structure rearrangement, the mentioned small rearranged genomic region does not interfere with fusion identification at the RNA level. We recently tested 2 Ewing sarcoma cases where EWSR1 FISH break-apart negativity and NGS fusion positivity were observed, which corroborates with the idea of identifying gene fusions at the RNA level without interference of a cryptic DNA rearrangement. Conversely, the exchange of promoter regions between 2 genomic loci or promoter swapping can be easily revealed by break-apart FISH probes in general. Although this event can introduce a strong promoter to the new target for the gene product overexpression, the lack of actual fusion transcripts from 2 separated gene coding regions would not make the NGS fusion detection possible at the RNA level.

In summary, a customized NGS fusion gene panel, with content chosen based on careful relevant literature review, was designed and analytically validated. Results demonstrated robust and acceptable analytical performance for introduction into a clinical molecular pathology service as a laboratory developed test. Furthermore, NGS-based testing may be more sensitive than FISH for identification of *EWSR1*-related fusion events.

SUPPLEMENTAL MATERIAL

Supplemental material is available at *The Journal of Applied Laboratory Medicine* online.

Nonstandard Abbreviations: FISH, fluorescence in situ hybridization; IHC, immunohistochemistry; NGS, next-generation sequencing; FFPE, formalin-fixed paraffin-embedded; TRK, tyrosine kinase; NTC, no template control; TNA, total nucleic acid; LoD, limit of detection.

Human Genes: *BCR*, BCR activator of RhoGEF and GTPase; *ABL1*, ABL proto-oncogene 1, non-receptor tyrosine kinase; *EML4*, EMAP like 4; *ALK*, ALK receptor tyrosine kinase; *SLC34A2*, solute carrier family 34 member 2; *ROS1*, ROS proto-oncogene 1, receptor tyrosine kinase; *TPM3*, tropomyosin 3; *NTRK1*, neurotrophic receptor tyrosine kinase 1; *ETV6*, ETS variant transcription factor 6; *NTRK3*, neurotrophic receptor tyrosine kinase 3; *FUS*, FUS RNA binding protein; *DDIT3*, DNA damage inducible transcript 3; *SS18*, SS18 subunit of BAF chromatin remodeling complex; *SSX4*, SSX family member 4; *SSX2*, SSX family member 2; *SSX1*, SSX family member 1; *EWSR1*, EWS RNA binding protein 1; *FLI1*, Fli-1 proto-oncogene, ETS transcription factor; *TFE3*, transcription factor binding to IGHM enhancer 3; *DVL2*, dishevelled segment polarity protein 2; *CREB3L2*, cAMP responsive element binding protein 3 like 2; *WWTR1*, WW domain containing transcription regulator 1; *CAMTA1*, calmodulin binding transcription activator 1.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.*

B.P. Rubin, provision of study material or patients.

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:* **Employment or Leadership:** None declared. **Consultant or Advisory Role:** B.P. Rubin, Caris Life Sciences, Avalon Healthcare Solutions. **Stock Ownership:** None declared. **Honoraria:** Y.-W. Cheng, River West Meeting Association, Inc. **Research Funding:** None declared. **Expert Testimony:** None declared. **Patents:** None declared.

Role of Sponsor: No sponsor was declared.

REFERENCES

1. Potter M. Brief historical sketch of chromosomal translocations and tumors. J Natl Cancer Inst Monogr 2008;2008:2–7.
2. Soverini S, Mancini M, Bavaro L, Cavo M, Martinelli G. Chronic myeloid leukemia: the paradigm of targeting oncogenic tyrosine kinase signaling and counteracting resistance for successful cancer therapy. Mol Cancer 2018;17:49.
3. Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. Nat Rev Cancer 2015;15:371–81.
4. Ducray SP, Natarajan K, Garland GD, Turner SD, Egger G. The transcriptional roles of ALK fusion proteins in tumorigenesis. Cancers (Basel) 2019;11:1074.
5. Solomon JP, Linkov I, Rosado A, Mullaney K, Rosen EY, Frosina D, et al. NTRK fusion detection across multiple assays and 33,997 cases: diagnostic implications and pitfalls. Mod Pathol 2020;33:38–46.
6. Kazandjian D, Blumenthal GM, Chen HY, He K, Patel M, Justice R, et al. FDA approval summary: crizotinib for the treatment of metastatic non-small cell lung cancer with anaplastic lymphoma kinase rearrangements. Oncologist 2014;19:e5-11.
7. Scott LJ. Larotrectinib: first global approval. Drugs 2019; 79:201–6.
8. Cocco E, Scaltriti M, Drilon A. NTRK fusion-positive cancers and TRK inhibitor therapy. Nat Rev Clin Oncol 2018;15:731–47.
9. Laetsch TW, Hawkins DS. Larotrectinib for the treatment of TRK fusion solid tumors. Expert Rev Anticancer Ther 2019;19:1–10.
10. Drilon A, Nagasubramanian R, Blake JF, Ku N, Tuch BB, Ebata K, et al. A next-generation TRK kinase inhibitor overcomes acquired resistance to prior TRK kinase inhibition in patients with TRK fusion-positive solid tumors. Cancer Discov 2017;7:963–72.
11. Hoffner B, Elias AD, Villalobos VM. Targeted therapies in the treatment of sarcomas. Targ Oncol 2018;13:557–65.
12. Xiao X, Garbutt CC, Hornicek F, Guo Z, Duan Z. Advances in chromosomal translocations and fusion genes in sarcomas and potential therapeutic applications. Cancer Treat Rev 2018;63:61–70.
13. Schaefer IM, Cote GM, Hornick JL. Contemporary sarcoma diagnosis, genetics, and genomics. J Clin Oncol 2018;36:101–10.
14. Demetri GD, Blay JY, Casali PG. Advances and controversies in the management of soft tissue sarcomas. Future Oncol 2017;13:3–11.
15. Cantile M, Marra L, Franco R, Ascierto P, Liguori G, De Chiara A, Botti G. Molecular detection and targeting of EWSR1 fusion transcripts in soft tissue tumors. Med Oncol 2013;30:412.
16. Tai HC, Chuang IC, Chen TC, Li CF, Huang SC, Kao YC, et al. NAB2-STAT6 fusion types account for

clinicopathological variations in solitary fibrous tumors. Mod Pathol 2015;28:1324–35.
17. Agaram NP, Zhang L, Sung YS, Singer S, Antonescu CR. Extraskeletal myxoid chondrosarcoma with non-EWSR1-NR4A3 variant fusions correlate with rhabdoid phenotype and high-grade morphology. Hum Pathol 2014;45:1084–91.
18. Koeppel F, Bobard A, Lefebvre C, Pedrero M, Deloger M, Boursin Y, et al. Added value of whole-exome and transcriptome sequencing for clinical molecular screenings of advanced cancer patients with solid tumors. Cancer J 2018;24:153–62.
19. Levy MA, Santos S, Kerkhof J, Stuart A, Aref-Eshghi E, Guo F, et al. Implementation of an NGS-based sequencing and gene fusion panel for clinical screening of patients with suspected hematologic malignancies. Eur J Haematol 2019;103:178–89.
20. Sussman RT, Oran AR, Paolillo C, Lieberman D, Morrissette JJD, Rosenbaum JN. Validation of a next-generation sequencing assay targeting RNA for the multiplexed detection of fusion transcripts and oncogenic isoforms. Arch Pathol Lab Med 2020;144:90–8.
21. Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, et al. Anchored multiplex PCR for targeted next-generation sequencing. Nat Med 2014;20:1479–84.
22. Mertens F, Antonescu CR, Mitelman F. Gene fusions in soft tissue tumors: recurrent and overlapping pathogenetic themes. Genes Chromosomes Cancer 2016;55:291–310.
23. Doyle LA, Moller E, Dal Cin P, Fletcher CD, Mertens F, Hornick JL. MUC4 is a highly sensitive and specific marker for low-grade fibromyxoid sarcoma. Am J Surg Pathol 2011;35:733–41.
24. Mohamed M, Fisher C, Thway K. Low-grade fibromyxoid sarcoma: Clinical, morphologic and genetic features. Ann Diagn Pathol 2017;28:60–7.
25. Flucke U, Vogels RJ, de Saint Aubain Somerhausen N, Creytens DH, Riedl RG, van Gorp JM, et al. Epithelioid hemangioendothelioma: clinicopathologic, immunhistochemical, and molecular genetic analysis of 39 cases. Diagn Pathol 2014;9:131.
26. Lessnick SL, Ladanyi M. Molecular pathogenesis of Ewing sarcoma: new therapeutic and transcriptional targets. Annu Rev Pathol Mech Dis 2012;7:145–59.
27. Wang WL, Patel NR, Caragea M, Hogendoorn PC, Lopez-Terrada D, Hornick JL, Lazar AJ. Expression of ERG, an ETS family transcription factor, identifies ERG-rearranged Ewing sarcoma. Mod Pathol 2012;25:1378–83.