Am J Cancer Res 2022;12(5):2376-2386
www.ajcr.us /ISSN:2156-6976/ajcr0134427

Original Article

The potential and limitation of targeted chromosomal breakpoint sequencing for the *ROS1* fusion gene identification in lung cancer

Ming-Szu Hung[1,2,3], Yu-Ching Lin[1,2,3], Fen-Fen Chen[4], Yuan-Yuan Jiang[1], Yu-Hung Fang[1], Ming-Shian Lu[5], Chin-Kuo Lin[2], Tsung-Ming Yang[2], Jrhau Lung[6], Chih-Cheng Chen[2,7], Kuan-Der Lee[8], Ying-Huang Tsai[9,10]

[1]Department of Pulmonary and Critical Care Medicine, Chang Gung Memorial Hospital, Chiayi Branch, Taiwan; [2]Department of Medicine, College of Medicine, Chang Gung University, Taoyuan, Taiwan; [3]Department of Respiratory Care, Chang Gung University of Science and Technology, Chiayi Campus, Chiayi, Taiwan; [4]Department of Pathology, Chang Gung Memorial Hospital, Chiayi Branch, Taiwan; [5]Department of Surgery, Division of Thoracic and Cardiovascular Surgery, Chang Gung Memorial Hospital, Chiayi Branch, Taiwan; [6]Department of Medical Research and Development, Chang Gung Memorial Hospital, Chiayi Branch, Taiwan; [7]Department of Hematology and Oncology, Chang Gung Memorial Hospital, Chiayi Branch, Taiwan; [8]Department of Hematology and Oncology, Taipei Medical University Hospital, Taipei 110, Taiwan; [9]Department of Respiratory Care, College of Medicine, Chang Gung University, Taoyuan, Taiwan; [10]Department of Pulmonary and Critical Care Medicine, Chang Gung Memorial Hospital, Linkou Branch, Taiwan

Received April 18, 2021; Accepted March 20, 2022; Epub May 15, 2022; Published May 30, 2022

Abstract: *ROS1* fusion genes are rare but important driver genes in lung cancer. Owing to their rarity, many clinicopathological features and treatment responses for each *ROS1* fusion variant are still largely unknown and require further investigation. RNA is the preferable template for the *ROS1* fusion gene screening, but deterioration of RNA in FFPE often makes the detection challenging. To resolve the difficulty, a targeted chromosomal breakpoint sequencing method was developed for searching the *ROS1* fusion gene, and was compared with fluorescence in situ hybridization, immunohistochemistry, RT-qPCR using 260 lung cancer samples of Southern Taiwan. The results showed that *ROS1*-altered cases were present at low frequencies, did not share distinct clinicopathological features, and often carried other driver mutations. The performance of the targeted sequencing assay was superior to the RT-qPCR in *ROS1* fusion gene identification when the cDNAs were from FFPE samples, but long-read DNA sequencing and fresh-frozen samples would be better to revolve all fusion genes. Precise determination of all *ROS1* fusion variants and concomitant driver mutations using both genomic DNA and RNA would be required to help improve the treatment of patients with *ROS1* alterations.

Keywords: Lung cancer, *ROS1* fusion gene, targeted chromosomal breakpoint sequencing, FISH, IHC, RT-qPCR

Introduction

Fusion genes rising from chromosomal rearrangement are potent divers prompting strong cancer cell growth [1]. Through joining other genes, fusion genes are liberated from the normal regulatory circuit and exert their pathogenic activities by changing expression, eliminating regulatory domain, forcing oligomerization, or changing subcellular localization. Tyrosine kinases and transcription regulators are the two biggest categories of fusion genes, but genes in other functional categories also exist, such as *IgH-BCL2*. In view of the importance of fusion genes in tumor pathogenesis, their detection, clinicopathological features, and mechanism of tumorigenesis are always under extensive study in the field of oncology.

ROS1 is a receptor tyrosine kinase belonging to the ALK/leukocyte tyrosine kinase and insulin receptor RTK families [2]. *ROS1* fusion genes in the solid tumor were first identified from a glioblastoma cell line [3], and were subsequently identified in various types of cancer, including inflammatory myofibroblastic tumor [4], cholangiocarcinoma [5], ovarian cancer [6], gastric cancer [7], colorectal cancer [8], angiosarcoma

[9], spitzoid melanoma [10], and non-small cell lung cancer (NSCLC) [11]. Currently, more than 20 different *ROS1* fusion partners have been identified. Some contain only the N-terminal portion of ROS1 without the kinase domain, such as those fused with *RNPC3* [12] and *CEP72* [13], and their functional importance awaits further elucidation. Others create novel fusion oncogenic kinases, such as those with *CD74*, *SLC34A2*, *SDC4*, *EZR*, *GOPC* (or *FIG*), *TPM3*, *LRIG3*, *KDELR2*, *CCDC6*, *MSN*, *TMEM106B*, *TPD52L1*, *CLTC*, and *LIMA1* [14]. With the dimerization domain lacking in most ROS1 fusion kinases, the activation mechanisms of these ROS1 fusion kinases are not fully understood [15]. In lung cancers, *ROS1* fusion genes tend to be present in patients with adenocarcinoma, no smoking history, young ages, and mutual exclusivity to other driver mutations. The prevalence rate of the *ROS1* fusion gene is around 2-3% in NSCLC, in which adenocarcinoma and advanced stage are more prevalent than non-adenocarcinoma and early stage [16]. A great proportion of *ROS1* fusion gene-positive patients respond well to ROS1 target therapy and the PFS can reach from 10 to 19.6 months in different studies [17-19]. Despite these encouraging results, many *ROS1* fusion genes exist only at low frequencies, and some have even been reported just once. Clinicopathological features and treatment response for many *ROS1* fusion genes and whether different *ROS1* fusion genes confer different levels of kinase activation and oncogenicity remain unclear, awaiting more in-depth study. Currently, detection of *ROS1* fusion genes relies most on FISH and IHC. To better understand the common and unique features of each *ROS1* fusion genes, an efficient molecular diagnostic tool is highly anticipated.

While detection of the fusion gene by FISH is very sensitive and specific, it is quite labor-intensive and could even become false-negative if the fusion partner is located close to each other on the same chromosome arm. To improve the detection of fusion genes, a more efficient and accurate method is highly anticipated. Despite the explicit junction of a fusion transcript could make it become a better template for fusion gene detection, the quality of RNA often becomes deteriorated by the ischemia before fixation and the FFPE processing thereafter [20, 21]. This could make the detection of fusion genes through RNA become challenging, especially when the fusion transcript is at a low level. To overcome the difficulty, determination of the chromosomal breakpoint would be another potential strategy to identify fusion genes, since genomic DNA is comparatively more stable than RNA during tissue processing and storage due to a less hydroxyl group on the ribose. Although the sequence preference for the chromosomal rearrangement in fusion gene formation is not clear at this moment and the determination would need to analyze much more sequence regions than using RNA, the comprehensiveness of the next-generation sequencing should be able to achieve this goal by analyzing the genomic sequences near the new exon junction of the fusion gene. To establish efficient methodologies able to identify and discriminate all possible *ROS1* fusion genes, a RT-qPCR (Supplementary Figure 1 and Supplementary Table 1) and a targeted chromosomal breakpoint sequencing protocols (**Figure 1**) were established and compared with FISH and IHC using a Taiwanese lung cancer population. Study results show that *ROS1* fusion genes are present at low frequency, and do not share distinct clinicopathological features. The developed targeted sequencing has a higher success rate in identifying *ROS1* fusion gene from formalin-fixed paraffin-embedded (FFPE) samples compared with RT-qPCR. Comprehensive molecular mutation profiling uses both genomic DNA and RNA to uncover *ROS1* fusion variants and concomitant driver mutations would be important for further improving the outcome and survival of patients carrying *ROS1* fusion genes.

Materials and methods

Specimen and cell line collection

A total of 260 FFPE surgical lung cancer samples collected from 2006 to 2020 with signed informed consent were requested from the tissue bank and biobank of Chang Gung Memorial Hospital, Chiayi after approval from the institutional review board (IRB No: 201600631B0). Two tumor parts and one adjacent normal part were sampled from each specimen in a 1.5-mm diameter core format for tissue microarray construction and were used for the downstream FISH and IHC analyses. The clinicopathological features of these samples are summarized in **Table 1**. Two *ROS1* fusion gene-positive cell lines, U118 (*GOPC-ROS1*) and HCC78

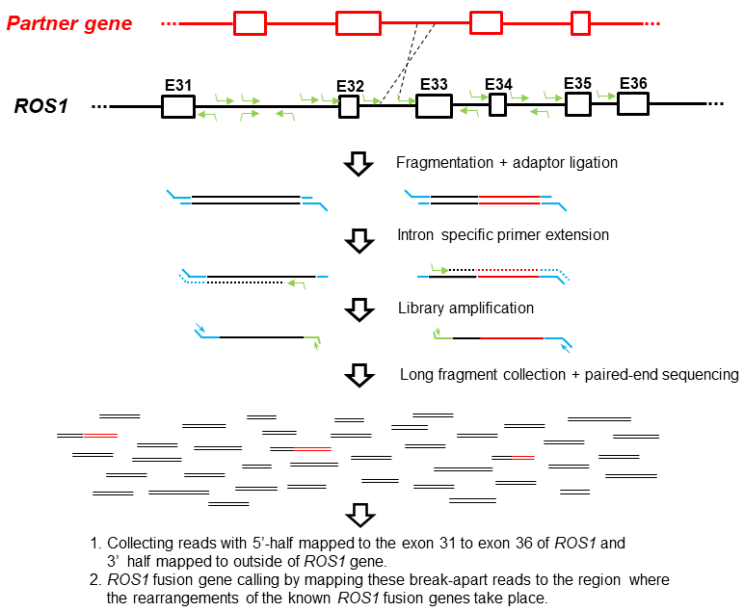


Figure 1. The workflow of the targeted chromosomal breakpoint sequencing to identify the *ROS1* fusion genes. The 17143 bp sequence around the new exon junctions of *ROS1* gene of all known *ROS1* fusion genes were amplified and sequenced by the primer on the universal adaptor (labeled with blue color) ligated to the terminals of DNA fragments and the gene-specific primer (labeled with green color) for the targeted sequencing. After PCR amplification, fragments over 500 bp were gel purified and resolved by paired-end sequencing. Reads with 5'- and 3'-half mapped within and outside the targeted sequencing regions, respectively, were collected and interpreted using a dataset composed of the position information of introns adjacent to the new exon junctions of all known *ROS1* fusion genes to identify the carried fusion genes in samples.

(*SLC34A2-ROS1*) were obtained from ATCC and DSMZ, respectively, and were maintained following the instructions from the cell banks.

Fluorescence in situ hybridization

For *ROS1* fusion gene detection, FISH was performed on tissue microarray using the SureFISH *ROS1* BA Probe Mix (Agilent Technologies, Glostrup, Denmark) with the Histology FISH Accessory Kit (Agilent Technologies). First, TMA slides were deparaffinized, rehydrated, and pretreated at 98°C for 10 min. After cooling to room temperature, the slides were washed twice with wash buffer, followed by air drying. Tissue sections were then digested with pepsin at room temperature for 18 min. Pepsin digestion was terminated by washing thrice with wash buffer. The reconstituted fluorescence probes were then overlaid onto the TMA sections and incubated inside the Dako Hybridizer. Hybridization proceeded with initial heating up to 90°C for 5 min, followed by gradual cooling to 37°C for overnight incubation. After hybridization, non-specific binding was eliminated by washing thrice with Stringent Wash buffer at 60°C. For signal visualization, tissue sections were dehydrated and mounted in an antifade mounting medium with DAPI (Invitrogen, Carlsbad, CA). FISH signals were observed using Axio Scope A1 fluorescence microscope (Carl Zeiss, Oberkochen, Germany) and recorded using the ISIS imaging software (MetaSystems, Altlussheim, Germany). *ROS1* fusion gene-positive was defined as the split of red and green signal greater than two signal widths apart, and/or loss of one green signal (5' probe) occurred in at least 15% of 100 counted nuclei.

Immunohistochemistry

For ROS1 protein detection, the anti-total ROS1 rabbit monoclonal antibody (clone D4D6, Cell Signaling Technology, Danvers, MA) was used. Staining was performed on the Ventana BenchMark XT platform (Ventana Medical Systems, Tucson, AZ) with the Ventana OptiView detection kit. The results were interpreted by pathologists blinded to the clinical information of the studied patients. Only cytoplasmic and membranous staining were interpreted as positive, and the intensity was classified on a four-tier system: 0, no staining; 1+, weak; 2+, moderate; and 3+, strong.

ROS1 fusion gene variant identification by RT-qPCR

Total RNA was purified from the FFPE samples using the AllPrep DNA/RNA Mini kit (Qiagen, 80204) according to the manufacturer's protocol. For cDNA preparation, 1 μg of total RNA from each sample was reverse transcribed using SuperScript III (Thermo Fisher Scientific, 18080093) with random hexamer and gene-specific reverse primers. Four separate PCR reactions were carried out to cover 14 *ROS1* fusion genes (20 different variants) according

Table 1. Clinicopathological features of lung cancer patients in the study

Variable	group	No. (%)
Age	Median	64
	Range	33-96
Gender	Male	155 (59.6)
	Female	105 (40.4)
Smoking†	Never or light	170 (65.4)
	Heavy	90 (34.6)
Stage	IA	49 (18.8)
	IB	102 (39.2)
	IIA	19 (7.3)
	IIB	23 (8.8)
	IIIA	49 (18.8)
	IIIB	8 (3.1)
	IV	10 (3.8)
Histology	Adeno	185 (71.2)
	Squamous	40 (15.4)
	Adenosquamous	10 (3.8)
	Large cell	8 (3.1)
	Sarcomatoid	5 (1.9)
	Others	12 (4.6)
Total		260 (100)

†Smoking status; Never: 0 pack-year; Light: 0-20 pack-years; Heavy: >20 pack-years.

to the joint exon of the *ROS1* gene (Supplementary Figure 1). The primer and probe were selected and evaluated using ThermoFisher Multiple primer analyzer, and MFEprimer-3.0 [22], and their sequence and labeling information are listed in the Supplementary Table 1. Reagents for *ACTB* gene are included as an internal control in each reaction to ensure the reaction setup and detection. The reaction mixture contains 1× Qiagen QuantiNova Probe master mix (Qiagen, 208254), 100 nM of internal control primers, 200 nM of specific primers, and TaqMan probes in a 20-μl reaction volume. Real-time PCR was performed on the Qiagen Rotor-gene Q qPCR machine with the following cycling condition: 95°C for 5 min, 45 cycles of 95°C for 5 s, and 60°C for 10 s. The qPCR products were electrophoresed in 4% agarose gel to confirm the results.

ROS1 fusion gene prediction by targeted chromosomal breakpoint sequencing

The 17143 bp sequence from intron 31 to 35 of *ROS1* gene (Chr6: 117641193-117658335, GRCh37) for *ROS1* FISH positive or IHC positive samples were analyzed by adaptor ligation and single gene-specific primer enrichment-based targeted sequencing (**Figure 1**). The primers used for amplification of DNA fragments in this region were designed using Qiagen QIAseq Target DNA Custom Panel designed tool, and 161 primers (listed in the Supplementary Table 2) were chosen to cover 89.21% of the interested region. The DNA sequence library was constructed using the QIAseq target DNA panel (Qiagen, 333512) and resolved using the MiniSeq High output kit (300-cycle) (Illumina, FC-420-1002). Sequencing results were filtered with a quality score over Q20 and a minimum length of 100 bp using Seqkit [23]. The overlapped and non-overlapped pair-end read were then merged using the bbmerge.sh and the fuse.sh in the bbmap package, respectively [24]. The merged paired-end sequencing results were then aligned against a locally built GRCh37 BLAST database using default settings with e-value less than 1e-40 and identity more than 98%. The BLASTN results were then filtered to collect broken-pair reads potentially coming from *ROS1* fusion genes, if the 5'-half mapped within the *ROS1* intron 31 to 35 and 3'-half mapped outside the *ROS1* intron 31 to 35. The information of the aligned position of these potential reads was then queried against a database containing the genomic coordinates of introns where rearrangements of known *ROS1* fusion genes took place to call putative *ROS1* fusion genes.

Results

Characteristics of ROS1 rearrangement cases identified by FISH

Seven FISH-positive cases were identified from the 260 lung cancer samples (**Table 2**). There are three females and four males, three smokers, and four non-smokers; and five patients aged below 62 and below the median ages of the studied population (64 years). Most of these cases are in their early cancer stage (4 in IB, and 1 in IIA), with two in advanced stage (IIIA and IV). The histology types of most *ROS1* rearrangements cases belong to adenocarcinoma (5/7), which comprises 2.7% (5/185) of the total studied adenocarcinoma cases, and are acinar predominant (2/5). The other two cases are squamous and large-cell types,

Table 2. The clinicopathological features of patients carrying *ROS1* fusion genes and overexpression

Case	FISH/IHC†	Fusion variant‡	Gender	Age	Stage	Histology	Smoking	Additional mutations
1	+/-	X	F	57	IB	Adeno (acinar)	NA	*EGFR* Ex19 Del
2	+/-	X	M	69	IB	Squamous	+	
3	+/3+3+	*CD74-ROS1* (E6:E34)	F	33	IB	Adeno (acinar)	NA	
4	+/-	X	M	61	IIA	Adeno (solid)	+	*KRAS* G13C
5	+/-	X	M	61	IV	Large cell	+	
6	+/2+2+	*EZR-ROS1* (E9:E32/E9:E34)	M	44	IIIA	Adeno (mucinous)	NA	
7	+/2+ (focal) -	X	F	74	IB	Adeno (lepidic)	NA	
8	-/2+-	X	M	59	IIA	Squamous	+	
9	-/2+2+ (focal)	X	M	70	IIIA	Adeno (solid+acinar)	+	
10	-/2+ (focal) -	X	M	56	IIIA	Adeno (acinar)	NA	*EGFR* Ex19 Del
11	-/3+2~3+	X	M	77	IIIA	Adeno (papillary)	NA	*EGFR* Ex19 Del

†The result of IHC staining was labeled according to 4-titer system for each tumor core (2 cores/sample). ‡Fusion variant was labeled with gene name and exon number closest to the fusion junction.

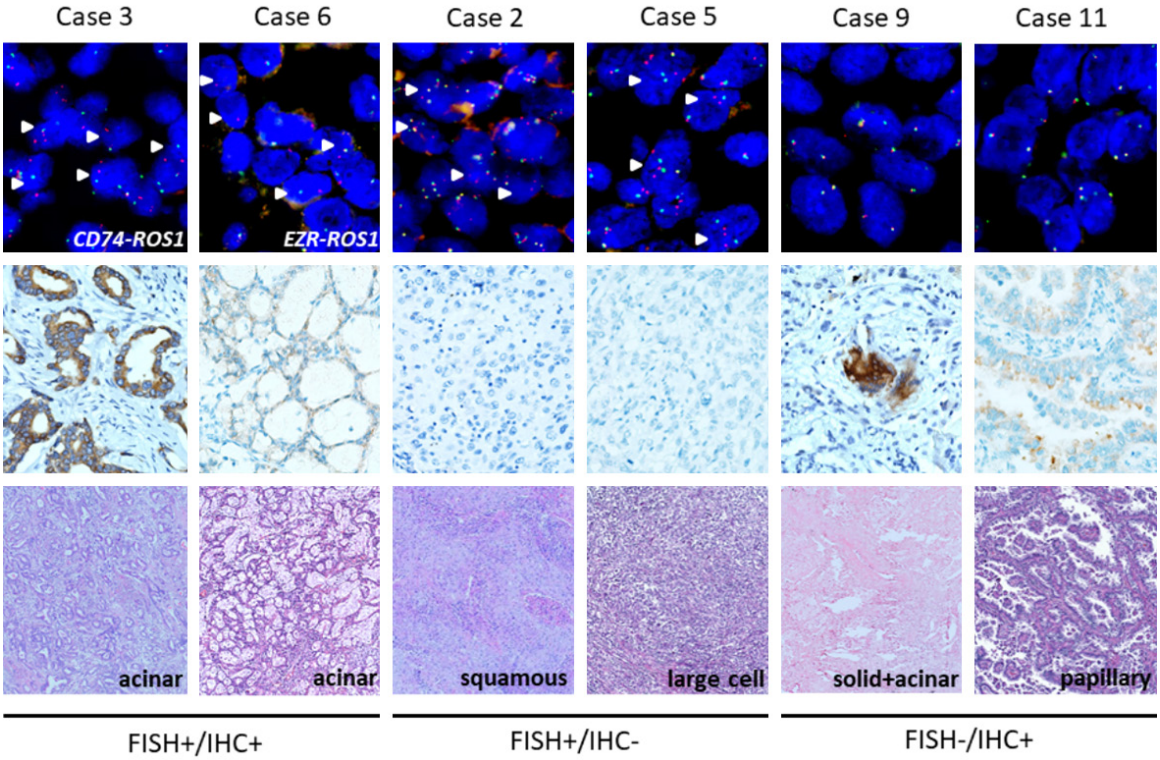


Figure 2. FISH, IHC, and H&E staining results for selected *ROS1* rearrangement and overexpression cases. FISH, IHC, and H&E staining images were captured with 63X, 40X, and 10X magnification respectively. White arrows in FISH images labeled break-apart signals of *ROS1* gene. The case numbers were labeled following **Table 2**. The histological types and identified *ROS1* fusion genes were labeled on the right bottom corner of the H&E and the FISH images, respectively.

which account for 2.5% (1/40) and 12.5% (1/8) in their histological groups (**Figures 1** and **2**).

ROS1 protein expression

To study the ROS1 protein in the 7 *ROS1* FISH positive cases and the other 253 FISH negative cases, IHC was performed on tissue microarray slides. Seven samples are found to have ROS1 overexpression (**Table 2**), in which three are from FISH positive cases and the remaining four are from FISH negative cases (**Figure 2** and **Table 2**). Among the three *ROS1* FISH positive cases, moderate to strong cyto-

plasmic staining throughout the tumor cells could be observed in two cases (**Figure 2**, Case 3 and 6) and focal moderate staining was observed in one case (Case 7, **Table 2**). In the four IHC positive/FISH negative cases, two have moderate staining intensities in the cytoplasm (**Figure 2**, Case 11) and the other two have focal strong staining in a small subset of cells (**Figure 2**, Case 9).

ROS1 fusion gene determination and prediction

The *ROS1* fusion gene variants in ten FISH positive and/or IHC positive cases were initially determined using TaqMan-based RT-qPCR by adopting amplicons as short as possible. Even so, no fusion signal was observed in these samples despite successful amplification of the internal control (*ACTB*) in each sample and the *ROS1* fusion transcripts in the U118 and HCC78 cells (data not shown). This is probably due to the no intact *ROS1* fusion transcript was left in tissue samples after the formalin fixation and paraffin-embedding procedure. To elucidate the molecular mechanism contributing to the *ROS1* alterations in FISH positive and protein overexpression cases, the sequences from the intron 31 to 35 of *ROS1* gene of these samples were analyzed using the next-generation sequencing. Due to being unable to cover the three low-complexity regions (chr6: 117652639-117652925, chr6: 117655667-117656144, and chr6: 117654495-117655315, GRCh37) by the developed sequencing panel notified by the NGS primer designing program, longer fragments in the library were collected for sequencing and hope to improve the chance to read through the three regions. U118 and HCC78 cells were also included to verify the performance of the assay. The base-calling was in good quality and over 80% was in Q30. With the minimum length of 100 bp and Q20 as cutoffs, the average length of reads after adaptor trimming of R1 and R2 reads were in 144.95 and 121.39, respectively. The read numbers for these samples were at least 1.22 million and 1.15 million for R1 and R2, respectively, and the coverage for each sample was all over 18000. Nevertheless, the coverage of the three low-complexity regions in intron 31 was not improved much and all samples only have less than 30 reads for each of these regions. After performing the bioinformatics procedures described in Materials and Methods, the fusion genes were successfully called out for U118 and HCC78 cells, and two FISH/IHC double-positive cases which carried *GOPC-ROS1* fusion, *SLC34A2-ROS1* fusion, *CD74-ROS1* fusion (Case 3) and *EZR-ROS1* fusion (Case 6), respectively (Supplementary Table 3). The results were further supported by the sequences of chromosomal breakpoints and exon junctions determined using cDNA and genomic DNA prepared from the two cell lines and the fresh frozen specimens of the two FISH and IHC double-positive patients, and Sanger sequencing (**Figure 3**).

Discussion

ROS1 fusion genes are rare but important mutant driver genes in lung cancers, but their clinical information remains largely unknown. To learn more about *ROS1* fusion genes, a novel targeted chromosomal breakpoint sequencing methodology was developed to facilitate fusion gene determination and was compared with FISH, IHC, and RT-qPCR in *ROS1* fusion gene screening. In view of the low frequency of *ROS1* fusion genes, enrichment strategy may help narrow down candidates for downstream diagnosis and treatment. Criteria for enrichment include adenocarcinoma, no smoking history, young age, and mutual exclusivity to other driver mutations. From the FISH and IHC positive cases in the current study, these enrichment criteria seem incapable of reliable identifications; with three FISH positive cases being smokers and two FISH positive cases not with adenocarcinoma histology (**Table 2**). These exceptions should urge caution in patient enrichment and mutation screening scopes in clinical practice, especially when using negativity of other driver gene mutations, such as EGFR, to select patients for *ROS1* screening. The concurrence of *ROS1* fusion genes and *EGFR* mutations are generally believed to be below 1% among *EGFR* mutation cases [25-27]. Nevertheless, significant frequency will be seen if inverse data interpretation by calculating the frequency of *EGFR* mutations in *ROS1* fusion gene-positive cases, since the relative frequencies of concurrent mutation in two intersected mutation populations would be always higher in the one of smaller size. In consistent with this speculation, a significant proportion of *EGFR* mutations were observed

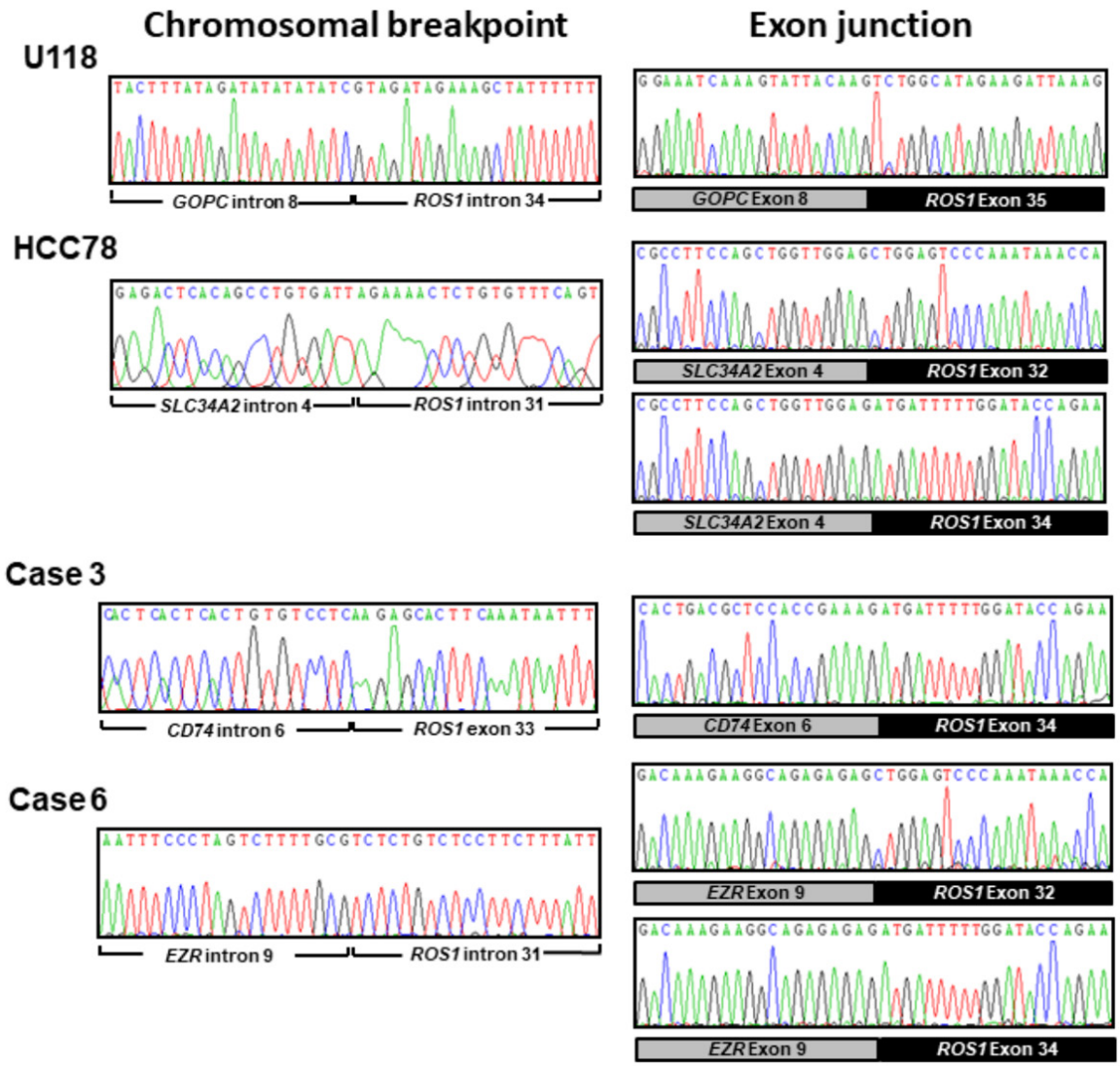


Figure 3. The chromosomal breakpoints and exon junctions of the identified *ROS1* fusion genes in U118 cell, HCC78 cell, and 2 FISH/IHC double-positive cases. To verify the targeted sequencing results, the chromosomal breakpoints and exon junctions of the fusion genes were determined by Sanger sequencing.

among *ROS1* fusion gene cases in this study and another report [28]. The same observation was also found in another lung cancer driver fusion gene, *ALK* [29, 30]. The same situation would also hold true for other co-driver mutations in *ROS1* positive cases, such as *KRAS*. The non-mutually exclusive examples in cases with *ROS1* fusion before treatment may imply possible occurrence of concomitance after treatment and may contribute to resistance. Such phenomenon began to be seen in lung cancer patients after treatment by adopting more sensitive and high-throughput techniques [31, 32]. The concurrence of driver mutations could generally lead to suboptimal response when targeting only one of them; hence, comprehensive profiling of somatic driver mutations should be performed both before and after treatment to further improve the outcome of *ROS1* positive patients. The best treatment strategy, whether sequential, combination, or using broad-spectrum TKI, for these patients merits more clinical studies to confirm.

Current detection of *ROS1* fusion genes relied mostly on FIHS and immunohistochemistry. There has been great hope for the more convenient-to-use immunohistochemistry to replace FISH for fusion gene screening because fusion genes seem to be controlled by stronger pro-

moters than their native ones. Nevertheless, inconsistent FISH and IHC results in *ROS1* detection are frequently found [33-35] and paracentric rearrangement could make the situation even more complicated. Currently, the expression levels of many *ROS1* fusion genes are still not fully deciphered, as many of them have only been reported once and their protein expressions were not measured in their original reports. According to the gene expression profile data of UniGene in the NCBI database, the transcript per million (TPM) values for some of the *ROS1* fusion partners, including *CD74, EZR, TPM3, CLTC, MSN, LC34A2, SDC4, and TPD52L1*, are higher, ranging from 77 to 992. It could be expected that the expression levels of these *ROS1* fusion genes will be significantly upregulated and show stronger IHC signals, as the TPM value of *ROS1* gene is only 14 in normal lung tissue. On the other hand, the TPM values for other *ROS1* fusion partners, such as *TFG, LIMA1, ZCCH8, TMEM106B, CCDC6, GOPC, RNPC3, and LRIG3,* in lung cancer tissue range from 0 to 38, and are close to that of *ROS1* gene in normal lung tissue. Although the transcription of these *ROS1* fusion partner genes may be augmented during tumorigenesis, it is likely that each *ROS1* fusion variant would not have an equal protein expression level and comparable oncogenic potential. Several *ROS1* genes have been shown to confer different expressions and oncogenicities in mouse model systems [36, 37]. This could be the reason why only two out of seven *ROS1* FISH-positive cases showed strong IHC signals in the current study (**Table 2**). However, resolving fusion genes in these cases is beyond the discrimination power of FISH and IHC and must rely on discrimination in the nucleic acid level. In addition, this would be the only way to answer whether different *ROS1* fusion genes pose different clinicopathological features. Among nucleic acids, RNA offers many advantages over genomic DNA for fusion gene identification, such as being explicit in fusion junction and more abundant templates for detection, but deterioration of RNA quality during the tissue processing always makes the task challenging as shown in the current study and many others. The targeted sequencing searching chromosomal breakpoint offers another potential solution to decoding the *ROS1* fusion genes, and the present findings prove the feasibility and successful identification of fusion genes for the two *ROS1* fusion gene-positive cell lines and two out of seven FISH positive cases. Over 75% of the *ROS1* fusion gene cases involve with the *ROS1* exon 32 and 34 [14], and the three low-complexity regions in the intron 31 contain several repeated sequences, including AT and GAA repeats, frequently found at the chromosomal translocation site [38]. Whether the low coverage in the three low-complexity by short-read sequencing leads to the false-negative results for the five FISH-positive cases remains speculative; and imply the drawback of the developed short pair-end targeted sequencing for fusion gene identification and may necessitate using long-read DNA sequencing technologies, such as Oxford Nanopore or Pacbio SMRT, with a fresh-frozen sample to further confirmation. Additionally, the repairment of double-strand break involved in chromosomal rearrangement may introduce new "template sequence insertions" (TSIs) derived from distant regions of the genome in lengths ranging from tens to 3 thousand base pairs [39]. This could make the chromosomal breakpoint become more implicit and further imply that a long-read sequencing may be better to help resolve fusion genes based on the breakpoint information. Since the CosmicFusionExport dataset only labels the coordinates of exons involved in the formation of *ROS*1 fusion genes, and the CosmicBreakpointsExport dataset does not include the breakpoint data for *ROS1* fusion genes, the direct answer for the speculation is not available at this moment. It could be anticipated, however, that further dissection of the translocation hotspots for *ROS1* and all other driver fusion genes may eventually turn fusion gene screening into sequencing recombinant hotspots for providing more precise treatment for each fusion gene. Despite promising, since the chromosomal breakpoint may not always take place only in the intron adjoining to the exons of the new exon junction as shown in the current study (HCC78, Case 3, and Case 6), it should also include the nearby exons and introns in the sequencing panel to identify all fusion variants. Besides, some breakpoints created by the rearrangement seem to change the splicing and may cause different degrees of exon skipping as shown in the cases of HCC78 and case 6 of the study, as mutations caused two and even multiple exon skipping have been reported before [40, 41]. This could cause misinterpretation of the fu-

sion variant based solely on the information of breakpoints. These results urge cautions in the future sequencing panel design and optimization of the bioinformatics algorithms for the fusion variant identification if using genomic DNA for searching.

In summary, the current study investigated the clinicopathological features of *ROS1* fusion genes and developed a targeted chromosomal breakpoint sequencing method and a RT-qPCR method, which were compared with other screening methods in *ROS1* fusion gene identification. Results show that *ROS1* fusion genes are present at low frequencies (2.7%) and do not share distinct clinicopathological features, such as no smoking history, adenocarcinoma, strong IHC staining, and mutual exclusivity to other driver mutations, and would require direct detection in the nucleic acid level to accurately identify and discriminate various *ROS1* fusion variants. When using FFPE samples, the developed targeted sequencing shows better performance in *ROS1* fusion gene identification compared with RT-qPCR. However, further improvement for resolving all variants would require long-read sequencing with fresh-frozen samples and further investigating the effects of the chromosomal rearrangement on splicing. Comprehensive somatic mutation profiling to uncover concomitant driver mutations and *ROS1* fusion variants would be the future direction for further improvement in treatment outcome and survival for patients carrying *ROS1* fusion genes.

Acknowledgements

The study was supported by COPRG6K0011-3, CORPG6K0031-3, CORPG6K0041-3, and CMRPG6F0471-3 from Chang Gung Memorial Hospital, Chiayi. We would like to acknowledge the Precious Instrumentation Core Laboratory, Department of Medical Research and Development, Chang Gung Memorial Hospital, Chiayi for providing tissue microarray construction service using the AutoTiss 1000 tissue microarrayer (EverBio, New Taipei City, Taiwan, ROC). We also would like to thank Dr. Chien-Hsing Lin of Feng Chi Biothech Corp (Taipei, Taiwan, ROC) for the assistance in primer design used for the targeted sequencing.

Disclosure of conflict of interest

None.

Abbreviations

NSCLC, Non-Small Cell Lung Cancer; FFPE, Formalin-Fixed Paraffin-Embedded; FISH, Fluorescence In Situ Hybridization; RT-qPCR, Real Time Quantitative Polymerase Chain Reaction; TPM, Transcript Per Million.

Address correspondence to: Jrhau Lung, Department of Medical Research and Development, Chang Gung Memorial Hospital, Chiayi Branch, No. 6 West Section, Chia-Pu Road, Pu-Tz City, Chiayi County 613, Taiwan (ROC). Tel: 886-53621000-2756; E-mail: jrhaulung@gmail.com

References

[1] Edwards PA. Fusion genes and chromosome translocations in the common epithelial cancers. J Pathol 2010; 220: 244-254.

[2] Robinson DR, Wu YM and Lin SF. The protein tyrosine kinase family of the human genome. Oncogene 2000; 19: 5548-5557.

[3] Birchmeier C, Sharma S and Wigler M. Expression and rearrangement of the ROS1 gene in human glioblastoma cells. Proc Natl Acad Sci U S A 1987; 84: 9270-9274.

[4] Lovly CM, Gupta A, Lipson D, Otto G, Brennan T, Chung CT, Borinstein SC, Ross JS, Stephens PJ, Miller VA and Coffin CM. Inflammatory myofibroblastic tumors harbor multiple potentially actionable kinase fusions. Cancer Discov 2014; 4: 889-895.

[5] Gu TL, Deng X, Huang F, Tucker M, Crosby K, Rimkunas V, Wang Y, Deng G, Zhu L, Tan Z, Hu Y, Wu C, Nardone J, MacNeill J, Ren J, Reeves C, Innocenti G, Norris B, Yuan J, Yu J, Haack H, Shen B, Peng C, Li H, Zhou X, Liu X, Rush J and Comb MJ. Survey of tyrosine kinase signaling reveals ROS kinase fusions in human cholangiocarcinoma. PLoS One 2011; 6: e15640.

[6] Birch AH, Arcand SL, Oros KK, Rahimi K, Watters AK, Provencher D, Greenwood CM, Mes-Masson AM and Tonin PN. Chromosome 3 anomalies investigated by genome wide SNP analysis of benign, low malignant potential and low grade ovarian serous tumours. PLoS One 2011; 6: e28250.

[7] Lee J, Lee SE, Kang SY, Do IG, Lee S, Ha SY, Cho J, Kang WK, Jang J, Ou SH and Kim KM. Identification of ROS1 rearrangement in gastric adenocarcinoma. Cancer 2013; 119: 1627-1635.

[8] Aisner DL, Nguyen TT, Paskulin DD, Le AT, Haney J, Schulte N, Chionh F, Hardingham J, Mariadason J, Tebbutt N, Doebele RC, Weickhardt AJ and Varella-Garcia M. ROS1 and ALK fusions in colorectal cancer, with evidence of intratumoral heterogeneity for molecular drivers. Mol Cancer Res 2014; 12: 111-118.

[9] Giacomini CP, Sun S, Varma S, Shain AH, Giacomini MM, Balagtas J, Sweeney RT, Lai E, Del Vecchio CA, Forster AD, Clarke N, Montgomery KD, Zhu S, Wong AJ, van de Rijn M, West RB and Pollack JR. Breakpoint analysis of transcriptional and genomic profiles uncovers novel gene fusions spanning multiple human cancer types. PLoS Genet 2013; 9: e1003464.

[10] Wiesner T, He J, Yelensky R, Esteve-Puig R, Botton T, Yeh I, Lipson D, Otto G, Brennan K, Murali R, Garrido M, Miller VA, Ross JS, Berger MF, Sparatta A, Palmedo G, Cerroni L, Busam KJ, Kutzner H, Cronin MT, Stephens PJ and Bastian BC. Kinase fusions are frequent in Spitz tumours and spitzoid melanomas. Nat Commun 2014; 5: 3116.

[11] Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, Hu Y, Tan Z, Stokes M, Sullivan L, Mitchell J, Wetzel R, Macneill J, Ren JM, Yuan J, Bakalarski CE, Villen J, Kornhauser JM, Smith B, Li D, Zhou X, Gygi SP, Gu TL, Polakiewicz RD, Rush J and Comb MJ. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell 2007; 131: 1190-1203.

[12] Chong CR, Bahcall M, Capelletti M, Kosaka T, Ercan D, Sim T, Sholl LM, Nishino M, Johnson BE, Gray NS and Janne PA. Identification of existing drugs that effectively target NTRK1 and ROS1 rearrangements in lung cancer. Clin Cancer Res 2017; 23: 204-213.

[13] Zhu YC, Zhou YF, Wang WX, Xu CW, Zhuang W, Du KQ and Chen G. CEP72-ROS1: a novel ROS1 oncogenic fusion variant in lung adenocarcinoma identified by next-generation sequencing. Thorac Cancer 2018; 9: 652-655.

[14] Lin JJ and Shaw AT. Recent advances in targeting ROS1 in lung cancer. J Thorac Oncol 2017; 12: 1611-1625.

[15] Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, Asaka R, Hamanaka W, Ninomiya H, Uehara H, Lim Choi Y, Satoh Y, Okumura S, Nakagawa K, Mano H and Ishikawa Y. RET, ROS1 and ALK fusions in lung cancer. Nat Med 2012; 18: 378-381.

[16] Zhu Q, Zhan P, Zhang X, Lv T and Song Y. Clinicopathologic characteristics of patients with ROS1 fusion gene in non-small cell lung cancer: a meta-analysis. Transl Lung Cancer Res 2015; 4: 300-309.

[17] Mazieres J, Zalcman G, Crino L, Biondani P, Barlesi F, Filleron T, Dingemans AM, Lena H, Monnet I, Rothschild SI, Cappuzzo F, Besse B, Thiberville L, Rouviere D, Dziadziuszko R, Smit EF, Wolf J, Spirig C, Pecuchet N, Leenders F, Heuckmann JM, Diebold J, Milia JD, Thomas RK and Gautschi O. Crizotinib therapy for advanced lung adenocarcinoma and a ROS1 rearrangement: results from the EUROS1 cohort. J Clin Oncol 2015; 33: 992-999.

[18] Shaw AT, Ou SH, Bang YJ, Camidge DR, Solomon BJ, Salgia R, Riely GJ, Varella-Garcia M, Shapiro GI, Costa DB, Doebele RC, Le LP, Zheng Z, Tan W, Stephenson P, Shreeve SM, Tye LM, Christensen JG, Wilner KD, Clark JW and Iafrate AJ. Crizotinib in ROS1-rearranged non-small-cell lung cancer. N Engl J Med 2014; 371: 1963-1971.

[19] Wu YL, Yang JC, Kim DW, Lu S, Zhou J, Seto T, Yang JJ, Yamamoto N, Ahn MJ, Takahashi T, Yamanaka T, Kemner A, Roychowdhury D, Paolini J, Usari T, Wilner KD and Goto K. Phase II study of crizotinib in East Asian patients with ROS1-positive advanced non-small-cell lung cancer. J Clin Oncol 2018; 36: 1405-1411.

[20] Evers DL, He J, Kim YH, Mason JT and O'Leary TJ. Paraffin embedding contributes to RNA aggregation, reduced RNA yield, and low RNA quality. J Mol Diagn 2011; 13: 687-694.

[21] Guerrera F, Tabbo F, Bessone L, Maletta F, Gaudiano M, Ercole E, Annaratone L, Todaro M, Boita M, Filosso PL, Solidoro P, Delsedime L, Oliaro A, Sapino A, Ruffini E and Inghirami G. The influence of tissue ischemia time on RNA integrity and patient-derived xenografts (PDX) engraftment rate in a non-small cell lung cancer (NSCLC) Biobank. PLoS One 2016; 11: e0145100.

[22] Wang K, Li H, Xu Y, Shao Q, Yi J, Wang R, Cai W, Hang X, Zhang C, Cai H and Qu W. MFEprimer-3.0: quality control for PCR primers. Nucleic Acids Res 2019; 47: W610-W613.

[23] Shen W, Le S, Li Y and Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One 2016; 11: e0163962.

[24] Bushnell B, Rood J and Singer E. BBMerge - accurate paired shotgun read merging via overlap. PLoS One 2017; 12: e0185056.

[25] Hu W, Liu Y and Chen J. Concurrent gene alterations with EGFR mutation and treatment efficacy of EGFR-TKIs in Chinese patients with non-small cell lung cancer. Oncotarget 2017; 8: 25046-25054.

[26] Lin JJ, Ritterhouse LL, Ali SM, Bailey M, Schrock AB, Gainor JF, Ferris LA, Mino-Kenudson M, Miller VA, Iafrate AJ, Lennerz JK and Shaw AT. ROS1 fusions rarely overlap with other oncogenic drivers in non-small cell lung cancer. J Thorac Oncol 2017; 12: 872-877.

[27] Mao Y and Wu S. ALK and ROS1 concurrent with EGFR mutation in patients with lung adenocarcinoma. Onco Targets Ther 2017; 10: 3399-3404.

[28] Chen RI, Yang JJ, Zhang XC, Lou NN, Chen H, Zhou Q, Su J, Guo W, Chen JG, Tu H, Sun Y and Wu YL. ROS1 rearrangement coexists with EGFR mutation in non-small-cell lung cancer. J Clin Oncol 2015; 33: e19003-e19003.

[29] Shin HJ, Kho BG, Kim MS, Park HY, Kim TO, Kim YI, Lim SC, Park CK, Kim YC, Choi YD and Oh IJ. Co-alteration of EGFR mutation and ALK

rearrangement in non-small cell lung cancer: case series. Medicine (Baltimore) 2019; 98: e14699.

[30] Yang JJ, Zhang XC, Su J, Xu CR, Zhou Q, Tian HX, Xie Z, Chen HJ, Huang YS, Jiang BY, Wang Z, Wang BC, Yang XN, Zhong WZ, Nie Q, Liao RQ, Mok TS and Wu YL. Lung cancers with concomitant EGFR mutations and ALK rearrangements: diverse responses to EGFR-TKI and crizotinib in relation to diverse receptors phosphorylation. Clin Cancer Res 2014; 20: 1383-1392.

[31] McCoach CE, Le AT, Gowan K, Jones K, Schubert L, Doak A, Estrada-Bernal A, Davies KD, Merrick DT, Bunn PA Jr, Purcell WT, Dziadziuszko R, Varella-Garcia M, Aisner DL, Camidge DR and Doebele RC. Resistance mechanisms to targeted therapies in ROS1(+) and ALK(+) non-small cell lung cancer. Clin Cancer Res 2018; 24: 3334-3347.

[32] Zhu YC, Wang WX, Zhang QX, Xu CW, Zhuang W, Du KQ, Chen G, Lv TF and Song Y. The KIF5B-RET fusion gene mutation as a novel mechanism of acquired EGFR tyrosine kinase inhibitor resistance in lung adenocarcinoma. Clin Lung Cancer 2019; 20: e73-e76.

[33] Boyle TA, Masago K, Ellison KE, Yatabe Y and Hirsch FR. ROS1 immunohistochemistry among major genotypes of non-small-cell lung cancer. Clin Lung Cancer 2015; 16: 106-111.

[34] Liang D, Wang W, Tang B, Dong D, Li F and Xu G. Detection of ROS1 translocation in lung adenocarcinoma by IHC, FISH and RT-PCR and its clinicopathologic features. Int J Clin Exp Pathol 2017; 10: 1801-1808.

[35] Shan L, Lian F, Guo L, Qiu T, Ling Y, Ying J and Lin D. Detection of ROS1 gene rearrangement in lung adenocarcinoma: comparison of IHC, FISH and real-time RT-PCR. PLoS One 2015; 10: e0120422.

[36] Inoue M, Toki H, Matsui J, Togashi Y, Dobashi A, Fukumura R, Gondo Y, Minowa O, Tanaka N, Mori S, Takeuchi K and Noda T. Mouse models for ROS1-fusion-positive lung cancers and their application to the analysis of multikinase inhibitor efficiency. Carcinogenesis 2016; 37: 452-460.

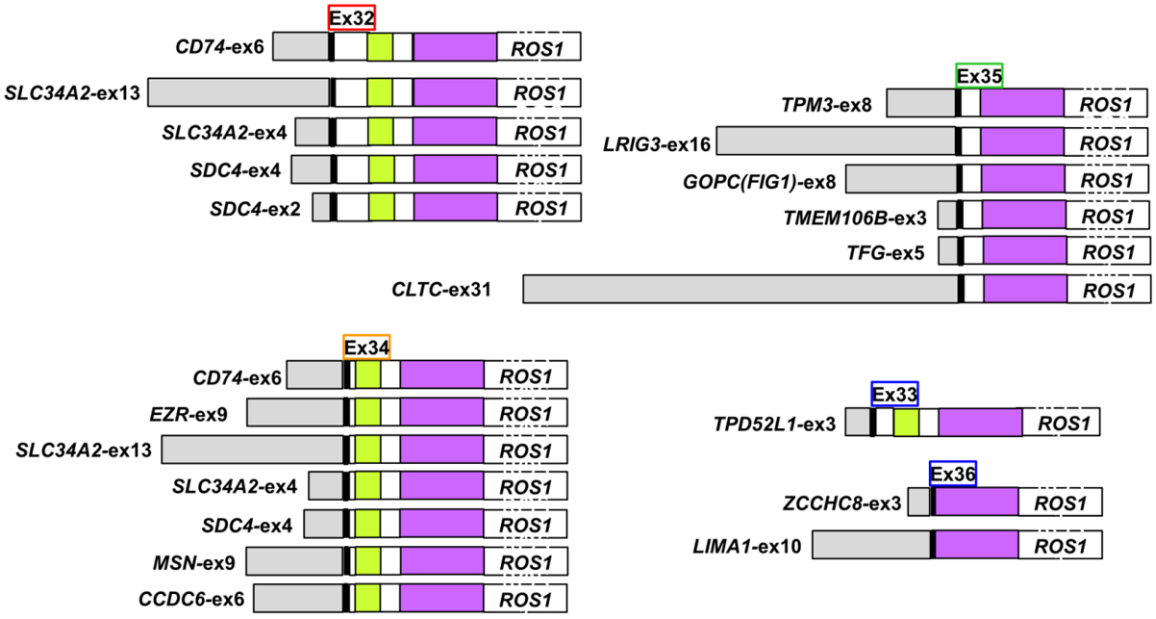
[37] Neel DS, Allegakoen DV, Olivas V, Mayekar MK, Hemmati G, Chatterjee N, Blakely CM, McCoach CE, Rotow JK, Le A, Karachaliou N, Rosell R, Riess JW, Nichols R, Doebele RC and Bivona TG. Differential subcellular localization regulates oncogenic signaling by ROS1 kinase fusion proteins. Cancer Res 2019; 79: 546-556.

[38] Bacolla A, Tainer JA, Vasquez KM and Cooper DN. Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. Nucleic Acids Res 2016; 44: 5673-5688.

[39] Onozawa M, Zhang Z, Kim YJ, Goldberg L, Varga T, Bergsagel PL, Kuehl WM and Aplan PD. Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. Proc Natl Acad Sci U S A 2014; 111: 7729-7734.

[40] Gualandi F, Trabanelli C, Rimessi P, Calzolari E, Toffolatti L, Patarnello T, Kunz G, Muntoni F and Ferlini A. Multiple exon skipping and RNA circularisation contribute to the severe phenotypic expression of exon 5 dystrophin deletion. J Med Genet 2003; 40: e100.

[41] Hori T, Fukao T, Murase K, Sakaguchi N, Harding CO and Kondo N. Molecular basis of two-exon skipping (exons 12 and 13) by c.1248+5g>a in OXCT1 gene: study on intermediates of OXCT1 transcripts in fibroblasts. Hum Mutat 2013; 34: 473-480.



Supplementary Figure 1. The grouping of the RT-qPCR assay for the *ROS1* fusion variant detection. The RT-qPCR reactions for *ROS1* fusion variants identification were grouped according to the common *ROS1* exon immediately downstream to the fusion junction. Since only a few variants involved the *ROS1* exon 33 and exon 36, the detection of the two types of variants was rearranged within a single reaction.

Supplementary Table 1. The primer and probe sequence information of the designed *ROS1* RT-qPCR assay

		Forward primer/probe name	Sequence/label	Reverse primer name	Sequence	Amplicon Size (bp)
Reaction 1	*ROS1* fusion gene with exon 32	*SLC34A2*-E4-R32	TCGTGTGCTCCCTGGATATT	*ROS1*-E32R	TTTTTACTCCCTTCTAGTAATTGG	100
		SLC34A2-E13-del2046	CTCCTGAGACCTTTGATAACATA			116
		CD74-E6	AATGAGCAGGCACTCCTT			102
		SDC4-E2	AGGAATCTGATGACTTTGAGC			94
		SDC4-E4-R32	GCAGGGCAGCAACATCTTT			92
		ROS1-E32-probe	FAM-AGTCCCAAATAAACCAGGCAT-BHQ1			
		ACTB-F	CCTTCCTTCCTGGGCATGGAGTC	*ACTB*-R	AGACAGCACTGTGTTGGCGT	111
		ACTB-probe	HEX-TGACGTGGACATCCGCAAAGAC-BHQ1			
Reaction 2	*ROS1* fusion gene With exon 34	*SLC34A2*-E4-R34	GGATATTCTTAGTAGCGCCTTC	*ROS1*-E34R	TTGTAACAACCAGAAATATTCCAAC	104
		SLC34A2-E13-del2046	CTCCTGAGACCTTTGATAACATA			133
		CD74-E6	GAGCAAAAGCCCACTGA			100
		SDC4-E4-R34	CAGCAACATCTTTGAGAGAAC			104
		EZR-E9	GCAGGACTATGAGGAGAAGA			107
		MSN-E9	AACAGACTAAGAAGGCTCAGC			93
		CCDC6-E6	GTCCAGCTTAGAAATGGACGACG			96
		ROS1-E34-probe	FAM-ATGATTTTTGGATACCAGAAACAAGTTTCA-BHQ1			
		ACTB-F	CCTTCCTTCCTGGGCATGGAGTC	*ACTB*-R	AGACAGCACTGTGTTGGCGT	111
		ACTB-probe	HEX-TGACGTGGACATCCGCAAAGAC-BHQ1			
Reaction 3	*ROS1* fusion gene with exon 35	*TPM3*-E8	TAGCCAAGCTGGAAAAGAC	*ROS1*-E35R	TCACCCCTTCCTTGGCACTTT	83
		LRIG3-E16	CACCAGTTTGTCACATCTTCA			106
		GOPC(*FIG1*)-E8	CACAAGTGGGGAAATCAAAGT			77
		TMEM106B-E3	GTCAGGGAACAGGAAGAATTCC			75
		CLCTC-E31	CAAGCTACAGAGACACAACCCAT			79
		TFG-E5	GTTTGGCTTAACAGATGATCAGG			77
		ROS1-E35-probe	FAM-CTGGCATAGAAGATTAAAGAATCA-BHQ1			
		ACTB-F	CCTTCCTTCCTGGGCATGGAGTC	*ACTB*-R	AGACAGCACTGTGTTGGCGT	111
		ACTB-probe	HEX-TGACGTGGACATCCGCAAAGAC-BHQ1			
Reaction 4	*ROS1* fusion gene with exon 33/36	*TPD52L1*-E3	GCAAAAGCTGGCATGACATGCAG	*ROS1*-E33-R	TGTCATCTTCCACCTTAAATTC	86
		ROS1-E33-PROBE	FAM-AGAGCACTTCAAATAATTTACAGAAC-BHQ1			
		ZCCHC8-E2	ATTGAACATTCTGACTCGACCG	*ROS1*-E36R	CGAGGGAAGGCAGGAAGATT	72
		LIMA1-E10	GCTCCTATTGCAACAACAAACTC			73
		ROS1-E36-probe	FAM-TACTCTTCCAACCCAAGAGGAG-BHQ1			
		ACTB-F	CCTTCCTTCCTGGGCATGGAGTC	*ACTB*-R	AGACAGCACTGTGTTGGCGT	111
		ACTB-probe	HEX-TGACGTGGACATCCGCAAAGAC-BHQ1			

Supplementary Table 2. The location and sequence information of the gene-specific primers used in the *ROS1* fusion gene breakpoint sequencing library construction

primer name*	PRIMER SEQUENCE#
>chr6:117640867-117640917	aatgtacagtattgcgttttgTGTGATTTCTGTAGCTATGGATAGGCTTCGACATT
>chr6:117640964-117641014	aatgtacagtattgcgttttgGGTCAATCTTTGTAGATATGGTGATATAATGTGTCAAGGAGTT
>chr6:117641029-117641079	aatgtacagtattgcgttttgGATTTCTCCACTTCCAACTCCTAAGATGTCCACT
>chr6:117641147-117641197	aatgtacagtattgcgttttgAGATTTTCAATCTCCTCTTGGGTTGGAAGAGTACTGT
>chr6:117641255-117641305	aatgtacagtattgcgttttgGTATTTGTGTTTGAGAGAGAAAGAAATATCTCATGGCTTTG
>chr6:117641366-117641416	aatgtacagtattgcgttttgCCCAGACAGAGCATGATGGTCCACTTCTA
>chr6:117641557-117641607	aatgtacagtattgcgttttgGGGTTTATAAGCTTTTCTCTTAGCTATTTCACACATGTGACAT
>chr6:117641648-117641698 reverse complement	aatgtacagtattgcgttttgGAGAAACTTGGCTTTTTAAAGTTAACCTTATGGTGATGG
>chr6:117641767-117641817	aatgtacagtattgcgttttgTGCTCAAAGGAAAGTCACCTCTAAACAGCTGATTTAG
>chr6:117641827-117641877 reverse complement	aatgtacagtattgcgttttgATTTATAATGCAAGGTGTAAAAACAAAGGATTTGAAGTAGACA
>chr6:117641973-117642023	aatgtacagtattgcgttttgCTCATCTATCTTAAGAAGGTTTTGTGGGCAATCAT
>chr6:117642048-117642098 reverse complement	aatgtacagtattgcgttttgGCCTTGTGTTACGTGGTCAGCCTTAAGTCAAT
>chr6:117642243-117642293	aatgtacagtattgcgttttgCCAGGTCACATTTGAGAGTGGAAGAGGAAGTTAT
>chr6:117642231-117642281 reverse complement	aatgtacagtattgcgttttgCCTGGGATGGATGGAAAATTATGTGGTTTTAAG
>chr6:117642341-117642391 reverse complement	aatgtacagtattgcgttttgAACAGTAAGGAGAGTGCCTATGAATTCCGCAT
>chr6:117642356-117642406	aatgtacagtattgcgttttgCCCTTTGCCTAGGTGCTCCATAATGATG
>chr6:117642481-117642531	aatgtacagtattgcgttttgCACTGTCACCCCTTCCTTGGCACTTTT
>chr6:117642609-117642659 reverse complement	aatgtacagtattgcgttttgCATAAAAACACTTCAAATGCACTGTTAACATTTCCTAAAGG
>chr6:117642752-117642802 reverse complement	aatgtacagtattgcgttttgACATTAAATTTTAAGAAGTGAGATTAAAAATGGTGTAGTATGATTTGTGT
>chr6:117642856-117642906 reverse complement	aatgtacagtattgcgttttgTTGGTTTATTTTGACTCGTTTATGGGTGATTTTGAC
>chr6:117642931-117642981 reverse complement	aatgtacagtattgcgttttgTCTTTTCTACTGAGGAATATAACCTGGATGTGACTGAA
>chr6:117643018-117643068 reverse complement	aatgtacagtattgcgttttgCTACATATTCGAGTATGTGTATATGAACAACAATGATAAACACTCTTGTA
>chr6:117643113-117643163 reverse complement	aatgtacagtattgcgttttgCTTATTATATAAGGAAGACAGTACAAATTACTCATGATTCTACAACTAGG
>chr6:117643221-117643271 reverse complement	aatgtacagtattgcgttttgTCTGAGCCAATAGTCCAGTACAGATGACAATCC
>chr6:117643313-117643363 reverse complement	aatgtacagtattgcgttttgATTGAATTTCATGTGAATAAAACTCTAACTTCTTGAACATAGAAGCTACT
>chr6:117643340-117643390	aatgtacagtattgcgttttgGAAATTCAATTTGAAGGGCTACACAGTGAAAGATCAC
>chr6:117643415-117643465 reverse complement	aatgtacagtattgcgttttgCGAGGTGGTGAGAACATTTCAAGTTGGA
>chr6:117643535-117643585	aatgtacagtattgcgttttgCCTCTCTGTGAGTGCTTATCCTATGCACTGTCTTAC
>chr6:117643627-117643677	aatgtacagtattgcgttttgCCACACAGGCATGTCTTAACTCCCTCATTAAAT
>chr6:117643623-117643673 reverse complement	aatgtacagtattgcgttttgGCCTGTGTGGGAAGACAAATAGCAAGATATG
>chr6:117643913-117643963 reverse complement	aatgtacagtattgcgttttgGCAGTGCATGTGGTTTTACTAGTAAGAGCGTGTG
>chr6:117644021-117644071 reverse complement	aatgtacagtattgcgttttgCCAAGACATCCTATGGATATCTGACTTTCTCAGGACTT
>chr6:117644116-117644166 reverse complement	aatgtacagtattgcgttttgCCTTTGCTTGCTTTCTTGTACATCACCATTGTT
>chr6:117644227-117644277 reverse complement	aatgtacagtattgcgttttgCCCTCCAGACTTACGCATATGACTGCAGTATATTT
>chr6:117644256-117644306	aatgtacagtattgcgttttgGGCACAAACATCAGCTGTGCAATACTTAGACTC
>chr6:117644337-117644387 reverse complement	aatgtacagtattgcgttttgCAGCAAACTTGGAAAGCTAAAAATTTCAGCTGA
>chr6:117644478-117644528	aatgtacagtattgcgttttgCAGTAAAGTGTTGGCTGTCTTTATCCTGAGAGTTCAAC
>chr6:117644538-117644588 reverse complement	aatgtacagtattgcgttttgGGGTAAGTTACAAGGGAAAGTTGATGGAAAAGTCA
>chr6:117644574-117644624	aatgtacagtattgcgttttgCAAAACAAGCCACAACTTTCTTAATATGTGTTTGAGTGTT
>chr6:117644825-117644875 reverse complement	aatgtacagtattgcgttttgATGCAGTAAAAACCTCAAAGGGAATGGTCAAGTAC
>chr6:117644909-117644959 reverse complement	aatgtacagtattgcgttttgCATTTATAAATGACCAGGAAGGCAGAAAAATTAGGTTAGGTA
>chr6:117644933-117644983	aatgtacagtattgcgttttgGAGCAGGGATTTGAAATAAGCAGATCAGATGTC
>chr6:117645013-117645063 reverse complement	aatgtacagtattgcgttttgGAGTAGGGCCTCTTTTGCAGAAGATCCATT
>chr6:117645041-117645091	aatgtacagtattgcgttttgCTGTCTGCTTAGAAACCAAAACTATCCCAATCAAA
>chr6:117645307-117645357 reverse complement	aatgtacagtattgcgttttgGCTATACATTTATAACCACTAACTTCTGTGAATTGTTTTCTGATCTAGG
>chr6:117645388-117645438 reverse complement	aatgtacagtattgcgttttgCTTAGTTGATCTAATCTCCCACATGAAATATTGCCTGA
>chr6:117645421-117645471	aatgtacagtattgcgttttgGAGATAAATCAATCAGGTATGATTAAGTAAACAGTTTGTTGCCTAT
>chr6:117645487-117645537 reverse complement	aatgtacagtattgcgttttgCTGGTTGTTACAATCCCACTGACCTTTGGTAAGTA
>chr6:117645608-117645658	aatgtacagtattgcgttttgGCCCCAGCTCTACCTAAGCACACAGAGTA
>chr6:117645668-117645718 reverse complement	aatgtacagtattgcgttttgGTCTAACAACTGGCTTGCAAAAATCCAGTAGTAGCTAGC
>chr6:117645708-117645758	aatgtacagtattgcgttttgCATCCATACAGACAGTGTTTATGCCATGTGAAAT
>chr6:117645863-117645913	aatgtacagtattgcgttttgGGACCAAGAAATCTCAGTCTTTGGATACTAAATAGTTG
>chr6:117645989-117646039 reverse complement	aatgtacagtattgcgttttgCGTGTTTGTTTCCTCTACACAACTGAAACTACCTAAGAGA

>chr6:117646073-117646123 reverse complement	aatgtacagtattgcgttttgGGTAAAACTTATTCATAGCTAATAGGGGGTACAAAATCAGGCTATA
>chr6:117646180-117646230 reverse complement	aatgtacagtattgcgttttgGCAGGCTACATGTGATCTCTGTAGCATGTTCTTC
>chr6:117646267-117646317 reverse complement	aatgtacagtattgcgttttgTTTATATTATAGTTAGTTATCTAGTTAGTTGTGTACAGAAGTTTGCTACC
>chr6:117646361-117646411 reverse complement	aatgtacagtattgcgttttgCATAATGCTTACCTGATGCTCCTTAGTCAAATGAATTATTG
>chr6:117646458-117646508 reverse complement	aatgtacagtattgcgttttgTACCCTGTGCCCCTTAGCTGTGATTTCCTATT
>chr6:117646499-117646549	aatgtacagtattgcgttttgGCAAACACAGGGCCAAAGACTAAGTGACAT
>chr6:117646559-117646609 reverse complement	aatgtacagtattgcgttttgCAATATTTATGTCTGTGCTGTAGCCATATCAGACCAGTA
>chr6:117646593-117646643	aatgtacagtattgcgttttgGATAAACCACAATTGGAGGATAAATTACAGAAGACCTCAC
>chr6:117646788-117646838	aatgtacagtattgcgttttgGAAAGCCAAGGTGGAAAGAGGAAGATGAGAA
>chr6:117646858-117646908 reverse complement	aatgtacagtattgcgttttgCCTCTACTCAGTAATCAGGTCCTGTTGATTTTACCAT
>chr6:117647010-117647060	aatgtacagtattgcgttttgTGTTATGTATAATGTAGGAGTGGTCATAAGGCTGGTATAATG
>chr6:117647067-117647117 reverse complement	aatgtacagtattgcgttttgGGCTGTTCACAATCTGTTCTATCCAGCCATATAATT
>chr6:117647099-117647149	aatgtacagtattgcgttttgCCTAATTAGGAGCATGGAATTGATGTGGTAGGAATTAAG
>chr6:117647141-117647191	aatgtacagtattgcgttttgGCTGATTATTATTGGGGAGAAAAATGACATGATGAT
>chr6:117647330-117647380	aatgtacagtattgcgttttgCTGAACCTTTAGGTAATAAGCTAGTGTGTAGACAGACATGGTA
>chr6:117647417-117647467	aatgtacagtattgcgttttgCCAAACCCTAGATTATTTGCAGCTACTACTCTGAACTGAA
>chr6:117647403-117647453 reverse complement	aatgtacagtattgcgttttgAATAATCTAGGGTTTGGTGAATATAGTGGAATCAGTGAGAA
>chr6:117647684-117647734 reverse complement	aatgtacagtattgcgttttgGATTATGGGAGAGGAATGAGGTCAAGGGTAAATT
>chr6:117647714-117647764	aatgtacagtattgcgttttgCAGATGCTAACAGCATGGTTTTATAGTGATCCTAGTGA
>chr6:117647808-117647858	aatgtacagtattgcgttttgGAAGAGAGAGCTGAGTACCTTTTTCCTATACTAACAGCATG
>chr6:117647795-117647845 reverse complement	aatgtacagtattgcgttttgCAGCTCTCTCTTCACCACTGATTACTCACAGTTTT
>chr6:117648079-117648129 reverse complement	aatgtacagtattgcgttttgGGATTGACTTTGAGGTGGAATGTAGGTCAATAC
>chr6:117648120-117648170	aatgtacagtattgcgttttgGAGCACAAGGGCTGGGTTCAAAAGACATAAC
>chr6:117648178-117648228 reverse complement	aatgtacagtattgcgttttgCCTTGAGCTTGCCTTTGATGATTCTGTAAAAAG
>chr6:117648332-117648382	aatgtacagtattgcgttttgGACTCATCCTTCTGAAGACCAACAGAGAGTGAA
>chr6:117648412-117648462 reverse complement	aatgtacagtattgcgttttgTGCCCTGCTGCTCCCGATGTTCTTA
>chr6:117648612-117648662 reverse complement	aatgtacagtattgcgttttgCAATGGCCATTGTGTAGAGCCACTTCA
>chr6:117648648-117648698	aatgtacagtattgcgttttgCTGATGTTGGTTGAGCACCTTCTATTTACAACATACTGTAC
>chr6:117648710-117648760	aatgtacagtattgcgttttgCGTTGTTTTAATTCACTGAGTCCACAACAGCTTAAAGAGATAAGATA
>chr6:117648711-117648761 reverse complement	aatgtacagtattgcgttttgAGCTGTTGTGGACTCAGTGAATTAAAACAACGTT
>chr6:117648987-117649037 reverse complement	aatgtacagtattgcgttttgCTCTGTGTCCAAATTTCCCTCTTCTTATGCAAG
>chr6:117649016-117649066	aatgtacagtattgcgttttgGAGGTGATGATGACAGTGATGCACCTACG
>chr6:117649091-117649141 reverse complement	aatgtacagtattgcgttttgCCATCTTTGCCTCTTAGCTTCTGGCGTTTAC
>chr6:117649119-117649169	aatgtacagtattgcgttttgCAGAACAAGCAGTCCCGCTGATACGTTGG
>chr6:117649334-117649384	aatgtacagtattgcgttttgGTGTCCTGACTGAAGGACTTATTTTGAGACACTGAAGTTA
>chr6:117649401-117649451 reverse complement	aatgtacagtattgcgttttgCCAGGTCTCCACTGTCAGGACATAGACTATAGGA
>chr6:117649503-117649553	aatgtacagtattgcgttttgCAAAAACCTGATGTGTGACCTGATTGATGAAC
>chr6:117649595-117649645	aatgtacagtattgcgttttgATGAAAAGCAGCAACAGAAACTTCAAGGAGGTT
>chr6:117649582-117649632 reverse complement	aatgtacagtattgcgttttgTCATGCAGAGTGAAATACATTCTAACCTGCAGAA
>chr6:117649802-117649852	aatgtacagtattgcgttttgCCTGTGAGAGGAGTGTCTTTGGAATTTGGAATT
>chr6:117649858-117649908 reverse complement	aatgtacagtattgcgttttgGCTTAGAAGGACTGATTTCATGCCAACTTGA
>chr6:117649917-117649967	aatgtacagtattgcgttttgTCAGTGGCTGGTTTTCATCGACGGTGTGT
>chr6:117650023-117650073	aatgtacagtattgcgttttgCTCTACACTTCAGACCATTTCAGAACCTGACCTAATTTTACA
>chr6:117650205-117650255	aatgtacagtattgcgttttgGTTGCATGAGCCTATGATCAAACAATTTCCAT
>chr6:117650301-117650351	aatgtacagtattgcgttttgGACCAACGTCTAGGAATTGAGCAAAGAACGT
>chr6:117650296-117650346 reverse complement	aatgtacagtattgcgttttgCCTAGACGTTGGTCTTGGCATGCTAATTATAAACAGAT
>chr6:117650450-117650500	aatgtacagtattgcgttttgGTCATTTACAAGTACTTTGCAAACACACATACCTTATCTCA
>chr6:117650583-117650633 reverse complement	aatgtacagtattgcgttttgCTAAAGCTGGAGTCCCAAATAAACCAGGCATT
>chr6:117650700-117650750 reverse complement	aatgtacagtattgcgttttgGCCCCGCCTCTGAATATTTCTTTAATGTTGT
>chr6:117650728-117650778	aatgtacagtattgcgttttgCTAGAAGCAACTCCGTTCAGAGGTTCCTATTGAAA
>chr6:117650810-117650860 reverse complement	aatgtacagtattgcgttttgCTTCTAGTCAATTTTGATGAGCGAATCTAGATACCTTGC
>chr6:117650984-117651034 reverse complement	aatgtacagtattgcgttttgGATCTTCCAATCCATGAGTTGCACAGTTCTT
>chr6:117651028-117651078	aatgtacagtattgcgttttgCCTAGCATCCCAACCCTAGAATGTACAGATTCTTATG
>chr6:117651098-117651148 reverse complement	aatgtacagtattgcgttttgCCTCCCCCAGGGAGTTCAGTAAGCTTAG
>chr6:117651247-117651297 reverse complement	aatgtacagtattgcgttttgCCTGGAGCTATAGAAATGGATGCTGCATTTCT
>chr6:117651325-117651375 reverse complement	aatgtacagtattgcgttttgGTTGGGTTGTGGCGACAAGTCTTGG

>chr6:117651353-117651403	aatgtacagtattgcgttttgCAGAGTGCCTCTTCAGGTCTAACCCTCAC
>chr6:117651436-117651486 reverse complement	aatgtacagtattgcgttttgCTGAGCCTCTGGCTGCAGTTACTGAAGATTC
>chr6:117651625-117651675 reverse complement	aatgtacagtattgcgttttgGACCCATTTAACGAAGCACTTTGTCCCCT
>chr6:117651700-117651750 reverse complement	aatgtacagtattgcgttttgAGTAGGATCGCTCCTGTATAGGGTGTCCTGAC
>chr6:117651815-117651865 reverse complement	aatgtacagtattgcgttttgGATTTGCTCCCACGCCTGGATATGTAAC
>chr6:117651899-117651949 reverse complement	aatgtacagtattgcgttttgGTTGTCACTTTCTGCTTATTTGTTTTTCTTTCAATAGTCACTTC
>chr6:117652017-117652067 reverse complement	aatgtacagtattgcgttttgGTTCTTTCTCACCTTCGTGAGTTTGTCTAGTTGTG
>chr6:117652122-117652172 reverse complement	aatgtacagtattgcgttttgCTCCGCCCAGTTCTGCACCCTTAAT
>chr6:117652235-117652285 reverse complement	aatgtacagtattgcgttttgATTCCTCCAACTTTTTATCAAGGTTCTTAGCTTCTTCG
>chr6:117652301-117652351 reverse complement	aatgtacagtattgcgttttgGCCCCATCAGGCCATTTGTGTTC
>chr6:117652364-117652414 reverse complement	aatgtacagtattgcgttttgCCACTTGGTCGATTCAGCTGTTGACA
>chr6:117652416-117652466 reverse complement	aatgtacagtattgcgttttgCTTTTTTCTCTCTTCTTGTCAGCATGTCTTATTTCAATAAGGTA
>chr6:117652520-117652570 reverse complement	aatgtacagtattgcgttttgCCCATCTCCTCCTGGTACTCCAAGCTATCAT
>chr6:117652555-117652605 reverse complement	aatgtacagtattgcgttttgGAAGGGTGTTTTCCCACGTGTTTCCAT
>chr6:117652640-117652690 reverse complement	aatgtacagtattgcgttttgCGTGGAGTATCTTAATGGTGTTCTCTGTGTTTGC
>chr6:117653135-117653185 reverse complement	aatgtacagtattgcgttttgAGCCCCTTTACATTTAAGGTAGGTATTGTTATGTGAGTTTGA
>chr6:117653253-117653303 reverse complement	aatgtacagtattgcgttttgTGAGCCTGTGTGTGTCTTTGCATGTAAGATAGAT
>chr6:117653539-117653589 reverse complement	aatgtacagtattgcgttttgTGAAGTTTCCCACTACTATTGCGTGGCAG
>chr6:117653710-117653760 reverse complement	aatgtacagtattgcgttttgAATTATGTGGTAGATTTTAGAATAAGTACCATGTGGCACTCAG
>chr6:117654042-117654092 reverse complement	aatgtacagtattgcgttttgGTTAGGGTGACGATTTGAGATCTTTCTGGCTTTCTAAT
>chr6:117654147-117654197	aatgtacagtattgcgttttgGAAGGAGACAAAGACACAAAAAGCCCTCCA
>chr6:117654203-117654253 reverse complement	aatgtacagtattgcgttttgGTTATTTATTTTATTTTTTAAAAAAAACGGCTCCTGGATGTGTTGAT
>chr6:117654257-117654307 reverse complement	aatgtacagtattgcgttttgGTCTATTTGATTCTTCTCTGTCTCCTTCTTTATTTCTCTAGCTAGTGGT
>chr6:117654434-117654484 reverse complement	aatgtacagtattgcgttttgGATTCAACTTCTTCCTGGTTTAGTCTTGGTGTATGC
>chr6:117654476-117654526 reverse complement	aatgtacagtattgcgttttgCCTCCATTCCAGAGCTTGCTACTGGTCTA
>chr6:117654496-117654546 reverse complement	aatgtacagtattgcgttttgGCTTTTTTTGGTTGGTAGTCTATTAATTACTGCCTCCATTC
>chr6:117655532-117655582 reverse complement	aatgtacagtattgcgttttgGATACTCATTATTCCTGTCCATGGGGATGGA
>chr6:117655610-117655660 reverse complement	aatgtacagtattgcgttttgAAATAACTTTTTCTAATTCTATGAAGAATGTCAACGGTAGTTAGATGAGA
>chr6:117655668-117655718 reverse complement	aatgtacagtattgcgttttgGTCTTGGCTACATGGGGTCTTCTTTGATTCTATGT
>chr6:117656359-117656409 reverse complement	aatgtacagtattgcgttttgATACGTTTATTGGCTGCATAAATCCCTTTTTGAGAA
>chr6:117656424-117656474 reverse complement	aatgtacagtattgcgttttgCGTGAGATGGTATCTCATTGTAGTTTTGATTTGTACTCCT
>chr6:117656464-117656514 reverse complement	aatgtacagtattgcgttttgGTTTCCTGACTTTTTAATAGTTGCCATTCTGACTCG
>chr6:117656739-117656789 reverse complement	aatgtacagtattgcgttttgGTCATTGATGGATTGGCTTGGTTCCAT
>chr6:117656868-117656918	aatgtacagtattgcgttttgCCTAACACAGGAACAGAGAACAAAACAGCACG
>chr6:117656997-117657047 reverse complement	aatgtacagtattgcgttttgGTTCCCTCCCCTCAGCTCCTATCCCTC
>chr6:117657186-117657236	aatgtacagtattgcgttttgGTTTTTATTAGTGATGGATTTGATGAATTGGTGATAAGATTAACA
>chr6:117657179-117657229 reverse complement	aatgtacagtattgcgttttgAATTCATCAAATCCATCACTAATAAAAACTTCAAATGCTCA
>chr6:117657285-117657335 reverse complement	aatgtacagtattgcgttttgGAAAAAAAGAATGGAGGGAAAATGCTCCAAAACTTATGATAG
>chr6:117657407-117657457 reverse complement	aatgtacagtattgcgttttgGGGCTTAATACCTAGGTGATGGTTTGCTAGGTGTAGTAA
>chr6:117657712-117657762 reverse complement	aatgtacagtattgcgttttgCCGTCAACATTAGACTGGTTGAAGCAAATG
>chr6:117657758-117657808 reverse complement	aatgtacagtattgcgttttgCACGTATGTTCACTGTAGCACTATTCACAATAATGAATCATTTG
>chr6:117657853-117657903	aatgtacagtattgcgttttgGTAATGTAATTTCTAGGTCAAATGGTATTTTTGGTTCTAGATCTTTGTTC
>chr6:117657840-117657890 reverse complement	aatgtacagtattgcgttttgGAACCAAAAATACCATTTGACCTAGAAATTACATTACTAGTTACATACCC
>chr6:117657918-117657968 reverse complement	aatgtacagtattgcgttttgTAAAAATTAACTAGTGTTCTGAAAGATATACCTTAACTCTGAGATGAGCT
>chr6:117658048-117658098	aatgtacagtattgcgttttgTTCATATAGTTATCACCATAAAATTGTCATAGCTAGACATGAAATAAGGA
>chr6:117658117-117658167	aatgtacagtattgcgttttgCCAGCCAGACCCAGCCAGTATTATTTCATTACT
>chr6:117658112-117658162 reverse complement	aatgtacagtattgcgttttgGGCTGGATCCTGAACTGGGCAAAATTAC
>chr6:117658158-117658208 reverse complement	aatgtacagtattgcgttttgCAGAAAATATAGGAGACCTTGCAGGCAGTAATGAA
>chr6:117658340-117658390	aatgtacagtattgcgttttgGGAAGTGAGGTGCTATTTTCTCCCGTCTTATAAACC
>chr6:117658326-117658376 reverse complement	aatgtacagtattgcgttttgCCTCACTTCCAGAAAGCTTTAAGACAAAAGGTGAGTAC
>chr6:117658433-117658483	aatgtacagtattgcgttttgGGACCTTGGCTGCATGAAGTTTTAACATGG
>chr6:117658433-117658483 reverse complement	aatgtacagtattgcgttttgGCAGCCAAGGTCCTGCTTATGTCTGTAATATCA

*Primer names are labeled according to their coordination in the GRCh37 human genome assembly; names labeled with reverse complement indicate that they are on minus-strand. #Sequences in small letter are the appended adaptor sequences used in QIAseq target sequence library construction.