ORIGINAL ARTICLE

Development and validation of an RNA sequencing panel for gene fusions in soft tissue sarcoma

Wanming Hu[1] | Li Yuan[2] | Xinke Zhang[1] | Yang Ni[3,4] | Dongchun Hong[5] |
Zhicai Wang[6] | Xiaomin Li[3] | Yuan Ling[3] | Chao Zhang[3] | Wanglong Deng[3] |
Minqi Tian[3] | Ran Ding[3] | Chao Song[3,4] | Jianmin Li[7] | Xing Zhang[5]

[1]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Department of Pathology, Sun Yat-sen University Cancer Center, Guangzhou, China

[2]Department of Pathology, Guangzhou Women and Children's Medical Center, Guangzhou, China

[3]State Key Laboratory of Translational Medicine and Innovative Drug Development, Jiangsu Simcere Diagnostics Co., Ltd, Nanjing, China

[4]Department of Medicine, Nanjing Simcere Medical Laboratory Science Co., Ltd, Nanjing, China

[5]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Department of Medical Melanoma and Sarcoma, Sun Yat-sen University Cancer Center, Guangzhou, China

[6]Department of General Surgery, Jiangsu Cancer Hospital & Jiangsu Institute of Cancer Research &, The Affiliated Cancer Hospital of Nanjing Medical University, Nanjing, China

[7]Department of Orthopedics, Qilu Hospital of Shandong University, Jinan, China

Correspondence
Chao Song, State Key Laboratory of Translational Medicine and Innovative Drug Development, Jiangsu Simcere Diagnostics Co., Ltd., No. 699-18 Xuanwu Avenue, Xuanwu District, Nanjing 210042, China.
Email: chao.song@simceredx.com

Jianmin Li, Department of Orthopedics, Qilu Hospital of Shandong University, No.107 Wenhua West Road, Jinan 250012, China.
Email: gkljm@163.com

Xing Zhang, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Department of Medical Melanoma and Sarcoma, Sun Yat-sen University Cancer Center, No. 651 Dongfeng East Road, Guangzhou 510060, China.
Email: zhangxing@sysucc.org.cn

Funding information
None

Abstract

Gene fusions are one of the most common genomic alterations in soft tissue sarcomas (STS), which contain more than 70 subtypes. In this study, a custom-designed RNA sequencing panel including 67 genes was developed and validated to identify gene fusions in STS. In total, 92 STS samples were analyzed using the RNA panel and 95.7% (88/92) successfully passed all the quality control parameters. Fusion transcripts were detected in 60.2% (53/88) of samples, including three novel fusions (*MEG3–PLAG1*, *SH3BP1–NTRK1*, and *RPSAP52–HMGA2*). The panel demonstrated excellent analytic accuracy, with 93.9% sensitivity and 100% specificity. The intra-assay, inter-assay, and personnel consistencies were all 100.0% in four samples and three replicates. In addition, different variants of *ESWR1–FLI*, *COL1A1–PDGFB*, *NAB2–STAT6*, and *SS18–SSX* were also identified in the corresponding subtypes of STS. In combination with histological and molecular diagnosis, 14.8% (13/88) patients finally changed preliminary histology-based classification. Collectively, this RNA panel developed in our study shows excellent performance on RNA from formalin-fixed, paraffin-embedded

----

samples and can complement DNA-based assay, thereby facilitating precise diagnosis and novel fusion detection.

KEYWORDS
fluorescence *in situ* hybridization, gene fusion, immunohistochemistry, next-generation sequencing, soft tissue sarcoma

1 | INTRODUCTION

Soft tissue sarcoma comprising more than 70 subtypes is a group of highly heterogeneous tumors characterized by local invasion, invasive or destructive growth, high recurrence, and distant metastasis.[1-3] Some subtypes have high pathological similarity and are difficult to identify. Here, ~30% STS have specific gene fusion variants, and NCCN guidelines recommend 44 specific fusion genes as diagnostic markers.[4]

Gene fusions are derived from the breakage and reconnections between chromosomes, or from intra-chromosomal rearrangements with deletions, insertions, inversions, duplications, or altered transcriptions.[5] They represent an important class of somatic alterations and often act as drivers for tumorigenesis and progression.[6,7] Detection of defined gene fusions in STS can guide their classification and targeted therapies.[8] In addition, identification of novel gene fusions can provide insights into the mechanism of tumorigenesis, allow for subclassification of histologically similar STS and may serve as useful biomarkers for disease progression and treatment.[9]

Traditional methods to detect gene fusions are immunohistochemistry (IHC) and FISH. However, these technologies have a common limitation in identifying multiple fused genes simultaneously.[10-12] The emergence of NGS technology and modern computational tools allows the identification of multiple fused genes in parallel.[13] Many large DNA-based NGS (DNA-NGS) panels have been developed to detect SNVs, insertions/deletions, and fused genes.[14] However, breakpoints are unpredictable, and any fusion detection requires a great many probes to cover a large range of genomes. Moreover, the rare fusion of *ALK* detected using DNA-NGS may be unable to be transcribed and translated into functional fusion proteins frequently sensitive to targeted therapies.[15] By contrast, RNA sequencing has been widely used to detect gene fusions without prior knowledge of the partner sequence or specific breakpoints in cancer cell lines, fresh frozen tissues, and FFPE samples.[16-19]

In this study, we presented the development and clinical validation of a custom-designed fusion panel for sarcoma diagnosis using RNA-based NGS (RNA-NGS). Detailed sensitivity and specificity, reproducibility, and tumor content LOD of the RNA panel to identify gene fusions in STS FFPE samples were also discussed.

2 | MATERIALS AND METHODS

2.1 | STS samples

In total, 99 independent samples were used for validating the RNA panel, including two known fusion-positive samples (one cell line and one standard FFPE sample from our laboratory), five negative controls, and 92 clinical samples. These FFPE samples from the archives of the Department of Pathology, Sun Yat-sen University Cancer Center (Guangzhou, China) were classified into adipocytic tumors, fibroplastic tumors, perivascular tumors, smooth muscle tumors, chondro-osseous tumors, peripheral nerve sheath tumors, and tumors of uncertain differentiation in accordance with sources. Paraffin-embedded tissue and anonymization of data were used in accordance with local ethical regulations. This study was approved by the Institutional Review Board of Sun Yat-sen University Cancer Center, with the informed consent of the patients. The validation was performed in a double-blinded manner to eliminate the effect from personnel factors.

2.2 | Panel design

In total, 67 genes were targeted from the presence of clinically relevant fusions or oncogenic isoforms present in STS. The genes targeted in this panel are listed in Table S1. Target-specific probes covering full exons were custom designed to identify known fusion transcripts and potential novel fusion transcripts associated with 26 cancer genes. To identify potential fusions involving 41 partner genes, multiple probes were designed to cover partial exon of partner genes, such as *ETV1*, *ETV4*, and *CIC*, enabling assessments of gene expression. Gene-specific primers were designed proximally to the exon–exon junctions involved in the fusions. These probes from Roche Company were typically 120 bp in length.

2.3 | RNA extraction and quality control

All included samples were pathologically assessed before RNA extraction, and the proportion of tumor cells was ≥20% except one case (Table S2). A minimum of 10 unstained slides from FFPE tissue were obtained for each sample and reviewed by a pathologist, who had

given a preliminary diagnosis combined with IHC and FISH results. Total RNA was isolated using RNeasy FFPE kit and RNeasy Mini Kit (Qiagen). The quality of each RNA sample was tested using Qubit 4.0 and Agilent 4200 TapeStation system prior to library preparation and sequencing. To improve the performance of our assay, some QC parameters were established, such as $OD_{260}/_{280}$, $OD_{260}/_{230}$, and value of DV_{200}. The values of $OD_{260}/_{280}$ in all clinical samples were >1.8, while 16 samples produced low-quality RNA ($DV_{200} < 40\%$). Next, no less than 100 ng of unsheared RNA ($DV_{200} \geq 40$) or 200 ng of RNA ($DV_{200} < 40$) was used for the assay whenever available, but testing was also attempted for all samples with at least 50 ng of input RNA.

2.4 | Library preparation and sequencing

rRNA depletion, cDNA synthesis, and library preparation were performed using the KAPA RNA HyperPrep Kit (Kapa Biosystems, KK8540) with 100 ng of RNA ($DV_{200} \geq 40$) or 200 ng of RNA ($DV_{200} < 40$). The total volume of the final library was at least 40 ng. No obvious joint contamination was detected in the final library using the Agilent 4200 TapeStation system, and the main peak was between 300 and 500 bp. After quantification, NGS was performed on an Illumina Novaseq 6000 instrument (Illumina).

2.5 | Fusion detection

A custom pipeline was developed to perform reads alignment, fusion detection, and QC on RNA sequencing data. The software fastp (v.2.20.0) was used for adapter trimming. The software STAR (v2.7.6a) was used to align reads to the reference genome (UCSC's hg19 GRCh37). Star-Fusion (v1.9.1) was applied to identify the primary fused genes. To avoid false-positive fusion results, we use FusionInspector in 'inspect' mode to re-score and filter the predicted fusions with the following parameters: min_junction_reads 1, min_novel_junction_support 3, min_spanning_frags_only 5, max_promiscuity 10, only_fusion_reads, fusion_contigs_only. Then we used a tier-based filter strategy: if fusion pairs were annotated to COSMIC or ChimerKB database, the cutoff was fusion fragment per million (FFPM) > 0.02 and support reads > 2, or else the cutoff was FFPM > 0.07 and support reads > 8. Fusion expression was calculated based on FFPM using the raw data from the RNA panel and the average number of reads in this cohort was 12,353,247. The "copies/ng" of the fusion transcript was calculated using the droplet digital polymerase chain reaction (ddPCR).

2.6 | Fusion confirmation by DNA-NGS or real-time quantitative polymerase chain reaction

DNA-NGS was applied to validate gene fusions in four clinical samples by the panel containing 539 genes. qRT-PCR was used to confirm the novel fusions in three samples. RNA was transcribed into cDNA and then performed PCR reaction using the HiScript II One Step qRT-PCR Probe Kit (Vazyme, Q222-01) under the following parameters: 50°C for 15 min (reverse transcription); 95°C for 30 s, 45 cycles at 95°C for 5 s, 60°C for 20 s; 4°C for hold. The primers used for RT-PCR are listed in Table S3. Fusion product sequences using DNA-NGS or qRT-PCR were then assessed to ensure that they aligned with the sequence predicted by the RNA panel.

3 | RESULTS

3.1 | Detection limit

A complete process was illustrated from the total RNA isolation to the report of data analysis focusing on gene fusion detection (Figure 1A). The RNA panel was designed to detect gene-specific fusions in STS.

To determine how tumor contents affected fusion detection, serial dilution experiments were performed. A standard RNA sample with an *ETV6–NTRK3* fusion was serially diluted with RNA from a normal control to generate various levels, from 50 copies/100 ng to 500 copies/100 ng. The *ETV6–NTRK3* fusion was identified in all replicates across all dilutions (Figure 1B). Then, standard FFPE samples (50 copies) and cell line samples (50 copies) were detected stably in five replicates. Notably, the FFPM to support *ETV6–NTRK3* and *EWSR1–FLI1* differed greatly (Figure 1C). In addition, interfering substances ethanol (1% V/V) and protease K (0.08 mg/ml) showed no effects on the experimental flow (Table S4). Finally, the samples were subjected to database construction, computer sequencing, and bioinformatics analysis.

3.2 | Validation based on clinical samples

To assess the clinical utility of the RNA panel, five fusion-positive, 28 FISH-positive, and 59 FISH-negative or no FISH clinical samples were selected. In our cohort, Ewing sarcoma, malignant small round cell tumors, and dermatofibrosarcoma protuberans (DFSP) were the three most common tumor types, occupying 10%, 10%, and 5% respectively (Table S2). The proportion of other subtypes ranged from 1% of lipofibromatosis-like neural tumor to 4% of inflammatory myofibroblastic tumor. In accordance with histological diagnosis, 55 cases were most likely to have translocation-related sarcomas, including 14 sarcoma subtypes, namely alveolar soft part sarcoma, clear cell sarcoma, congenital/infantile fibrosarcoma, DFSP, Ewing sarcoma, extraskeletal myxoid chondrosarcoma, inflammatory myofibroblastic tumor, low-grade fibromyxoid sarcoma, malignant peripheral nerve sheath tumor, malignant small round cell tumors, mesenchymal chondrosarcoma, pseudomyogenic/epithelioid hemangioendothelioma, solitary fibrous sarcoma, and synovial sarcoma. Most of the remaining

cases may be attributed to complex-genomics sarcomas. Four samples failed in the QC process and were excluded from further analysis. Therefore, the final clinical samples included 16 ($DV_{200} < 40\%$) and 72 ($DV_{200} \geq 40\%$) FFPE samples (Figure 2A).

The fusion transcripts were detected in 60.2% of cases (53/88) using the RNA fusion panel. The most common fused gene was *EWSR1* with different partners including *FLI1*, *WT1*, *ETV1*, *ERG*, *DDIT3*, and *CREB3L1* genes. Other commonly identified gene fusions included *ASPSCR1–TFE3* in alveolar soft part sarcoma, *COL1A1–PDGFB* in DFSP, *ETV6–NTRK3* in congenital/infantile fibrosarcoma,

FUS–CREB3L2 in low-grade fibromyxoid sarcoma, and so forth (Figure 2B). Notably, novel gene fusions were identified, including *MEG3–PLAG1*, *SH3BP1–NTRK1*, and *RPSAP52–HMGA2*, which were confirmed by qRT-PCR (Figure S1).

3.3 | Analytic sensitivity and specificity

Of the five fusion-positive (DNA-NGS), 28 FISH-positive samples and five negative controls, no fusions were found in the five negative
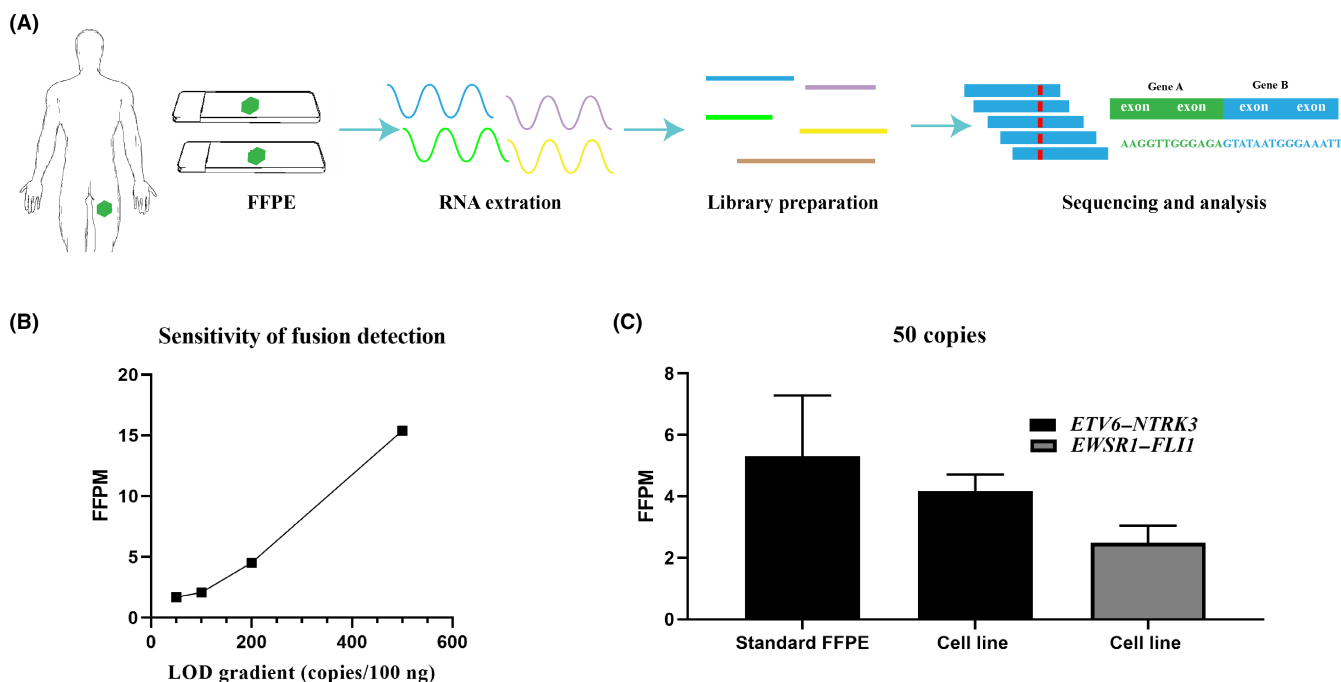

FIGURE 1 Performance of the RNA panel. (A) Workflow of the RNA fusion assay. (B) Detection limit of the RNA panel using a serial dilution assay. (C) FFPM of *ETV6–NTRK3* and *EWSR1–FLI1* fusions identified in standard FFPE and cell line samples in five replicates
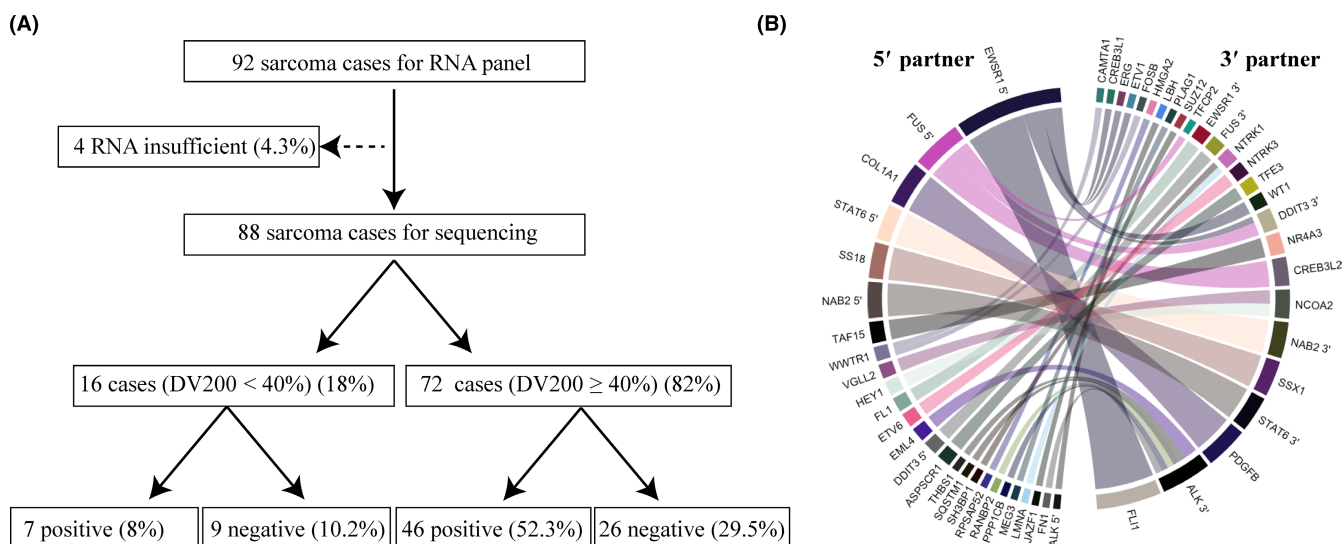

FIGURE 2 Clinical validation of the RNA panel. (A) Quality control summary of all the STS samples submitted for the RNA fusion assay. (B) Circos plot of the 73 gene fusions in all the STS identified by the RNA fusion assay 5′ partner

controls, demonstrating 100% specificity. As shown in Table 1, the sensitivity of fusions detected by the RNA fusion panel was 93.9% (31/33) compared with FISH and DNA-NGS analysis.

DNA-NGS was used to validate the inconsistency between FISH assay and RNA fusion panel. In the two samples, the fusions were not detected using the RNA fusion panel, whereas an *EWSR1* (intron)–*ZNF444* fusion was identified in one sample by the DNA-NGS. Additionally, in another sample, DNA-NGS assay identified an *EWSR1–DCTIN2* (intron) fusion, but not *EWSR1–DDIT3* fusion detected using the RNA fusion panel. The rare fusion detected using DNA sequencing assays may not be transcribed and translated into the functional fusion proteins of matched targeted therapies.[15] Interestingly, we identified an *FUS–TFCP2* fusion in one sample, which was initially diagnosed as inflammatory myofibroblastic tumor due to *ALK* disruption using FISH. However, we did not find any *ALK* fusion in the sample by DNA-NGS.

3.4 | High precision in detecting fusions by the RNA panel

To evaluate the precision of the RNA panel, two standard FFPE samples (100 copies and 200 copies) with *ETV6–NTRK3*, one clinical FFPE sample with *SS18–SSX1* and one clinical FFPE sample with negative fusion were assessed (Table 2). Three fusion-positive samples had the expected fusions detected in all three replicates with one sequencing run by examiner A. Furthermore, the same three samples were used across two sequencing runs (A and B) with complete concordance of gene fusion results for all samples. In addition to verifying the repeatability and reproducibility of the assay for the presence of gene fusions, we also observed the FFPM among replicates of the same sample in the same run and across different runs. The CV of FFPM was the largest among personnel consistencies, followed by that within inter-assay and the smallest within intra-assay.

TABLE 1 Specificity and sensitivity of the RNA panel in fusion detection

	FISH or DNA-NGS		
RNA-NGS	Positive	Negative	Performance
Positive	31	0	100.0% (PPV)
Negative	2	5	71.4% (NPV)
Performance	93.9 (sensitivity)	100.0% (specificity)	94.7% (accuracy)

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

TABLE 2 Intra-assay, inter-assay, and personnel-assay reproducibility

				CV (%)		
Sample	Fusion detected	Run and replicate	FFPM	Intra-assay	Inter-assay	Personnel assay
Standard FFPE 1 (100 copies)	ETV6–NTRK3	A-1-1	14.705			
		A-1-2	13.638	24.67		
		A-1-3	8.941			
		A-2	19.531		31.43	
		B	22.091			39.59
Standard FFPE 2 (200 copies)	ETV6–NTRK3	A-1-1	16.836			
		A-1-2	24.453	18.62		
		A-1-3	20.212			
		A-2	29.189		24.73	
		B	13.859			27.34
815400-1	SS18–SSX1	A-1-1	295.210			
		A-1-2	423.924	27.07		
		A-1-3	517.457			
		A-2	322.674		17.23	
		B	246.488			35.58
Negative control	NA	A-1-1	NA			
		A-1-2	NA	NA		
		A-1-3	NA			
		A-2	NA		NA	
		B	NA			NA

Note: NA, not applicable. "A" and "B" represented two examiners. Examiner A carried out two sequencing runs (A-1 and A-2), in which A-1 contained three replicates (A-1-1, A-1-2, and A-1-3). Examiner B carried out only one sequencing run.

These findings indicated that FFPE samples could be used for RNA-NGS to detect fusion genes and our RNA panel was very stable to identify gene fusions.

3.5 | Variants of the fusion transcript

EML4–ALK fusion has at least 15 different variants and the expression of particular variants directly impact the response of patients to ALK inhibitors.[20,21] Therefore, we investigated the breakpoint of the same fusion gene in patients of our cohort.

In our cohort, we identified 13 *ESWR1* fusions in nine patients harboring Ewing sarcoma. *EWSR1–FLI1* was found in seven patients, while *ESWR1–ERG* and *ESWR1–ETV1* only in one patient, respectively. Coexistence of *EWSR1–FLI1* and *FLI1–EWSR1* was present in two patients. These results were consistent with previous reports that ~90% of Ewing sarcoma harbored *EWSR1–FLI1* fusions, and other partners were from the ETS family, including *ERG* and *ETV1*.[22] *EWSR1* rearrangement most frequently occurred in *EWSR1* exon 8, while less frequently in other exons, including exons 7, 9, and 11. *EWSR1* most frequently fused to exon 7 of *FLI1*, exon 9 of *ERG*, and exon 10 of *ETV1* (Figure 3A).

COL1A1–PDGFB fusion was detected in six patients. The breakpoint in the *PDGFB* gene was exon 2 in 5 patients, while that in the *COL1A1* gene was exon 46 in two patients, and exons 17, 24, 29, 43 in one patient (Figure 3B). Interestingly, we found a novel *COL1A1–PDGFB* fusion with the breakpoint in exon 3 region of *PDGFB* in one patient.

We detected one *NAB2* exon 4–*STAT6* exon 2, one *NAB2* exon 4–*STAT6* exon 15, one *NAB2* exon 6–*STAT6* exon 16, and two *NAB2* exon 6–*STAT6* exon 17 fusion transcripts in five cases, respectively. Interestingly, several variants of *STAT6–NAB2* fusions coexisted in

three patients harboring *NAB2–STAT6* fusion, but their clinical significance was still unclear. For *STAT6–NAB2* fusions, *STAT6* exon 15 was the predominant breakpoint location and most frequently fused to exon 7 of *NAB2* (Figure 3C).

The t(X;18)(p11.2;q11.2) is the genetically hallmark of SS and produces SS18–SSX fusion protein.[23,24] Here, we also identified four *SS18–SSX1* and one *SS18–SSX2* fusions in four patients. The breakpoint in the *SS18* gene was exon 10 in four patients, while that in the *SSX1* and *SSX2* genes was exon 3 in three patients and exon 6 in one patient, respectively (Figure 3D). Interestingly, *SS18* exon 5–*SSX1* exon 9 fusions coexisted in one patient harboring *SS18* exon 10–*SSX1* exon 7 fusion.

3.6 | Sarcoma harboring novel fusions

The advantage of the RNA-NGS over the trapping based on DNA sequencing assay is that the RNA-NGS can recognize all fusion transcripts without the need to design probes to cover the fracture areas. In our study, three cases were found to have novel gene fusions. The novel gene of two fusions was identified as the partner gene fused to a gene known to be recurrently involved in the specific tumor type. However, another novel fusion contained two rare genes. These three cases were summarized as follows:

3.6.1 | Case one

A 2-year-old boy presented a local mucinous lesion in his shoulder zone. The tumor was composed of sparsely spindle or stellate fibroblastoid cells and many disordered collagen fibers and
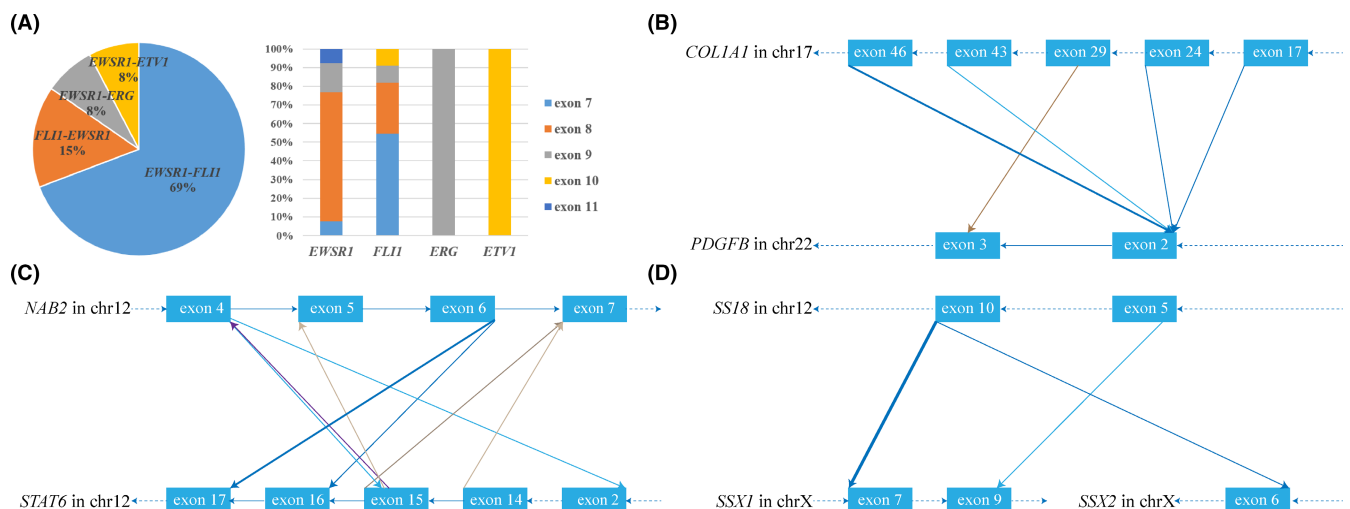

FIGURE 3 Breakpoint distribution in STS subtypes and corresponding fusion partners. (A) Frequency of *ESWR1* fusion variants and distribution of fusion breakpoint positions in Ewing sarcoma. (B) Distribution of *COLIA1–PDGFB* fusion breakpoints. The bold line (blue) represents two *COLIA1* exon 46–*PDGFB* exon 2 fusions. (C) Distribution of *NAB2–STAT6* fusion breakpoints. The bold line (blue) represents two *NAB21* exon 6–*STAT6* exon 17 fusions. (D) Distribution in *SS18* fusion and the fusion breakpoints. The bold line (blue) represents three *SS18* exon 10–*SSX16* exon 7 fusions
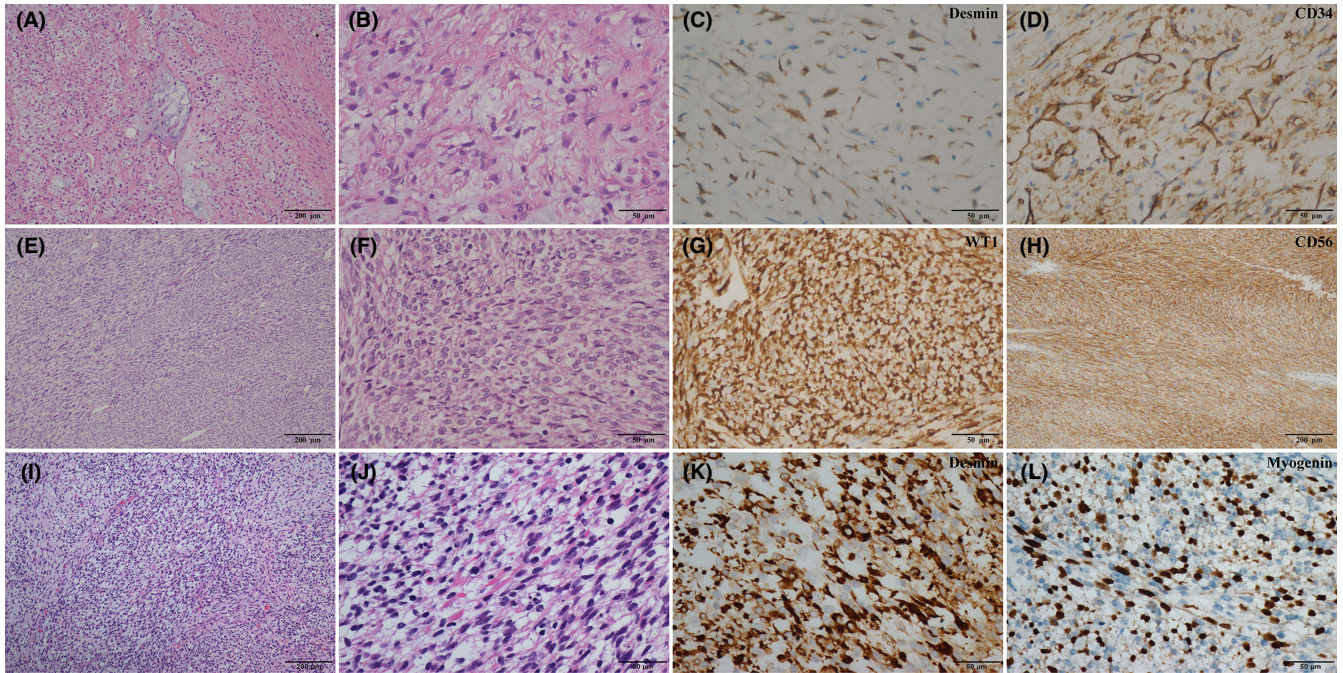
FIGURE 4 Representative HE and IHC of the STS harboring novel fusions. (A–D) Pediatric fibromyxoid soft tissue tumor (*MEG3–PLAG1*) displays relatively sparse spindle cells in the disorganized collagen fibers. The tumor is positive for Desmin and CD34 by IHC. (A: ×100, B–D: ×400). (E–H) *NTRK* fusion-positive sarcoma (*SH3BP1–NTRK1*) shows simple cell morphology and poor differentiation. The tumor is diffusely positive for WT1 and CD56 by IHC. (E, H: ×100; G, F: ×400). (I–L) Embryonal rhabdomyosarcoma (*RPSAP52–HMGA2*) composed of spindle cells with moderate amounts of eosinophilic or clear cytoplasm. IHC staining shows that the tumor is strongly and diffusely positive for Desmin and Myogenin. (I: ×100; J–L: ×400)

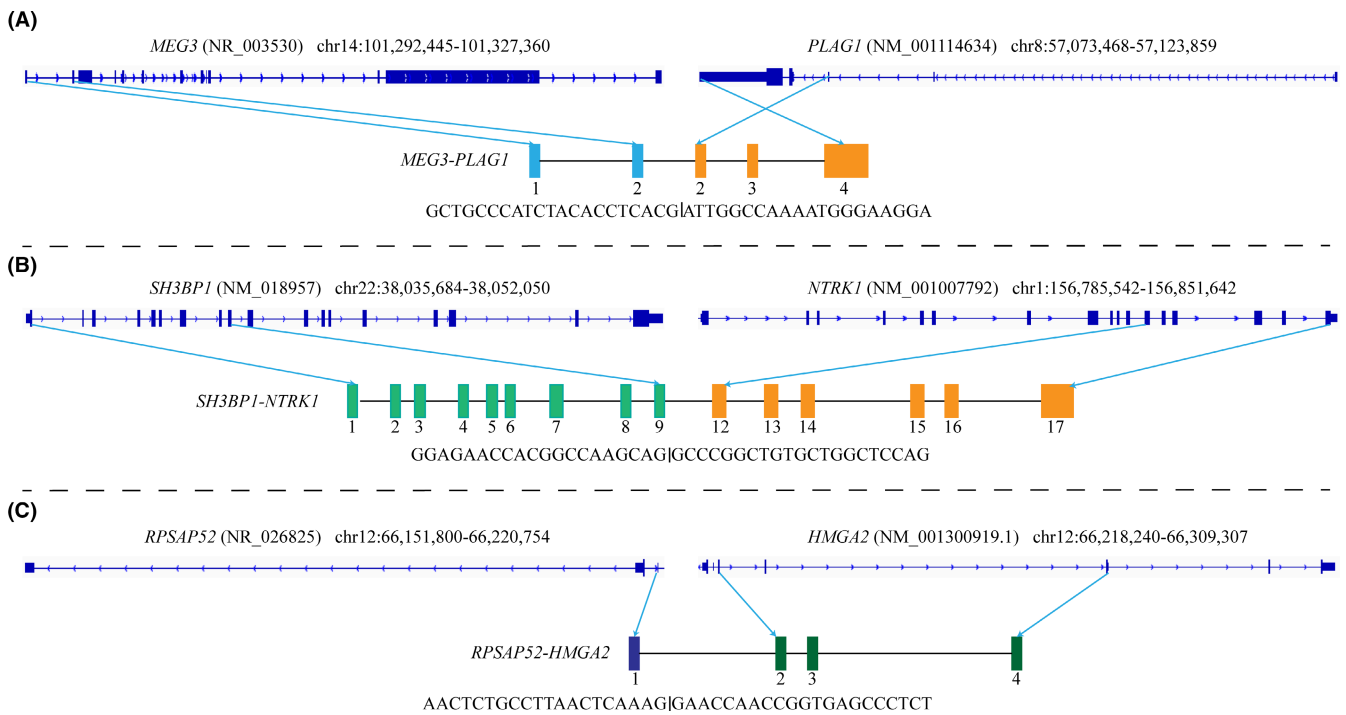**FIGURE 5** Schematic diagrams of the novel fusions identified by the RNA fusion assay. (A) The *MEG3–PLAG1* fusion between *MEG3* exon 2 and *PLAG1* exon 2 in a pediatric fibromyxoid soft tissue tumor. (B) A *SH3BP1–NTRK1* fusion between *SH3BP1* exon 9 and *NTRK1* exon 12 in a *NTRK* fusion-positive sarcoma. (C) A *RPSAP52–HMGA2* fusion between *RPSAP52* exon 1 and *HMGA2* exon 2 in an embryonal rhabdomyosarcoma

myxoid change (Figure 4A). The nucleus of stellate fibroblastoid cells was small and visible, while uniform chromatin and mitotic images were noted (Figure 4B). Tumor cells were double positive for Desmin and CD34 in IHC (Figure 4C,D). Ki-67 stain for tumor cells showed a relatively low proliferation activity of 10%. A novel *MEG3–PLAG1* fusion was detected between *MEG3* exon 2 and *PLAG1* exon 2 in the tumor, which was initially diagnosed as fibroblast-derived tumor (Figure 5A). In combination with histological patterns, IHC results and molecular analysis, we reclassified this lesion as pediatric fibromyxoid soft tissue tumor with *PLAG1* fusion now.[25]

3.6.2 | Case two

A novel *SH3BP1–NTRK1* fusion was found between *SH3BP1* exon 9 and *NTRK1* exon 12 in the renal mass of a 7-month-old infant (Figure 5B). The tumor cell showed simple spindle cell morphology with poorly differentiated and no three-phase differentiation structure of typical nephroblastoma was observed (Figure 4E,F). Tumor cells were positive for WT1 and CD56, and showed strong proliferation activity of 80% for Ki-67 stain (Figure 4G,H). FISH identified no *ETV6* gene rearrangement (not shown). Based on a novel *SH3BP1–NTRK1* fusion, the final diagnosis of this tumor was a *NTRK*-rearranged spindle cell neoplasm.

3.6.3 | Case three

A novel *RPSAP52–HMGA2* fusion was identified in a rhabdomyosarcoma with breakpoint in *RPSAP52* exon 1 and *HMGA2* exon 2 (Figure 5C). Microscopically, the tumor showed that spindle cells were arranged in bundles or braids with eosinophilic to clear cytoplasm and an active mitotic activity of 10/10 HPFs (Figure 4I,J). Regional interstitial mucinous lesion and small focal necrosis were also observed. Tumor cells were positive for rhabdomyosarcoma markers (Desmin, Myogenin and MyoD1), and H3K27me3 protein was retained. Ki-67 index showed high proliferation activity of 60% (Figure 4K,L). Combined with the current results of molecular analysis and IHC staining, the lesion was diagnosed as embryonal rhabdomyosarcoma, with a novel *RPSAP52–HMGA2* fusion.

3.7 | Application of RNA-NGS data for the integrated histological and molecular classification

In combination with histological and molecular diagnosis, 14.8% (13/88) patients changed preliminary histology-based classification. 53 cases showed fusion transcripts, and the integrated histology-based diagnosis was completely consistent in 75% (40/53) patients (Figure 6). The most pronounced changes occurred in the three unclassified sarcoma samples that split up into

Ewing sarcoma (*ESWR1–ETV1*), low-grade fibromyxoid sarcoma (*EWSR1–CREB3L1*), and epithelioid cell histiocytoma (*SQSTM1–ALK*). While all three liposarcoma of our cohort corresponded to myxoid/round cell liposarcoma with *DDIT3* fusion, the two malignant small round cell tumors were refined into desmoplastic small round cell tumors due to *EWSR1–WT1* fusion. In addition, five different subtypes of sarcomas were completely reclassified into pediatric fibromyxoid soft tissue tumor with *PLAG1* fusion (*MEG3–PLAG1*), lipofibromatosis-like neural tumor (*LMNA–NTRK1*), *NTRK* fusion-positive sarcoma (*SH3BP–NTRK1*), spindle cell rhabdomyosarcoma (*FUS–TFCP2*), and inflammatory myofibroblastic tumors (*THBS1–ALK*).

4 | DISCUSSION

As an increasing number of clinically significant fusions are being identified, an efficient method allowing the detection of many recurrent novel fusions in different tumors simultaneously is critically important. We developed and clinically validated a targeted RNA sequencing assay including most fusion genes in the central nervous system sarcomas of the new version, based on probe hybridization capture sequencing, to detect multiple gene fusions in our study. The rapid and easy automatable protocol enabled stranded RNA library construction in ~4 h, and was suitable for the library construction of purified total RNA and low-quality RNA samples, including FFPE. This custom-designed gene panel allowed high-throughput sequencing to identify rapidly both recurrent and novel gene fusions without the need for primers specific to both fusion partners of a given gene.

Our assay can successfully detect fusions even when the RNA level of tumors is as low as 50 copies. The LOD of the assay functions on FFPE tissue and requires as little as 100 ng of RNA input. Additionally, the assay can detect oncogenic fusion transcripts, which is not affected by interfering substances, including ethanol and proteinase K. In clinical validation, the assay could detect the fusions in FFPE samples among sarcomas using RNA-NGS, with a sensitivity of 93.9% and a specificity of 100%. Currently, the discordance of fusion detection between RNA-NGS and DNA-based assay may be explained by the following several reasons. First, fusion detection by the FISH or DNA panel assay is limited to the genes covered by probes. By contrast, RNA-NGS can detect novel fusions that cannot be detected using the DNA-based assay. During the assay development, we identified new fusion transcripts in 22 patients of 59 FISH-negative or not done patients, which indicated a great advantage in detecting unknown fusions for this assay. Second, some uncommon fusions detected using DNA-based assay generated no aberrant transcripts or proteins, therefore failing in the response to targeted therapies.[15] In our cohort, the RNA fusion assay detected *ESWR1–DDIT3* instead of *EWSR1–DCTN2* using a DNA panel assay in one case. Therefore, validation assays may be necessary for the cases harboring rare fusions detected using DNA-based assay to ensure that the

FIGURE 6 Schematic illustration of the diagnostic changes in 53 STS samples originally classified by the WHO classification 2020 (left side) and then reclassified by integrating histology and gene fusion markers
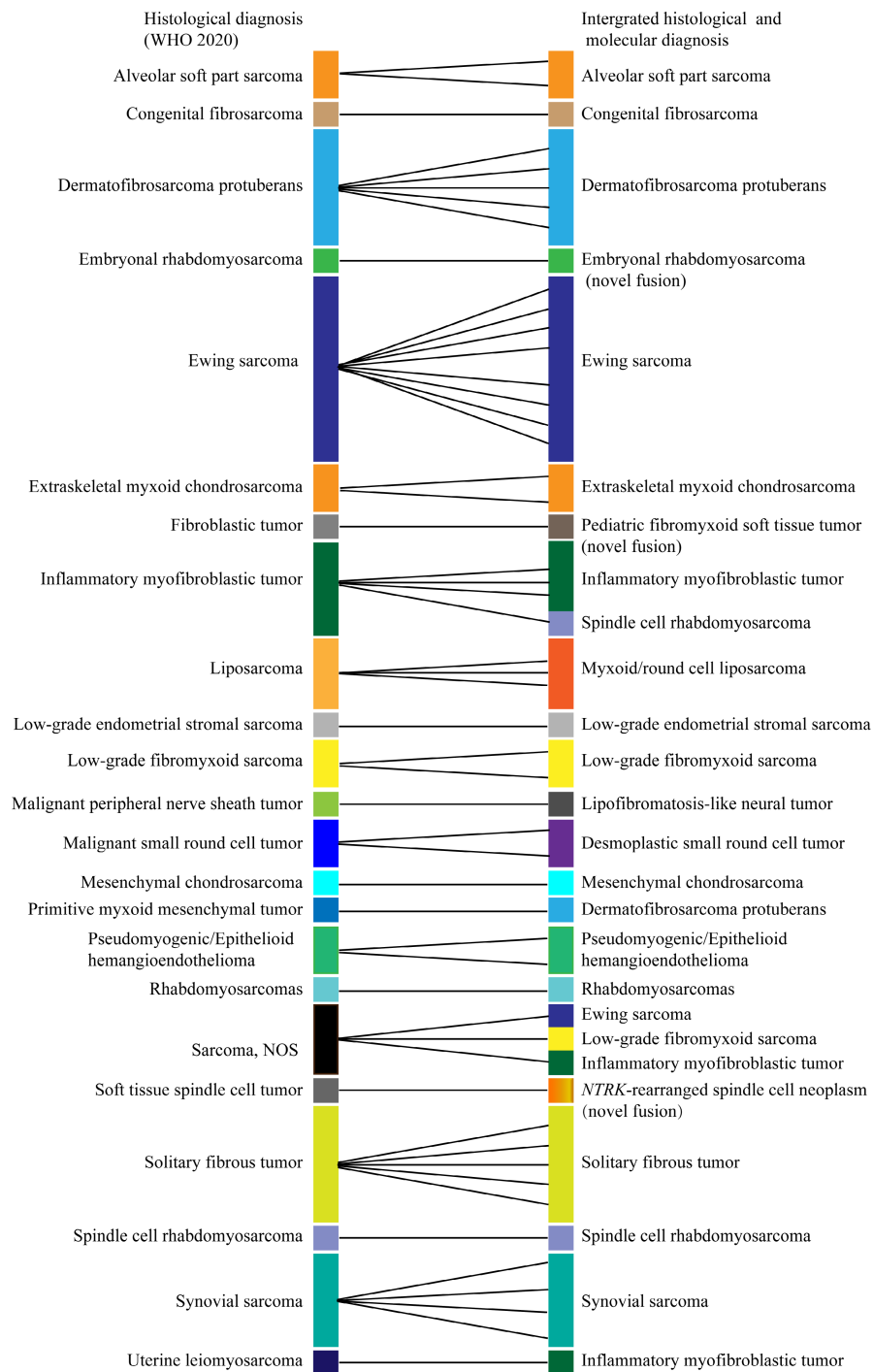
rearrangement was intact and could generate an active chimeric transcript. Third, the tissue heterogeneity and other factors may also contribute to the discordance.

In our study, different variants including *ESWR1–FLI* in Ewing sarcoma and *COL1A1–PDGFB* in DFSP were identified. Ewing sarcoma is the first group of STS to be demonstrated to have characteristic molecular genetic variation, such as t(11;22)(q24;q12) translocation.[26] The most breakpoints (80%) of the *EWSR1* gene are located at exon 7 or 8, causing the N-terminal of *EWSR1* to fuse with the heterologous DNA-binding domain of the partner gene.[27,28] The tumorigenic mechanism of DFSP is mainly defined

by the formation of a fusion gene between *COL1A1* and *PDGFB* genes.[29] Despite no correlation between *COL1A1* breakpoints and clinical findings,[30] the clinical significance of *PDGFB* breakpoints needs to be further investigated.

A solitary fibrous tumor (SFT) is a rare *NAB2–STAT6* fusion-associated neoplasm, which drives *STAT6* nuclear relocation.[31-33] The common fusion variants of *NAB2–STAT6* were the variant 1 (*NAB2* exon 4–*STAT6* exon 2/15) and variant 2 (*NAB2* exon 6–*STAT6* exon 16/17). Histologically, SFTs with the most common variant 1 are often located in the pleuropulmonary region with less cells and abundant collagen, while SFTs with variant 2 often occur in the deep

soft tissue of the retroperitoneum and intra-abdominal pelvic region or in the meninges with displaying a round or ovoid cell morphology. Furthermore, there were also significant differences in gene expression associated with anatomic localization and *NAB2–STAT6* gene fusion variants.[34] Therefore, identification of distinct *NAB2–STAT6* gene fusion variants may establish a potential molecular biologic basis for clinicopathologic differences in SFTs.

Interestingly, we also identified four intron–exon gene fusions in our cohort, including one *ETV6–NTRK3* in congenital fibrosarcoma, two *FUS–CREB3L2* in low-grade fibromyxoid sarcoma, and one *EML4–ALK* fusion in inflammatory myofibroblastic tumor. Coexistence of exon–exon gene fusions was also present in all the patients. The reason for intron–exon gene fusion was intron retention occurring near the breaking point and the breakpoint was very close to the corresponding exon.[35] To date, no summarized reports have described these breakpoints at the RNA level and its relationship with clinical findings, except for *EML4–ALK* fusion.

Novel fusions were identified in three cases, two long noncoding RNA (lncRNA) fusions and one *NTRK* fusion. The *MEG3* gene transcribes a 1.6 kb lncRNA that acts as an antitumor component in different cancer cells such as breast, liver, glioma, colorectal, cervical, gastric, lung, ovarian, and osteosarcoma.[36] Tumorigenesis in *PLAG1* fusion tumors often results from PLAG1 overexpression driven by constitutively expressed fusion partners.[37,38] The *PLAG1*-rearranged STS displays a purely fibromyxoid histology, with high expression of CD34 and Desmin.[25] In addition, *RPSAP52* is a pseudogene-transcribed RNA that runs antisense to the oncofetal gene *HMGA2* and positively regulates HMGA2 expression through the formation of an R loop structure.[39,40] These findings were consistent with the important role of *HMGA2* in the oncogenicity of fusion-negative rhabdomyosarcoma.[41] The three novel fusions all contained their functional domains. the *MEG3–PLAG1* fusion transcript contained a serine-rich transcriptional activation domain of the *PLAG1* gene, which encoded a developmentally regulated transcription factor with seven canonical C2H2 zinc finger domains in the N-terminus. The *SH3BP1–NTRK1* fusion transcript contained NTRK1 kinase domain, while *RPSAP52–HMGA2* fusion transcript included DNA-binding domains of the *HMGA2* gene. These fusions are likely to activate downstream signaling pathways that lead to tumorigenesis.

In other *NTRK* fusions, *SH3BP1* was identified as a novel partner gene fused to *NTRK1*. In addition to the fusion described above, eight sarcomas were identified to harbor gene fusions containing the kinase domain of a protein kinase. The common gene fusions *LMNA–NTRK1* and *ETV6–NTRK3* were detected in lipofibromatosis-like neural tumor and infantile fibrosarcoma, encoding TRK receptors A and C, respectively.[42,43] *EML4–ALK*, *FN1–ALK*, *RANBP2–ALK*, *THBS1–ALK*, and *SQSTM1–ALK* were found respectively in five inflammatory myofibroblastic tumors, while the most breakpoints of *ALK* were located at exon 20.[44] The remaining one case was *PPP1CB–ALK* in rhabdomyosarcoma. The detection of gene fusions involving kinase genes can identify a target for therapies using agents that are approved or available in the setting of clinical trials.

Based on the excellent performance of our RNA panel on routinely processed tissue samples, we used this method for molecular analysis of 53 STS samples with different WHO histological types (Table S5). Unsupervised analysis of the NGS data corresponded largely to the previously characterized biologically distinct STS subtypes such as Ewing sarcoma, DFSP, SFT, and SS. We also used our RNA panel for an integrated histological and molecular classification of the STS. The results exhibited that the clinical diagnosis of 13 patients changed and targeted therapies may be considered for the second-line treatment in nine patients harboring *ALK* or *NTRK* fusion. In the 2020 WHO Classification of STS, there have been the tumors named in combination with genetic variations such as *CIC*-rearranged sarcomas and sarcomas with *BCOR* genetic alterations. These findings illustrated the high impact of molecular markers on future sarcoma classification, and the RNA gene panel NGS may become an important tool for sarcoma diagnosis in the future through timely and reliable molecular profiling.

Several limitations of this assay should be of concern. First, degradation in FFPE samples led to low-quality RNA, which seriously affected the performance for fusion detection and increased the negative risk. Second, the RNA panel assay cannot detect fusions of the two genes excluded from our panel, which may miss some fusions that may play an important role in tumorigenesis. To better service clinical diagnosis, we will continuously upgrade our panel by incorporating newly discovered fusion genes in the future.

In summary, we established and validated a sarcoma-tailored 67-gene RNA panel for the improvement of STS molecular diagnosis. This panel can detect known or unknown gene fusions, thereby serving as a good supplement for DNA-based assays. Moreover, the method is applicable to routinely processed FFPE tissue samples and bears great potential for facilitating future integrated STS diagnostics as molecular classification.

DISCLOSURE
The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS
WMH, LY, and XKZ were responsible for study conception and original draft preparation. YN, DCH, and ZCW participated in sample collection and investigation. XML, YL, CZ, WLD, MQT, and RD analyzed and interpreted the data. CS, JML, and XZ reviewed and edited the manuscript. All the authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT
The data sets used and/or analyzed during the current study are available from the corresponding authors on reasonable request.

ORCID
Xinke Zhang https://orcid.org/0000-0002-3527-2140
Xing Zhang https://orcid.org/0000-0003-2665-6940

REFERENCES

1. Burningham Z, Hashibe M, Spector L, et al. The epidemiology of sarcoma. *Clin Sarcoma Res*. 2012;2:14.

2. Taylor BS, Barretina J, Maki RG, et al. Advances in sarcoma genomics and new therapeutic targets. *Nat Rev Cancer*. 2011;11:541-557.

3. Kallen ME, Hornick JL. The 2020 WHO classification: what's new in soft tissue tumor pathology? *Am J Surg Pathol*. 2021;45:e1-e23.

4. von Mehren M, Kane JM, Benjamin RS, et al. Soft tissue sarcoma, version 2.2020, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2020.

5. Mertens F, Johansson B, Fioretos T, et al. The emerging complexity of gene fusions in cancer. *Nat Rew Cancer*. 2015;15:371-381.

6. Chougule A, Taylor MS, Nardi V, et al. Spindle and round cell sarcoma with EWSR1-PATZ1 gene fusion: a sarcoma with polyphenotypic differentiation. *Am J Surg Pathol*. 2019;43:220-228.

7. Le Loarer F, Pissaloux D, Watson S, et al. Clinicopathologic features of CIC-NUTM1 sarcomas, a new molecular variant of the family of CIC-fused sarcomas. *Am J Surg Pathol*. 2019;43:268-276.

8. Mosse YP, Voss SD, Lim MS, et al. Targeting ALK with crizotinib in pediatric anaplastic large cell lymphoma and inflammatory myofibroblastic tumor: a children's oncology group study. *J Clin Oncol*. 2017;35:3215-3221.

9. Yamazaki F, Nakatani F, Asano N, et al. Novel NTRK3 fusions in fibrosarcomas of adults. *Am J Surg Pathol*. 2019;43:523-530.

10. Hamard C, Mignard X, Pecuchet N, et al. IHC, FISH, CISH, NGS in non-small cell lung cancer: what changes in the biomarker era? *Rev Pneumol Clin*. 2018;74:327-338.

11. Aydin HA, Pestereli E, Ozcan M, et al. A study detection of the ROS1 gene fusion by FISH and ROS1 protein expression by IHC methods in patients with ovarian malignant or borderline serous tumors. *Pathol Res Pract*. 2018;214:1868-1872.

12. Niu X, Chuang JC, Berry GJ, et al. Anaplastic lymphoma kinase testing: IHC vs. FISH vs. NGS. *Curr Treat Options Oncol*. 2017;18:71.

13. Zhao G, Chen L, Xiao M, et al. Rare coexistence of three novel CDCA7-ALK, FSIP2-ALK, ALK-ERLEC1 fusions in a lung adenocarcinoma patient who responded to Crizotinib. *Lung Cancer*. 2021;152:189-192.

14. Izevbaye I, Liang LY, Mather C, et al. Clinical validation of a myeloid next-generation sequencing panel for single-nucleotide variants, insertions/deletions, and fusion genes. *J Mol Diagn*. 2020;22:208-219.

15. Li W, Guo L, Liu Y, et al. Potential unreliability of uncommon ALK, ROS1, and RET genomic breakpoints in predicting the efficacy of targeted therapy in NSCLC. *J Thorac Oncol*. 2021;16:404-418.

16. Winters JL, Davila JI, McDonald AM, et al. Development and verification of an RNA sequencing (RNA-Seq) assay for the detection of gene fusions in tumors. *J Mol Diagn*. 2018;20:495-511.

17. Chang F, Lin F, Cao K, et al. Development and clinical validation of a large fusion gene panel for pediatric cancers. *J Mol Diagn*. 2019;21:873-883.

18. Zhu G, Benayed R, Ho C, et al. Diagnosis of known sarcoma fusions and novel fusion partners by targeted RNA sequencing with identification of a recurrent ACTB-FOSB fusion in pseudomyogenic hemangioendothelioma. *Mod Pathol*. 2019;32:609-620.

19. Peng H, Huang R, Wang K, et al. Development and validation of an RNA sequencing assay for gene fusion detection in formalin-fixed, paraffin-embedded tumors. *J Mol Diagn*. 2020;23:223-233.

20. Sabir SR, Yeoh S, Jackson G, et al. EML4-ALK Variants: biological and molecular properties, and the implications for patients. *Cancers (Basel)*. 2017;9:118.

21. Lin JJ, Zhu VW, Yoda S, et al. Impact of EML4-ALK variant on resistance mechanisms and clinical outcomes in ALK-positive lung cancer. *J Clin Oncol*. 2018;36:1199-1206.

22. Antonescu CR, Dal Cin P. Promiscuous genes involved in recurrent chromosomal translocations in soft tissue tumours. *Pathology*. 2014;46:105-112.

23. Clark J, Rocques PJ, Crew AJ, et al. Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. *Nat Genet*. 1994;7:502-508.

24. Crew AJ, Clark J, Fisher C, et al. Fusion of SYT to two genes, SSX1 and SSX2, encoding proteins with homology to the Kruppel-associated box in human synovial sarcoma. *EMBO J*. 1995;14:2333-2340.

25. Chung CT, Antonescu CR, Dickson BC, et al. Pediatric fibromyxoid soft tissue tumor with PLAG1 fusion: a novel entity? *Genes Chromosomes Cancer*. 2021;60:263-271.

26. Aurias A, Rimbaut C, Buffe D, et al. Translocation involving chromosome 22 in Ewing's sarcoma. A cytogenetic study of four fresh tumors. *Cancer Genet Cytogenet*. 1984;12:21-25.

27. Riggi N, Knoechel B, Gillespie SM, et al. EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer Cell*. 2014;26:668-681.

28. Arvand A, Denny CT. Biology of EWS/ETS fusions in Ewing's family tumors. *Oncogene*. 2001;20:5747-5754.

29. Llombart B, Serra-Guillen C, Monteagudo C, et al. Dermatofibrosarcoma protuberans: a comprehensive review and update on diagnosis and management. *Semin Diagn Pathol*. 2013;30:13-28.

30. Otsuka-Maeda S, Kajihara I, Kanemaru H, et al. Retrospective study of COL1A1-PDGFB fusion gene-positive dermatofibrosarcoma protuberans in Kumamoto university. *Clin Exp Dermatol*. 2020;45:1067-1068.

31. Chmielecki J, Crago AM, Rosenberg M, et al. Whole-exome sequencing identifies a recurrent NAB2-STAT6 fusion in solitary fibrous tumors. *Nat Genet*. 2013;45:131-132.

32. Robinson DR, Wu YM, Kalyana-Sundaram S, et al. Identification of recurrent NAB2-STAT6 gene fusions in solitary fibrous tumor by integrative sequencing. *Nat Genet*. 2013;45:180-185.

33. Doyle LA, Vivero M, Fletcher CD, et al. Nuclear expression of STAT6 distinguishes solitary fibrous tumor from histologic mimics. *Mod Pathol*. 2014;27:390-395.

34. Bieg M, Moskalev EA, Will R, et al. Gene expression in solitary fibrous tumors (SFTs) correlates with anatomic localization and NAB2-STAT6 gene fusion variants. *Am J Pathol*. 2021;191:602-617.

35. Choi YL, Takeuchi K, Soda M, et al. Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. *Cancer Res*. 2008;68:4971-4976.

36. Al-Rugeebah A, Alanazi M, Parine NR. MEG3: an oncogenic long non-coding RNA in different cancers. *Pathol Oncol Res*. 2019;25:859-874.

37. Hibbard MK, Kozakewich HP, Dal Cin P, et al. PLAG1 fusion oncogenes in lipoblastoma. *Cancer Res*. 2000;60:4869-4872.

38. Logan SJ, Schieffer KM, Conces MR, et al. Novel morphologic findings in PLAG1-rearranged soft tissue tumors. *Genes Chromosomes Cancer*. 2021;60:577-585.

39. Boque-Sastre R, Soler M, Oliveira-Mateos C, et al. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc Natl Acad Sci USA*. 2015;112:5785-5790.

40. Oliveira-Mateos C, Sanchez-Castillo A, Soler M, et al. The transcribed pseudogene RPSAP52 enhances the oncofetal HMGA2-IGF2BP2-RAS axis through LIN28B-dependent and independent let-7 inhibition. *Nat Commun*. 2019;10:3979.

41. Ouchi K, Miyachi M, Yagyu S, et al. Oncogenic role of HMGA2 in fusion-negative rhabdomyosarcoma cells. *Cancer Cell Int*. 2020;20:192.

42. Bartenstein DW, Coe TM, Gordon SC, et al. Lipofibromatosis-like neural tumor: case report of a unique infantile presentation. *JAAD Case Rep*. 2018;4:185-188.

43. Yamamoto H, Yoshida A, Taguchi K, et al. ALK, ROS1 and NTRK3 gene rearrangements in inflammatory myofibroblastic tumours. *Histopathology*. 2016;69:72-83.

44. Lovly CM, Gupta A, Lipson D, et al. Inflammatory myofibroblastic tumors harbor multiple potentially actionable kinase fusions. *Cancer Discov.* 2014;4:889-895.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.