

RNA sequencing of sarcomas with simple karyotypes: identification and enrichment of fusion transcripts

Jakob Hofvander¹, Johnbosco Tayebwa¹, Jenny Nilsson¹, Linda Magnusson¹, Otte Brosjö², Olle Larsson³, Fredrik Vult von Steyern⁴, Henryk A Domanski⁵, Nils Mandahl¹ and Fredrik Mertens¹

Gene fusions are neoplasia-associated mutations arising from structural chromosomal rearrangements. They have a strong impact on tumor development and constitute important diagnostic markers. Malignant soft tissue tumors (sarcomas) constitute a heterogeneous group of neoplasms with >50 distinct subtypes, each of which is rare. In addition, there is considerable morphologic overlap between sarcomas and benign lesions. Several subtypes display distinct gene fusions, serving as excellent biomarkers. The development of methods for deep sequencing of the complete transcriptome (RNA-Seq) has substantially improved the possibilities for detecting gene fusions. With the aim of identifying new gene fusions of biological and clinical relevance, eight sarcomas with simple karyotypes, ie, only one or a few structural rearrangements, were subjected to massively parallel paired-end sequencing of mRNA. Three different algorithms were used to identify fusion transcripts from RNA-Seq data. Three novel (*KIAA2026-NUDT11*, *CCBL1-ARL1*, and *AFF3-PHF1*) and two previously known fusions (*FUS-CREB3L2* and *HAS2-PLAG1*) were found and could be verified by other methods. These findings show that RNA-Seq is a powerful tool for detecting gene fusions in sarcomas but also suggest that it is advisable to use more than one algorithm to analyze the output data as only two of the confirmed fusions were reported by more than one of the gene fusion detection software programs. For all of the confirmed gene fusions, at least one of the genes mapped to a chromosome band implicated by the karyotype, suggesting that sarcomas with simple karyotypes constitute an excellent resource for identifying novel gene fusions.

Laboratory Investigation (2015) 95, 603–609; doi:10.1038/labinvest.2015.50; published online 13 April 2015

Sarcomas are malignant tumors that arise in bone or soft tissues. They are classified according to their degree of resemblance to normal mesenchymal cells, and diagnosis of the >50 histological subtypes is challenging owing to morphological overlap and the rarity of the tumors.¹ In some cases, however, genetic features, in particular gene fusions, are helpful in separating differential diagnostic entities. Gene fusions are cancer-associated mutations that have attracted much attention because of their pathogenic and diagnostic importance.² They occur in all types of neoplasia and arise from chromosomal rearrangements in the form of translocations, insertions, inversions, or interstitial deletions. Chromosomal rearrangements giving rise to gene fusions can often be seen as the only structural rearrangement at chromosome banding analysis, and deep sequencing of fusion-positive leukemias and sarcomas has shown that they often are accompanied by only a small number of mutations;^{3,4} hence, gene

fusions are generally supposed to have a strong impact on tumor development. Experimental animal models and *in vitro* studies have strengthened this theory as they have shown that a gene fusion can sometimes be sufficient for malignant transformation.^{5,6}

Gene fusions were previously identified through a time-consuming multi-step procedure, starting with the identification of recurrently involved chromosome bands in metaphase spreads. Breakpoint regions could then be narrowed down with fluorescence *in situ* hybridization (FISH) and potential chimeric transcripts directly tested with reverse-transcriptase PCR (RT-PCR). Recently, the development of methods for deep sequencing of the transcriptome (RNA-Seq) has not only bypassed the need for cell culturing and subsequent analyses of metaphase chromosomes but has also made it possible to detect fusions arising through cytogenetically cryptic rearrangements.⁷ RNA-Seq is based on the sequencing

¹Department of Clinical Genetics, University and Regional Laboratories, Lund University, Lund, Sweden; ²Department of Orthopedics, Karolinska University Hospital, Solna, Sweden; ³Department of Pathology, Karolinska University Hospital, Solna, Sweden; ⁴Department of Orthopedics, Skåne University Hospital, Lund University, Lund, Sweden and ⁵Department of Pathology, University and Regional Laboratories, Lund University, Lund, Sweden

Correspondence: J Hofvander, MSc, Department of Clinical Genetics, University and Regional Laboratories, Lund University, SE-221 85 Lund, Sweden.
E-mail: jakob.hofvander@med.lu.se

Received 7 November 2014; revised 24 February 2015; accepted 26 February 2015

of the complete set of RNA transcripts in a tissue or cell sample to give a greater understanding of the gene expression profile, allowing for improved mapping of transcription start sites, as well as identification of alternative splicing events and gene fusions.⁸

By taking advantage of prior cytogenetic information, nine sarcoma samples from eight patients were subjected to massively parallel paired-end sequencing of RNA to identify new gene fusions. Samples were selected on the basis of their simple karyotypes, harboring only one or a few structural chromosome aberrations, none of which corresponded to any known gene fusion. RNA-Seq data from these tumors were then analyzed using three state-of-the-art gene fusion-detecting algorithms: TopHat,⁹ SOAPfuse,¹⁰ and ChimeraScan.¹¹ Potential chimeric transcripts were correlated with the karyotypes and verified with RT-PCR.

MATERIALS AND METHODS

Tumor Samples and Chromosome Banding Analysis

The study was based on cytogenetic findings in nine tumor samples from eight sarcoma patients (Table 1). As part of the diagnostic routines, all tumors had been sent to the Department of Clinical Genetics in Lund for cytogenetic analysis. Portions of the samples that had been stored at -80°C for 2–20 years were used for RNA extraction. All samples were obtained after written consent and all studies were approved by the institutional ethical committees.

Cell culturing, harvesting, and G-banding were performed according to established methods.¹² Karyotypes were written according to the recommendations of the International System for Human Cytogenetic Nomenclature 2013.¹³

RNA-Seq and Bioinformatical Analysis

Total RNA was extracted from frozen tumor samples using the RNeasy Lipid tissue kit (Qiagen, CA, USA) and mRNA libraries were prepared as described¹⁴ using the TruSeq RNA sample preparation kit v 2 (Illumina, CA, USA). Briefly, poly-A-tailed RNA was enriched using oligo-dT beads. RNA

was fragmented to a median size of 200 nucleotides and cDNA was synthesized from these fragments using Super-script II reverse-transcriptase (Invitrogen, CA, USA). Double-stranded cDNA was produced using DNA polymerase I and RNase H. Oligonucleotide adaptors were ligated to the double-stranded cDNA and the adaptor-bound fragments were enriched using a 15-cycle PCR. Paired-end 101 base pair (bp) reads were generated from the mRNA libraries using the HiScanSQ System (Illumina).

Identification of potential fusion transcripts was performed on fastq files using TopHat version 2.0.7 (<http://tophat-fusion.sourceforge.net>), SOAPfuse version 1.26 (<http://soap.genomics.org.cn/SOAPfusion.html>), and ChimeraScan version 0.4.5 (<http://code.google.com/p/chimerascan>). Further details regarding settings are given in Supplementary Tables S1–3. The GRCh37/hg19 build was used as the human reference genome.

Potential fusion transcripts obtained from the output files (Supplementary Tables S1–3) were reduced to a list containing only those that were selected for further analysis by RT-PCR (Supplementary Table S4). The filtering was primarily based on the cytogenetic information for the individual cases. Thus, when any of the suggested 5' and 3' genes were located close to any of the breakpoints in the karyotype, it was kept for further investigation. Also, fusion transcripts involving genes that had previously been reported in fusions were considered.¹⁵ Chimeric transcripts that did not fit the above criteria and had no reads spanning the fusion junction (spanning reads) and less than five reads bordering the fusion junction (flanking reads), as well as those that were regarded as read-through transcripts or involved pseudogenes, were discarded.

RT-PCR Analysis

Reverse transcription and PCR amplifications were performed as described.^{16,17} Primers specific for each gene were designed to detect possible fusion transcripts (Supplementary Table S5). Transcripts were amplified using an initial denaturation for 2 min at 94°C , followed by 30 cycles of 30 s at 94°C , 30 s at 58°C , and 3 min at 72°C , and a final extension for 3 min at

Table 1 Cases studied by RNA sequencing

Case	Diagnosis	Karyotype	No. of reads
1a	Osteosarcoma	46,XY,der(5)t(5;15)(p15;q21)	11 146 076
1b	Osteosarcoma, metastasis	46,XY,der(5)t(5;15)/46,idem,der(6)t(6;11)(p24;q12)	12 137 370
2	Myxofibrosarcoma	46,XX,t(2;17;7)(p22;q12;q11),ins(17;11)(p11;p11p13)	11 184 324
3	Myxofibrosarcoma	44–46,X,-X,t(2;6)(q13;p21),add(9)(p22),add(12)(p13),-16,der(21)t(16;21)(p11;p11),+2r	12 252 963
4	Low-grade fibromyxoid sarcoma	46,XX,t(7;13;11)(q32;q34;q23)	43 161 828
5	Fibrosarcoma	46,XY,t(4;22)(p15;q12)	14 341 816
6	Undifferentiated pleomorphic sarcoma	48–52,X,-Y,add(7)(p22)x2,+8,+add(12)(p11)x2,+19,+20/49–50,idem,-12,-12,+add(12)(p13)x2	11 158 929
7	Glomus tumor	46,XX,t(3;7)(q21;q32)	10 469 292
8	Myxoid liposarcoma, reclassified as lipoblastoma	46,XX,der(5)t(5;8)(q35;q13)inv(8)(q22q24),der(8)t(5;8)	16 951 072

72 °C. Amplified fragments were purified from agarose gels and directly sequenced using the Big Dye v1.1 cycle sequencing kit (Applied Biosystems, CA, USA) on an ABI-3130 genetic analyzer (Applied Biosystems). BLASTN software (<http://www.ncbi.nlm.nih.gov/blast>) and ORF-finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) were used for the analysis of sequence data.

Rapid Amplification of cDNA Ends (RACE)

A 5' RACE was used to detect a potential partner to the *NOTCH3* gene in Case 8.

The SMARTer RACE cDNA Amplification Kit (Clontech, CA, USA) was used to generate RACE-Ready cDNA. The buffer mix was prepared by mixing 2.5 × First-Strand Buffer, 5 mM DTT, and 2.5 mM dNTP MIX to a final volume of 4 μl. In a separate tube, 1 μg of RNA, 1 μl 5'-CDS Primer A, and 1.75 μl ddH₂O were mixed. The tube was incubated at 72 °C for 2 min, followed by 42 °C for 2 min, and then cooled on ice. Thereafter, 1 μl of SMARTer IIA oligo, 4 μl of the previously prepared buffer mix, 1 U RNase inhibitor, and 10 U SMARTScribe reverse-transcriptase were added to a final volume of 10 μl. The solution was then incubated in a thermal cycler at 42 °C for 90 min, followed by 10 min at 70 °C.

RACE was performed with the Advantage 2 PCR Kit (Clontech) according to the following protocol: Mix 2.5 μl of the RACE-Ready cDNA, 34.5 μl PCR-Grade water, 1 × Advantage PCR buffer, 0.2 mM dNTP mix, 1 × Advantage 2 polymerase mix, 1 × Universal primer mix, and 0.2 μM primer to a final volume of 50 μl. The PCR-reaction was carried out according to the following protocol: 5 cycles of 30 s at 94 °C and of 5 min at 72 °C, 5 cycles of 30 s at 94 °C, 30 s at 70 °C, and 3 min at 72 °C, and 25 cycles of 30 sec at 94 °C, 30 s at 68 °C, and 3 min at 72 °C.

Results

ChimeraScan reported a total of 1329 potential fusions, whereas SOAPfuse reported 81 and TopHat 26 (Supplementary Tables S1–3). After filtering, six fusions from ChimeraScan, two from SOAPfuse, and six from TopHat were kept for verification. A summary of the results from RT-

PCR analysis and sequencing of these potential fusion transcripts is displayed in Table 2.

Samples 1a (primary tumor) and 1b (metastasis) were two samples from an osteosarcoma with an unbalanced t(5;15). None of the algorithms indicated fusions involving genes located in the breakpoint regions of the chromosomes involved in the translocation. ChimeraScan, but not any of the other two algorithms, reported a *CIQTNF6-HIF3A* transcript in both samples. The transcript lacked spanning reads and was only supported by a few flanking reads. It could not be confirmed by RT-PCR.

The myxofibrosarcoma of Case 2 had two structural rearrangements, involving six breakpoints in four different chromosomes. Only one reported fusion—*KLHL29-PER1* detected by TopHat—showed some correspondence to the cytogenetic data, but could not be detected by RT-PCR.

The myxofibrosarcoma of Case 3 had ring chromosomes involving chromosomes 9 and 12, as well as a translocation involving chromosomes 2 and 6. The reported *AFF3-PHF1* fusion, detected by TopHat, correlated well with the translocation, and was confirmed by RT-PCR. The sequenced fragment revealed that a fusion had occurred between *AFF3* exon 11 and *PHF1* exon 13 (Figure 1), generating an out-of-frame transcript. Several reported fusions involving chromosomes 9 and 12 were analyzed. The *DNM1-GBA2*, *DCN-CUX2*, and *ELK3-RIC8B* fusions were not detected by RT-PCR, whereas *CCBL1-ARL1* (exon 1 with exon 5, out of frame) and *KIAA2026-NUDT11* (exon 1 with exon 2, in frame) could be confirmed (Figure 1).

The tumor of Case 4 was initially diagnosed as a benign fibroblastic-myofibroblastic lesion on the basis of preoperative fine- and core-needle aspirates. Histopathologic analysis of the excised tumor was more compatible with a malignant tumor, but a precise diagnosis could not be reached. Also the cytogenetic results were inconclusive, identifying a balanced t(7;13;11)(q32;q34;q23) at G-banding analysis. An external review of the morphology suggested a low-grade fibromyxoid sarcoma (LGFMS), which was in agreement with the RNA-Seq data; ChimeraScan detected a *FUS-CREB3L2* fusion,

Table 2 Verified gene fusions

Case	Fusion	Algorithm ^a	ORF-finder	BLAST
3	<i>AFF3-PHF1</i>	T	Out of frame	<i>AFF3</i> exon 11 fused with <i>PHF1</i> exon 13
3	<i>CCBL1-ARL</i>	T, C, S	Out of frame	<i>CCBL1</i> exon 1 fused with <i>ARL1</i> exon 5
3	<i>KIAA2026-NUDT11</i>	T, C, S	In frame	<i>KIAA2026</i> exon 1 fused with <i>NUDT11</i> exon 2
4	<i>FUS-CREB3L2</i>	C	In frame	<i>FUS</i> exon 6 fused with <i>CREB3L2</i> exon 6
7	<i>NOTCH3-?</i>	C		<i>NOTCH3</i> exon 27 fused with inverted <i>NOTCH3</i> exon 33 and 7 unidentified bp
8	<i>HAS2-PLAG1</i>	C	In frame	<i>HAS2</i> exon 1 fused with <i>PLAG1</i> exon 3

^aT = TopHat; C = ChimeraScan; S = SOAPfuse.

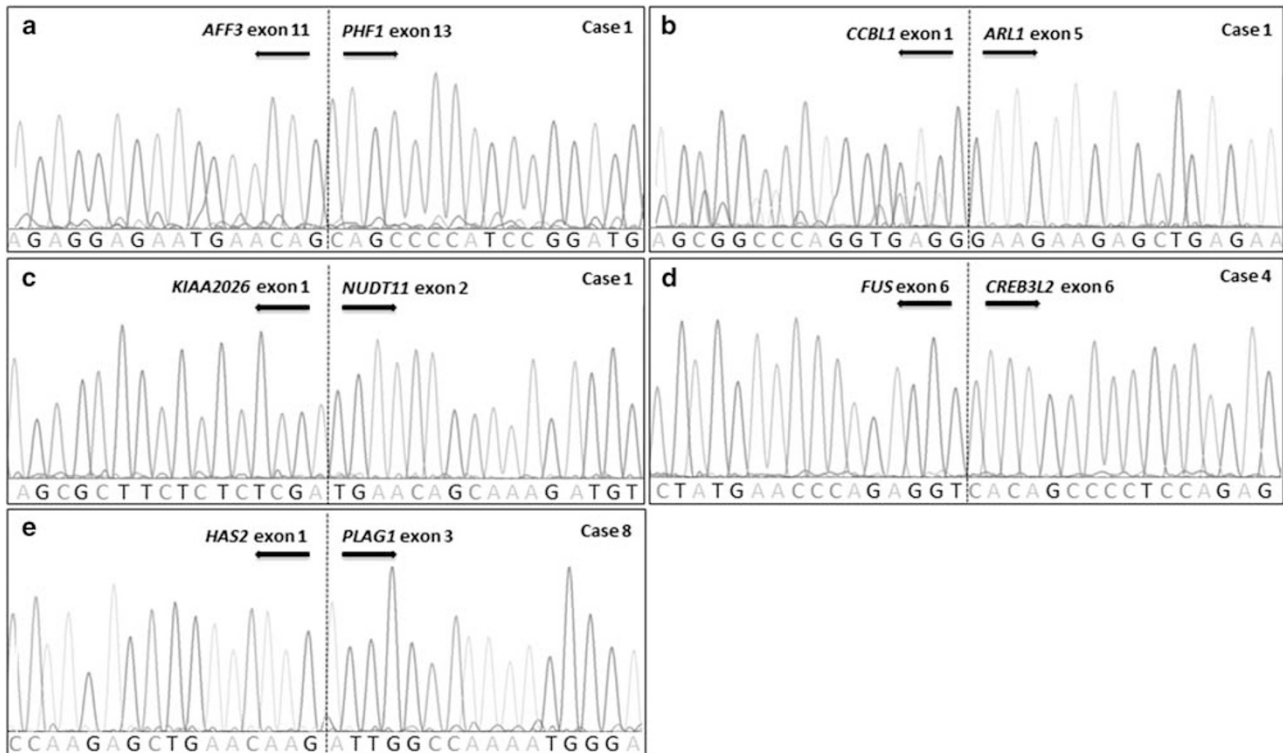


Figure 1 (a–e): Partial chromatograms of RT-PCR-amplified fusion transcripts from chimeric genes identified by RNA-Seq.

which is characteristic for LGFMS, that could be confirmed with RT-PCR. Sequencing of the PCR-amplified product identified an in-frame fusion between *FUS* exon 6 and *CREB3L2* exon 6 (Figure 1).

In Cases 5 and 6 (a fibroblastic sarcoma and an undifferentiated pleomorphic sarcoma, respectively) no fusion transcripts of potential pathogenetic importance or corresponding to the breakpoint regions of the chromosomes involved in the structural rearrangements could be detected by any of the three algorithms.

The glomus tumor of Case 7 showed a balanced $t(3;7)(q21;q32)$ at G-banding analysis. ChimeraScan identified a *NOTCH3-AGBL3* fusion that could not be confirmed by RT-PCR. A 5'RACE PCR starting from exon 29 of *NOTCH3* generated a fragment containing *NOTCH3* exons 29–27 preceded by an inverted part of *NOTCH3* exon 33 and ending with 7 bp that could not be mapped to the reference genome (ACATGGG).

The tumor of Case 8, an intramuscular tumor from a 26-year-old woman, was initially diagnosed as myxoid liposarcoma (MLS). However, G-banding analysis of the excised tumor identified a complex exchange of material between chromosomes 5 and 8 as the sole anomaly (Table 1). FISH for the genes (*FUS* and *DDIT3*) involved in the $t(12;16)(q13;p11)$ that is pathognomonic for MLS was negative, ruling out a cryptic or variant *FUS-DDIT3* fusion. Despite the cytogenetic results, which were clearly at odds with the diagnosis but not specific for any other entity, the diagnosis was kept.

ChimeraScan and subsequent RT-PCR identified an in-frame fusion between *HAS2* exon 1 and *PLAG1* exon 3 (Figure 1). The two genes are located in bands 8q24 and 8q12, respectively. Thus, the rearrangement of chromosome arm 8q, including a deletion of the sequence between the two genes, must have been more complex than suspected from the karyotype. Prompted by these results, the morphology was re-reviewed and the diagnosis was changed to lipoblastoma.

DISCUSSION

The aim of the study was to evaluate the potential advantages of combining cytogenetic and RNA-Seq data when searching for new gene fusions. It is already well known that recurrent balanced structural rearrangements often result in gene fusions and that RNA-Seq is an excellent method for unguided detection of fusion transcripts.^{2,7} However, there has been no systematic analysis of the extent to which a seemingly unique structural aberration results in a functional fusion transcript. Furthermore, RNA-Seq typically results in a large number of potential fusion events that could be difficult to evaluate without additional information. To date, 85 gene fusions have been reported in sarcomas.¹⁵ They are found in all major lineages (apart from nerve sheath tumors), covering >30 distinct entities. By focusing on sarcomas with unique structural rearrangements accompanied by few or no additional chromosomal aberrations we were able to detect five different fusions, three of them being novel.

RNA-Seq has dramatically increased the pace at which gene fusions are detected.⁷ However, RNA-Seq also has several drawbacks that must be kept in mind. For instance, potential errors can arise when converting RNA into cDNA, including the generation of cDNA artifacts due to template switching, leading to false-positive gene fusions. Reverse-transcriptase can also synthesize cDNA in a primer-independent manner, generating random cDNA.⁸ Furthermore, some gene fusions are difficult to detect with RNA-Seq; chromosomal rearrangements leading to exchange of only the regulatory sequences, ie the fusion breakpoints being located outside of the mature mRNA molecules, do not generate proper fusion transcripts and are therefore not identified.^{8,18} Another potential pitfall with RNA-Seq is the presence of so-called read-through transcripts—ie, two neighboring genes that are located on the same strand and are transcribed in the same direction form a single mRNA. Such events are more likely to occur when the genes are located close to each other and when the 5' gene is highly expressed.¹⁹ An example of this phenomenon in the present study was the finding of a *CTBS-GNG5* transcript in four cases; these are located 46 kb from each other on chromosome 1 and are transcribed in the same direction.

The technique is rapidly improving and the main challenge today is not the sequencing as such but the analysis of the huge amounts of data generated. There have, however, been major accomplishments also in the field of bioinformatics, and a large number of different softwares for the detection of fusion transcripts have been developed. The different strategies employed to detect such transcripts have been outlined in recent reviews.^{20,21} As there are several parameters that could influence the specificity and sensitivity of the software, such as mapping tools, cut-offs for distance between genes located within the same chromosome, requirements with regard to the number of supporting reads, etc, it is not surprising that the output varies greatly among the software programs.²¹ Indeed, part of the different outcomes of the softwares used in the present study—identifying as few as 26 putative fusion transcripts with TopHat and as many as 1329 with ChimeraScan—could be explained by different settings (Supplementary Tables S1–3). For instance, with TopHat, the requirements for fusion transcripts were that the two genes were separated by at least 100 000 bp, that the fusion anchor length was ≥ 13 bp, and that there was at least one fusion-spanning read and at least two fusion-spanning mate pairs; the corresponding values for SOAPfuse were 1,000 bp, 10 bp, and ≥ 2 supporting reads, at least one of which was a spanning read, respectively. In contrast to the other two algorithms, ChimeraScan did not require spanning reads,¹¹ explaining why a much larger number of putative fusion transcripts were identified with this tool.

From the results of the present study it is clear that use of any one of the three programs is not sufficient for detecting all potential fusion genes. ChimeraScan detected in total four fusions that were confirmed by PCR (*KIAA2026-NUDT11*, *CCBL1-ARL1*, *FUS-CREB3L2*, and *HAS2-PLAG1*); in

addition, a fusion involving *NOTCH3* was indicated. TopHat detected three fusions (*KIAA2026-NUDT11*, *CCBL1-ARL1*, and *AFF3-PHF1*) and SOAPfuse detected two (*KIAA2026-NUDT11* and *CCBL1-ARL1*). Thus, the *CCBL1-ARL1* and *KIAA2026-NUDT11* fusions were the only fusions independently detected by more than one algorithm. Needless to say, it cannot be excluded that there were additional fusion transcripts that remained undetected by all three algorithms. A recent study by Panagopoulos *et al* exemplifies the bioinformatic problems that remain to be solved.²² They analyzed RNA from a small round cell tumor with cytogenetic features strongly indicative of the characteristic *CIC-DUX4* fusion with three different software programs (ChimeraScan, FusionMap, and FusionFinder). However, none of them identified the suspected fusion transcript. Only when the reads specifically aligning to the part of the *CIC* gene where previous fusion breakpoints have been mapped were retrieved (so-called grep command) could a fusion with *DUX4* be detected. Needless to say, such an approach could not be used in the present study, as there were no strong candidate target genes to scrutinize. In contrast to our results and those of Panagopoulos *et al*,²² a recent study evaluating the reliability of RNA-Seq for detecting clinically relevant fusion genes in leukemias detected all previously known fusions, as well as nine novel ones, with a single algorithm (ChimeraScan).²³ Possibly, the higher admixture of normal cells in solid tumors, such as sarcomas, makes it advisable to use more than one fusion algorithm when searching for gene fusions.

The fusion between *HAS2* exon 1 and *PLAG1* exon 3 detected in Case 8 is a good example of the clinical relevance of RNA-Seq. The tumor, obtained from a 26-year-old woman, was originally diagnosed as MLS, in spite of the lack of cytogenetic support for this diagnosis; molecular confirmation is not mandatory. Consequently, after surgery, the patient was checked regularly for local recurrences and lung metastases. The *HAS2-PLAG1* fusion is, however, specific for another adipocytic tumor, lipoblastoma, which is benign and never metastasizes.²⁴ Prior genetic analyses have shown that lipoblastomas almost invariably have gene fusions leading to transcriptional upregulation of the *PLAG1* gene.²⁴ In contrast, MLS never shows *PLAG1* fusions, but instead displays a *FUS-DDIT3* or *EWSRI-DDIT3* fusion in close to 100% of the cases. Lipoblastomas are extremely rare in adults, with 90% occurring in children below 3 years of age and $< 5\%$ after the age of 10 years.²⁵ As lipoblastoma and myxoid liposarcoma are morphologically very similar, a lesion with such features in an adult is easily mistaken for a MLS. The present finding of a *HAS2-PLAG1* fusion illustrates that RNA-Seq can provide differential diagnostic information of vital importance in the management of patients with sarcomas.

Case 3 was a myxofibrosarcoma having ring chromosomes involving chromosomes 9 and 12,²⁶ as well as a translocation involving chromosomes 2 and 6. The *AFF3-PHF1* fusion, correlating well with the t(2;6) translocation, has previously not been reported. The *PHF1* and *AFF3* genes have both been

described in other fusion events before.¹⁵ The present fusion, however, produces an out-of-frame transcript, making it difficult to speculate on its pathogenetic impact. The *AFF3* gene codes for a transcriptional activator that may function in lymphoid development and oncogenesis, although very little is known about its transcriptional targets. The PHF1 protein acts as an accessory component of the polycomb repressive complex 2 (PRC2), which catalyzes trimethylation of histone H3 Lys27 (H3K27me3) to repress gene expression.²⁷ As *PHF1* rearrangements are believed to be important for tumor development in other sarcomas,^{28,29} it is not unlikely that loss of PHF1 function may also lead to epigenetic deregulation of PRC2 target genes.

It is not surprising that fusions involving the chromosomes included in the ring chromosomes were found in Case 3, as rings undergo a series of breakage-fusion-bridge events, causing the DNA molecule to frequently break and rejoin at cell division. This could cause the formation of several gene fusions that lack driver mutation qualities.^{30,31} The *CCBL1-ARL1* fusion resulted in an out-of-frame transcript, suggesting that it was a chance event with little impact on tumor development. The *KIAA2026-NUDT11* fusion transcript was in frame, but speculation on the importance, if any, for tumor development is premature, not least because *KIAA2026* is an uncharacterized gene.

In Case 4, eventually diagnosed as a LGFMS, ChimeraScan reported a fusion between *FUS* and *CREB3L2*, which could be verified by RT-PCR. This fusion has been detected in 76–96% of all LGFMS and is cytogenetically seen as a t(7;16)(q33;p11) or in the form of a ring chromosome.³² The findings in this case serve to illustrate that characteristic gene fusions sometimes are cytogenetically cryptic or masked as more complex rearrangements.²

The RNA-Seq data in Case 7, diagnosed as a glomus tumor, supported a fusion between *AGBL3* and *NOTCH3*. This could, however, not be verified by RT-PCR. Glomus tumors belong to the pericytic subgroup of soft tissue tumors, and fusions involving the micro RNA *MIR143HG* as the 5' gene with any of *NOTCH1*, *NOTCH2*, or *NOTCH3* as the 3' partner gene have previously been reported in these tumors.³³ We thus investigated the fusion further. Visualizing the reads generated from RNA-Seq in the Integrative Genomics Viewer (IGV; <https://www.broadinstitute.org/igv/home>) showed that there were more reads in the 3' part of the *NOTCH3* gene, exons 25–33, indicating that this part was expressed at higher levels than the rest of the gene (Supplementary Figure 1) and in agreement with a rearrangement resulting in a split of the gene. *NOTCH3* was thus further analyzed with 5'-RACE-PCR, revealing that exon 27 of *NOTCH3* was fused with a small inverted part of *NOTCH3* exon 33 followed by 7 bp that could not be aligned to any known genomic location. Previous reports have shown that *MIR143HG* exon 1 fuses with exon 29 of *NOTCH3* leading to an increased expression of the 3' part of *NOTCH3*.³³ These results are in line with our data, although we could not identify the 5' partner. It could be noted that

neither *NOTCH3* nor *MIR143HG* maps to chromosome bands involved in the translocation t(3;7) in this case. The lack of cytogenetic support for the involvement of *NOTCH3* might be explained by the fact that it is located in a chromosome band (19p13) that is difficult to detect when translocated.

No fusions were identified in samples 1a, 1b, 2, 5, and 6. Whether this is because they truly lacked gene fusions or because we were not able to detect them is a moot point. As mentioned above, some chimeric genes might not have been detected because of bioinformatic limitations, and some because of the methodological aspects of RNA-Seq. Still, the present study shows that there is an increased likelihood of finding novel gene fusions by sequencing tumors that show simple structural rearrangements at chromosome banding analysis, and it emphasizes that, in the absence of whole-genome sequencing data, karyotypes are valuable when evaluating the significance of detected fusion transcripts. Finally, there still seem to be bioinformatic limitations when handling RNA-Seq data as none of the algorithms used in this study was able to identify all confirmed fusions. It is therefore advisable to use more than one algorithm to detect chimeric genes in sarcomas and, reasonably, in other solid tumors.

Supplementary Information accompanies the paper on the Laboratory Investigation website (<http://www.laboratoryinvestigation.org>)

ACKNOWLEDGMENTS

This work was supported by Swedish Cancer Society, the Gunnar Nilsson Cancer Foundation, the Medical Faculty of Lund University, and Region Skåne.

DISCLOSURE/CONFLICT OF INTEREST

The authors declare no conflict of interest.

1. Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens F. WHO Classification of Tumours of Soft Tissue and Bone. IARC Press: Lyon, 2013.
2. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 2007;7:233–245.
3. Welch JS, Ley TJ, Link DC et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell* 2012;150:264–278.
4. Joseph CG, Hwang H, Jiao Y et al. Exomic analysis of myxoid liposarcomas, synovial sarcomas and osteosarcomas. *Genes Chromosomes Cancer* 2014;53:15–24.
5. Riggi N, Cironi L, Suvà ML et al. Sarcomas: genetics, signalling, and cellular origins. Part 1: the fellowship of TET. *J Pathol* 2007;213:4–20.
6. Straessler KM, Jones KB, Hu H et al. Modeling clear cell sarcomagenesis in the mouse: cell of origin differentiation state impacts tumor characteristics. *Cancer Cell* 2013;23:215–227.
7. Mertens F, Tayebwa J. Evolving techniques for gene fusion detection in soft tissue tumours. *Histopathology* 2013;64:151–162.
8. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 2011;12:87–98.
9. Kim D, Perteza G, Trapnell C et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;14:R36.
10. Jia W, Qiu K, He M et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Genome Biol* 2013;14:R12.
11. Iyer MK, Chinnaiyan AM, Maher CA. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* 2011;27:2903–2904.
12. Mandahl N, Mertens F, Willén H et al. A new cytogenetic subgroup in lipomas: loss of chromosome 16 material in spindle and pleomorphic lipomas. *J Cancer Res Oncol* 1994;120:707–711.

13. Shaffer LG, McGowan-Jordan J, Schmid M. An International System for Human Cytogenetic Nomenclature. Karger: Basel, 2013.
14. Walther C, Tayebwa J, Lilljebjorn H *et al*. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangi endothelioma. *J Pathol* 2014;232:534–540.
15. Mitelman F, Johansson B, Mertens F eds. Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. Available from: URL: <http://cgap.nci.nih.gov/Chromosomes/Mitelman2015>.
16. Panagopoulos I, Storlazzi CT, Fletcher CD *et al*. The chimeric FUS/CREB3L2 gene is specific for low-grade fibromyxoid sarcoma. *Genes Chromosomes Cancer* 2004;40:218–228.
17. Mohajeri A, Tayebwa J, Collin A *et al*. Comprehensive genetic analysis identifies a pathognomonic *NAB2/STAT6* fusion gene, non-random secondary genomic imbalances, and a characteristic gene expression profile in solitary fibrous tumor. *Genes Chromosomes Cancer* 2013;52:873–886.
18. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol*; 12:R72.
19. Nacu S, Yuan W, Kan Z *et al*. Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. *BMC Med Genomics* 2011;4:11.
20. Wang Q, Xia J, Jia P *et al*. Application of next generation sequencing to human gene fusion detection: computational tools, features and perspectives. *Brief Bioinform* 2012;14:506–519.
21. Carrara M, Beccuit M, Lazzarato F *et al*. State-of-the-art fusion-finder algorithms sensitivity and specificity. *Biomed Res Int* 2013;2013:340620.
22. Panagopoulos I, Gorunova L, Bjerkehagen B *et al*. The 'Grep' command but not FusionMap, FusionFinder or ChimeraScan captures the *CIC-DUX4* fusion gene from whole transcriptome sequencing data on a small round cell tumor with t(4;19)(q35;q13). *PLoS One* 2014;9:e99439.
23. Lilljebjorn H, Agerstam H, Orsmark-Pietras C *et al*. RNA-seq identifies clinically relevant fusion genes in leukemia including a novel MEF2D/CSF1R fusion responsive to imatinib. *Leukemia* 2014;28:977–979.
24. Hibbard MK, Kozakewich HP, Dal Cin P *et al*. PLAG1 fusion oncogenes in lipoblastoma. *Cancer Res* 2000;60:4869–4872.
25. Weiss SW, Goldblum JR. Enzinger and Weiss's Soft Tissue Tumors. Mosby Elsevier: Amsterdam, 2008.
26. Nord KH, Macchia G, Tayebwa J *et al*. Integrative genome and transcriptome analyses reveal two distinct types of ring chromosome in soft tissue sarcomas. *Hum Mol Genet* 2014;23:878–888.
27. Hong Z, Jiang J, Lan L *et al*. A polycomb group protein, PHF1, is involved in the response to DNA double-strand breaks in human cell. *Nucleic Acids Res* 2008;36:2939–2947.
28. Gebre-Medhin S, Nord KH, Möller E *et al*. Recurrent rearrangement of the *PHF1* gene in ossifying fibromyxoid tumors. *Am J Pathol* 2012;181:1069–1077.
29. Antonescu CR, Sung YS, Chen CL *et al*. Novel *ZC3H7B-BCOR*, *MEAF6-PHF1*, and *EPC1-PHF1* fusions in ossifying fibromyxoid tumors—molecular characterization shows genetic overlap with endometrial stromal sarcoma. *Genes Chromosomes Cancer* 2014;53:183–193.
30. Gebhart E. Ring chromosomes in human neoplasias. *Cytogenet Genome Res* 2008;121:149–173.
31. Gisselsson D, Hoglund M, Mertens F *et al*. The structure and dynamics of ring chromosomes in human neoplastic and non-neoplastic cells. *Hum Genet* 1999;104:315–325.
32. Folpe AL, Hornick JL, Mertens F. Low-grade fibromyxoid sarcoma. In: Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens F eds. WHO Classification of Tumours of Soft Tissue and Bone. 4th (ed.) IARC Press: Lyon, 2013, p 95–96.
33. Mosquera JM, Sboner A, Zhang L *et al*. Novel MIR143-NOTCH fusions in benign and malignant glomus tumors. *Genes Chromosomes Cancer* 2013;52:1075–1087.