



A novel next generation sequencing approach to improve sarcoma diagnosis

Lauren McConnell¹ · Oisín Houghton² · Peter Stewart¹ · Jana Gazdova¹ · Shambhavi Srivastava¹ · Chang Kim¹ · Mark Catherwood² · Anna Strobl^{3,4} · Adrienne M. Flanagan^{3,4} · Anca Oniscu⁵ · Leonie I. Kroeze⁶ · Patricia Groenen⁶ · Philippe Tanriere⁷ · Manuel Salto-Tellez^{1,2} · David Gonzalez^{1,2}

Received: 16 October 2019 / Revised: 12 January 2020 / Accepted: 13 January 2020 / Published online: 11 February 2020
© The Author(s), under exclusive licence to United States & Canadian Academy of Pathology 2020

Abstract

Sarcoma is a rare disease affecting both bone and connective tissue and with over 100 pathologic entities, differential diagnosis can be difficult. Complementing immune-histological diagnosis with current ancillary diagnostic techniques, including FISH and RT-PCR, can lead to inconclusive results in a significant number of cases. We describe here the design and validation of a novel sequencing tool to improve sarcoma diagnosis. A NGS DNA capture panel containing probes for 87 fusion genes and 7 genes with frequent copy number changes was designed and optimized. A cohort of 113 DNA samples extracted from soft-tissue and bone sarcoma FFPE material with clinical FISH and/or RT-PCR results positive for either a translocation or gene amplification was used for validation of the NGS method. Sarcoma-specific translocations or gene amplifications were confirmed in 110 out of 113 cases using FISH and/or RT-PCR as gold-standard. *MDM2/CDK4* amplification and a total of 25 distinct fusion genes were identified in this cohort of patients using the NGS approach. Overall, the sensitivity of the NGS panel is 97% with a specificity of 100 and 0% failure rate. Targeted NGS appears to be a feasible and cost-effective approach to improve sarcoma subtype diagnosis with the ability to screen for a wide range of genetic aberrations in one test.

Supplementary information The online version of this article (<https://doi.org/10.1038/s41379-020-0488-1>) contains supplementary material, which is available to authorized users.

✉ David Gonzalez
D.GonzalezdeCastro@qub.ac.uk

- ¹ CCRCB, Queen's University Belfast, Belfast BT9 7AE, UK
- ² Belfast Health & Social Care Trust, Belfast BT9 7AB, UK
- ³ Royal National Orthopedic Hospital Stanmore, Middlesex HA7 4LP, UK
- ⁴ UCL Cancer Institute, London WC1E 6BT, UK
- ⁵ Pathology Department at the Royal Infirmary of Edinburgh, Edinburgh EH16 4SA, UK
- ⁶ Department of Pathology, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands
- ⁷ Pathology Department at Queen's Elizabeth Hospital Birmingham, Birmingham B15 2TH, UK

Introduction

Sarcomas are a rare group of heterogeneous tumors that arise within bone or soft tissues including connective tissue, muscle, nerves, blood vessels, and fat [1]. Sarcomas account for roughly 1% of all adult cancers and around 20% of pediatric solid malignancies [2, 3]. Soft tissue tumors are conventionally classified according to morphological, immunohistochemical, and in many cases molecular characteristics into more than 100 distinct subtypes, which can make differential diagnosis difficult [1]. At a molecular level they are classified into two main categories: sarcomas with specific genetic alterations and sarcomas displaying multiple complex karyotypic abnormalities [1]. Focusing on the first category, most sarcoma-specific genetic alterations involve chromosomal translocations that result in fusion genes. Fusion genes arise from the juxtaposition of two separate genes, resulting in the translation of deregulated or chimeric proteins with altered function and potential oncogenic activity [4]. It has been estimated that approximately one third of sarcomas carry a detectable driver fusion gene [5, 6].

Whilst some fusion genes in tumors of different origin appear to be passenger mutations resulting from genetic instability, in sarcomas the majority are considered strong driver alterations [4]. Certain sarcoma subtypes display recurrent, often pathognomonic translocations resulting in fusion genes, the identification of which can support or define diagnosis [7, 8]. The detection of recurrent translocations that are present in some benign soft tissue and bone tumors is also of diagnostic utility [1]. Copy number changes in particular genes can also support diagnosis. For example, amplification of genes within 12q13-15 including *MDM2* and *CDK4* genes in adipocytic neoplasms has been shown to be characteristic of well differentiated and dedifferentiated liposarcomas [9]. Detection of *MDM2* amplification using FISH is therefore an extremely useful ancillary diagnostic tool [10, 11].

In addition to their diagnostic potential, it has been demonstrated that detection of these chromosomal aberrations and fusion genes may be of use in the treatment of sarcoma patients, highlighting the importance of accurate diagnosis [12–18]. This is emphasized by the 2016 Clinical Effect of Molecular Methods in Sarcoma Diagnosis (GENSARC) study, in which 53 out of 384 (13.8%) sarcoma diagnoses were amended following molecular findings. These new findings not only modified patient subtype diagnosis but also had implications for patient management as well as prognosis assessment [19].

Due to the vast number of sarcoma histological subclassifications and overlapping histological and immunohistochemical features, conventional hematoxylin and eosin staining and immunohistochemistry (IHC) are used to support diagnosis. In addition, molecular analysis using reverse transcription polymerase chain reaction (RT-PCR), fluorescence in situ hybridization (FISH) and more recently next generation sequencing (NGS) are extremely valuable diagnostic tools [20, 21]. Whilst RT-PCR and FISH are considered the gold standard tools for fusion gene detection in sarcomas in the clinical setting, there are certain disadvantages to these tests. Required prior knowledge of alterations, the lability of RNA and scalability constraints can all pose challenges using these methods.

NGS however is a high-throughput, cost-effective tool allowing for multiple chromosomal regions to be sequenced in parallel to interrogate a wide range of genetic variation across the genome [22, 23]. Targeted NGS, in particular, allows for a set of genes or regions to be enriched and sequenced in a single assay, reducing cost, turnaround time, and data analysis burden [24]. However, there are currently no commercially available targeted NGS panels for fusion gene detection in sarcoma using DNA, only NGS tests focused on RNA sequencing and fusion gene analysis, most of which do not support copy number variation (CNV) or mutation detection [25, 26].

For this study, we designed a custom capture-based NGS panel for use with genomic DNA extracted from soft tissue tumor and bone tumor formalin-fixed paraffin embedded (FFPE) tissue. Our aims were to assess the feasibility of this panel and determine whether this assay can detect clinically-relevant fusion genes and copy number variations in diagnostic FFPE tissue samples that have previously been characterized by FISH or RT-PCR in clinical laboratories.

Methods

Ethics approval and consent to participate

Local approval was obtained for the molecular analysis of all clinical material in this study according to standard clinical practice. Ethics approval was obtained from the Northern Ireland Biobank (NIB) (NIB17-0231). The RNOH Biobank was approved by the National Research Ethics Committee of the Health Research Application System (IRAS) project identifier: 18309. This specific project was approved by the National Research Ethics Committee approved UCL/UCLH Biobank Ethics Committee (specific project reference no. EC17.14).

Tumor specimen collection

A total of 113 resection ($n = 83$) and core biopsies ($n = 30$) were selected from FFPE preserved sarcoma cases dating from 2013–2018 from the Belfast Health and Social Care Trust (Belfast, United Kingdom), Queen Elizabeth Hospital (Birmingham, United Kingdom), Radboud University Medical Center (Nijmegen, Netherlands), Royal Infirmary of Edinburgh (Edinburgh, Scotland) and Royal National Orthopedic Hospital (London, United Kingdom). The selection criteria required all samples to have previously tested positive by FISH/RT-PCR for either a fusion gene or *MDM2* amplification during routine diagnostics at the respective collaborating institution, as part of the standard of care for those patients. The 113 samples included 22 tumor subtypes harboring either *MDM2* gene amplification or one of 25 distinct chromosomal translocations including *ALK-EML4*, *ASPSCR1-TFE3*, *BCOR-CCNB3*, *C11orf95-MKL2*, *COL1A1-PDGFB*, *COL1A1-USP6*, *ETV6-NTRK3*, *EWSR1-ATF1*, *EWSR1-ERG*, *EWSR1-FLI1*, *EWSR1-NR4A3*, *EWSR1-WT1*, *FUS-CREB3L2*, *FUS-DDIT3*, *FUS-ERG*, *HEY1-NCOA2*, *NAB2-STAT6*, *PAX3-FOXO1*, *RANBP2-ALK*, *SS18-SSX1*, *SS18-SSX2*, *TAF15-NR4A3*, *USP6-FGFR1*, *USP6-MYH9*, and *WWTR1-CAMTA1*.

DNA Isolation from FFPE

Tumor content was determined by a pathologist by reviewing the hematoxylin and eosin (H&E) stained tissue sections and estimating the percentage of malignant nuclei out of the total nucleated cellularity. DNA was extracted from scrolls or macrodissected unstained sections of FFPE tissue (3–8 × 5 μm thick sections per extraction) using the Maxwell 16 FFPE Tissue LEV DNA Purification Kit (Promega, Wisconsin, US) according to manufacturer's instructions.

For FFPE sections on slides, prior to extraction, sections were dewaxed on the Tissue-Tek® Film® Prism (Sakura, Japan) involving two xylene 2-min incubations, one xylene 1-min incubation, a 95% ethanol 1-min incubation and a 90% ethanol 1-min incubation. Immediately after dewaxing, using an annotated H&E reference slide to highlight tumor areas, the unstained sections were macro-dissected to enrich for tumor tissue before placing in a collection tube containing 200 μL of incubation buffer. For tissue received in scrolls, no dewaxing or macro-dissection was performed and 200 μL of incubation buffer was added directly to the sample. Proteinase K treatment was performed by incubating the sample at 70 °C overnight before completing the automated DNA extraction using the Maxwell® 16 Instrument, configured for LEV operation, with final elution into 50 μL of nuclease-free water. DNA was quantified using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, MA, US) and quality assessed using either the Fragment Analyzer™ Automated CE System or 4200 TapeStation System.

Panel design

An in-depth literature review was performed to identify gene translocations involved in sarcomas. Fusion breakpoint sequences were collected from relevant case reports and precise chromosome locations were validated with Ensembl GRCh38.p12 (<https://www.ensembl.org>) and Integrative Genomics Viewer (IGV) (<http://software.broadinstitute.org/software/igv>) [27, 28]. The start and end chromosomal location of each intron for all 202 fusion variants was identified using Ensembl. For copy number variation, probes were designed for the chromosomal location of the exons of the genes of interest and surrounding regions spanning 200 kb. All chromosomal locations were checked against areas of poor mappability and CpG islands using the ENCODE Duke Mappability and Uniqueness track and CpG island track available for download from the University of California Santa Cruz (UCSC) Genome Browser. The final design was manufactured through the EZ Choice Probe design (Roche Nimblegen, Madison, WI, USA).

NGS library construction and sequencing

Where available, library preparation was performed on 200 or 400 ng of DNA. This decision was based on average fragment size of the DNA measured using either the Fragment Analyzer™ Automated CE System or 4200 TapeStation System. Either 200 ng of good quality DNA (>1000 bp average length) or 400 ng of poorer quality DNA (<1000 bp average length) was loaded into the preparation reaction.

The KAPA HyperPlus Kit (Roche Sequencing Solutions, Inc., California, USA), SeqCap EZ adapters and associated SeqCap target enrichment reagents (Roche Sequencing Solutions, Inc., California, US) were used according to the manufacturer's protocol, including dual size selection of the libraries (250–450 bp). Following enzymatic fragmentation, DNA was end-repaired, A-tailed, indexed adapters ligated and amplified using PCR (6 cycles) before dual size selection. For each run, 24 libraries were hybridized overnight (16–20 h) using 1 μg of the pooled libraries and custom designed biotinylated DNA baits complementary to the genomic regions of interest (NimbleGen SeqCap EZ library, Roche Nimblegen, Madison, WI, US). Following hybridization and stringency washes, the resulting DNA was PCR amplified (11 cycles) and purified using AMPure XP beads (Beckman Coulter, Danvers, MA, US) at a ratio of 1.8× and quantified using the KAPA Quantification q-PCR Kit (KAPA Biosystems, Wilmington, MA, US) and Qubit High Sensitivity (HS) assay kit with the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, MA, US). The quality and molarity of the PCR-amplified hybridized library was measured using the Fragment Analyzer™ Automated CE System or 4200 TapeStation System.

Sequencing was performed on a NextSeq 500 (Illumina, San Diego, CA, USA) using the NextSeq 550 System Mid-Output Kit v2 (150 cycles) with 76 bp paired-end read cycles according to the manufacturer's instructions.

Low input NGS library construction and sequencing

FFPE samples with less than 200 ng of DNA input available were also prepared for sequencing, irrespective of DNA quality, with DNA input varying between 15 and 170 ng. Samples with a DNA input between 40 and 200 ng were processed exactly as above. In samples where DNA input was lower than 40 ng, three additional PCR cycles (9 cycles total) were performed following enzymatic fragmentation, DNA end repair and A-tailing and indexed adapters ligation.

Structural variant data analysis

Base calls and quality scores were provided by the NextSeq 500 using real time analysis (RTA 2.0). FASTQ data were aligned to the hg38 build of the human reference genome

using the Isaac Aligner (iSAAC-03.16.02.20) [29]. Translocations were detected from the aligned sequence data using Manta Structural Variant Caller (v0.28.0) [30]. The analysis was performed blinded to the original results obtained by each respective Center. Any samples with undetected translocations with the analysis pipeline were manually analyzed using integrative genomics viewer (IGV) by color coding reads by insert size and grouping alignments by chromosome of mate [28, 31].

Copy number variation data analysis

The fastq sequences were aligned to human reference genome version GRCh38/hg38 using Burrow-Wheeler Aligner (BWA, v0.7.17) [32]. The aligned reads were sorted, merged, and filtered using SAMtools (v4.0.12.0) [33]. Picard tools from GATK was used to sort, mark, and remove the duplicates. Local realignment of reads around INDELs and base recalibration were performed using Genome Analysis Toolkit (GATK) [34]. DNA copy number information was first acquired using CopywriteR R package version 2.16.0 utilizing off target reads from targeted sequencing data [35]. The results from this confirmed that *MDM2/CDK4* genes were not amplified in fusion positive sarcoma cases (Supplementary File 1: Fig. S1) CNVpanelizer R package Version 1.16.0, which is based on subsampling strategy to predict the CNVs, was then used to detect *MDM2* and *CDK4* amplification in the respective cases. CNVpanelizer compares *MDM2/CDK4* amplified samples with the non-amplified pool of fusion positive samples.

Results

Panel design

The overall objective was to develop a sarcoma-specific targeted NGS gene panel to detect both fusion genes and common copy number variants in FFPE-extracted DNA. This gene panel targets 202 unique fusion junctions involving 87 sarcoma-related fusion genes and 7 genes with known copy number variation in sarcoma (Supplementary File 2: Table S1). Fourteen additional fusion junctions were identified during the literature search but were excluded from final panel design due to their large intron size, rarity and/or diagnostic value (Supplementary File 2: Table S1).

Sequencing quality

A validation set of 100 sarcoma samples with >200 ng of DNA characterized either by FISH ($n = 61$) or RT-PCR ($n = 39$), consisting of 22 soft tissue and bone tumor

subtypes covering 24 distinct chromosomal translocations and *MDM2/CDK4* gene amplifications were used. The average tumor content was 70% (range 30–90%). An initial repeatability assessment was performed by repeating the first sequencing analysis (with 24 clinical samples) from library preparation on two different hybridizations and two different sequencing runs on different dates. Both sequencing runs were able to identify the same fusion genes following analysis. The targeted NGS assay yielded an average of 7,595,994 reads per sample (range 2,590,654–13,667,938). The mean depth of unique coverage was $602.6 \times \pm 161.6$ SD.

Detection of structural variants

Out of the 100 patient samples, 89 harbored a fusion gene and 87/89 (97.8%) fusion genes were detected by the NGS assay (Table 1). All fusions had previously been confirmed during diagnostic testing using FISH, RT-PCR, or both. In the two discrepant cases, the NGS assay failed to detect *WWTR1-CAMTA1* and *CIC-DUX4* translocations which were detectable using RT-PCR and FISH, respectively. The analytical sensitivity of the panel for detection of fusion genes in this validation set was 97.8%, and the analytical specificity was 100%, with no false positives observed across any of the fusion genes detectable by the NGS panel in any of the cases.

Eleven of the 100 patient samples were characterized by *MDM2* amplification confirmed by FISH. The NGS assay detected both *MDM2* and *CDK4* amplification in 10/11 (91%) of these samples, failing to detect *MDM2* or *CDK4* amplification in one sample. As only *MDM2* amplification is tested diagnostically and therefore comparable to our test, the analytical sensitivity of the NGS assay in detecting *MDM2* amplification in this small cohort of samples is 91% and analytical specificity 100%, with no false positives being detected (Table 1).

To determine minimum DNA input required for the assay, DNA extracted from a further 13 samples from which 200 ng of DNA could not be obtained was used. The DNA quantity available from these samples ranged from 15 to 171 ng with an average fragment size of 1379 bp. For this cohort of patients, the targeted NGS assay yielded an average of 5,049,051 reads per sample (range 2,227,787–7,074,052) and the mean unique target coverage depth was $233 \times \pm 120.8$ SD. Out of the 13 samples with lower DNA input, the NGS assay detected all fusion genes previously identified using FISH or RT-PCR (Table 2).

Out of the 113 samples combined, a total of 25 fusion genes out of a possible 87 fusions targeted by the panel were identified (Fig. 1). A full sample list including translocations identified and their location are also provided (Supplementary File 3: Table S2).

Table 1 Table of sarcoma subtypes and abnormalities detected.

Tumor subtype	Abnormalities detected	Number of cases with known abnormalities	Number of cases with abnormalities detected
Synovial sarcoma	<i>SS18-SSX1, SS18-SSX2</i>	25	25
Ewing's sarcoma	<i>EWSR1-FLI1, EWSR1-ERG, FUS-ERG</i>	12	12
Myxoid liposarcoma	<i>FUS-DDIT3</i>	10	10
Well/Dedifferentiated liposarcoma ^a	<i>MDM2/CDK4</i> amplification	10	9
Alveolar rhabdomyosarcoma	<i>PAX3-FOXO1</i>	6	6
Extraskeletal myxoid chondrosarcoma	<i>EWSR1-NR4A3, TAF15-NR4A3</i>	5	5
Clear cell sarcoma	<i>EWSR1-ATF1</i>	4	4
Dermatofibrosarcoma protuberans	<i>COL1A1-PDGFB</i>	4	4
Low-grade fibromyxoid sarcoma	<i>FUS-CREB2L2</i>	3	3
Solitary fibrous tumor	<i>NAB2-STAT6</i>	3	3
Congenital fibrosarcoma	<i>ETV6-NTRK3</i>	2	2
Desmoplastic small round cell tumor	<i>EWSR1-WT1</i>	2	2
Epithelioid haemangioendothelioma	<i>WWTR1-CAMTA1</i>	2	1
Inflammatory myofibroblastic tumor	<i>ALK-EML4, RANBP2-ALK</i>	2	2
Mesenchymal chondrosarcoma	<i>HEY1-NCOA2</i>	2	2
Nodular fasciitis	<i>COL1A1-USP6, USP6-MYH9</i>	2	2
Alveolar soft part cell sarcoma	<i>ASPSCR1-TFE3</i>	1	1
Aneurysmal bone cyst	<i>USP6-FGFR1</i>	1	1
BCOR-CCNB3 sarcoma	<i>BCOR-CCNB3</i>	1	1
Chondroid lipoma	<i>C11orf95-MKL2</i>	1	1
Intimal sarcoma ^a	<i>MDM2</i> amplification	1	1
<i>CIC</i> -rearranged round cell sarcoma	Not applicable	1	0
Total		100	97

^aCharacterized by *MDM2/CDK4* gene amplification.

Total number of sarcoma subtypes detected using the NGS panel assay including the abnormalities detected (structural variants and copy number variations), total number of cases and number of abnormalities detected.

Test sensitivity and specificity

When combining both fusion gene detection and *MDM2/CDK4* amplification in all 113 samples, the assay sensitivity was 97% (92.5–99.1% 95% CI) and the observed specificity remained at 100%. The failure rate of the panel, even when using low DNA inputs, such as those obtained from core biopsies (10 of 13 samples), remained at 0%. A sample would be considered failed if the minimum read depth was less than 100× and no alterations were detected.

Analysis performance

Overall, 80% ($n = 82$) of translocations were called by structural variant caller Manta during automated analysis, with manual IGV analysis only required for the remaining 20% ($n = 20$) of fusion positive cases. *MDM2* and *CDK4* amplification was detected in the respective ten cases using R package CNVPanelizer, with a \log_2 ratio of 1.5 or above considered as gene amplification [36]. Figure 2 shows the *MDM2* and *CDK4* amplified samples including

centromeric and telomeric control regions. Significant difference was observed between *MDM2* with its centromeric (p -value = 0.002) and telomeric (p -value = 0.008) region, demonstrating gene amplification instead of polysomy.

Discussion

The aim of this study was to design and validate a DNA-based custom targeted sequencing panel to detect known and clinically relevant fusion genes and copy number variants in FFPE material from soft-tissue and bone tumors and to assess its potential clinical applicability. In total, 25 different fusions were identified in 102 cases using the targeted gene panel in a cohort of 113 patients, representing only a third of the fusions that the panel can potentially detect. *MDM2/CDK4* amplification was detected in 10 out of the 11 cases with *MDM2* amplification detected by FISH and in none of the fusion gene positive cases. An expanded patient cohort would be required to evaluate the additional fusion genes and copy number alterations.

Table 2 Low input DNA samples.

Sample number	DNA input (ng)	Fusion detected	Unique reads	Number of alternative reads ^a	Mean target coverage depth
1	171	<i>FUS-CREB2L2</i>	7,456,276	48	452
2	158	<i>EWSR1-FLII</i>	6,462,694	84	357
3	150	<i>SS18-SSX1</i>	6,098,438	62	305
4	125	<i>FUS-DDIT3</i>	6,176,276	173	294
5	99	<i>FUS-DDIT3</i>	3,803,180	14	156
6	98	<i>FUS-DDIT3</i>	4,573,458	26	203
7	96	<i>EWSR1-NR4A3</i>	5,421,508	84	271
8	94	<i>EWSR1-FLII</i>	8,109,076	70	436
9	83	<i>FUS-DDIT3</i>	4,839,058	115	216
10	39	<i>FUS-ERG</i>	4,331,932	19	150
11	26	<i>FUS-DDIT3</i>	3,133,298	45	78
12	16	<i>EWSR1-FLII</i>	3,009,217	15	72
13	15	<i>EWSR1-NR4A3</i>	2,721,404	10	64

Table of 13 low input DNA samples showing DNA input, fusion genes detected by NGS, unique reads, number of alternative reads, and the mean target coverage depth.

^aAlternative reads: number of unique reads that either spanned the breakpoint or contained the breakpoint for the translocation.

FISH can detect translocations by probing for either breakage of one gene using break-apart probes or the joining of two specific gene partners using fusion probes, however multiple tests per sample are generally required as it lacks multiplexing capability, resulting in increased tissue requirements, labor and cost [37]. RT-PCR on the other hand can be more sensitive than FISH in specimens with lower tumor content [38–40]. RT-PCR however may not detect rare fusion genes or those with unusual breakpoints and cannot easily detect copy number variation, for example *MDM2* amplification, making FISH a continual requirement for comprehensive sarcoma testing [41].

In a previous study of 772 *EWSR1*-rearranged soft tissue sarcoma FFPE samples, FISH failed in 2.5% of cases whereby RT-PCR failed in 18% [41]. In another study involving 50 synovial sarcoma cases, RT-PCR was found to be more sensitive than FISH, yet concordant results were only reported in 76% of cases [42]. Failure rates of up to 25% have also been reported in fusion positive tumor samples using RNA NGS sequencing, mainly attributed to small sample size, tissue heterogeneity and tissue fixation processes [25, 43]. Conclusive molecular diagnostic test results therefore cannot be achieved in up to a quarter of fusion-positive sarcoma patients with current methods [41, 42, 44].

Due to the lower input requirements, sequencing costs and data storage, panel-based targeted sequencing is currently more suited to the clinical setting than more comprehensive approaches, such as whole genome sequencing

(WGS), and there are many examples of the development and implementation of targeted sequencing panels in oncology [45–48]. The sarcoma-specific NGS panel described here can detect both fusion gene partners without any prior knowledge of the diagnosis, irrespective of whether both, or only one gene partner are included in the panel design. This has advantages over both FISH and RT-PCR: In Ewing's sarcoma, for example, *EWSR1* is most frequently fused with *FLII* or *ERG* accounting for over 90% of cases [49]. RT-PCR primer design generally focuses on the more common fusion type and false negatives can therefore be reported in cases with rarer fusion transcripts (10% of all Ewing's sarcomas). Whilst *EWSR1* FISH break-apart probes can combat this problem for the majority of cases, a small subset of Ewing's sarcomas harbor *FUS* gene rearrangements, substituting for *EWSR1*, which would consequently result as a false negative using this technique [50]. In addition, the fusion gene partner remains unknown, which can be crucial for subtype diagnosis, since *EWSR1* rearrangements are also present in multiple sarcoma subtypes other than Ewing's sarcoma. Similarly, being unable to identify different partners that have an effect on prognosis, such as *SSX1* and *SSX2* in synovial sarcomas or *PAX3* and *PAX7* in alveolar rhabdomyosarcomas, can limit the usefulness of FISH [51, 52]. FISH analysis may also be limited by nuclear truncation artefacts that can lead to false positive or negative results. The presence of a split signal caused by a non-functional rearrangement may also lead to false positive results [53]. NGS on the other hand, provides

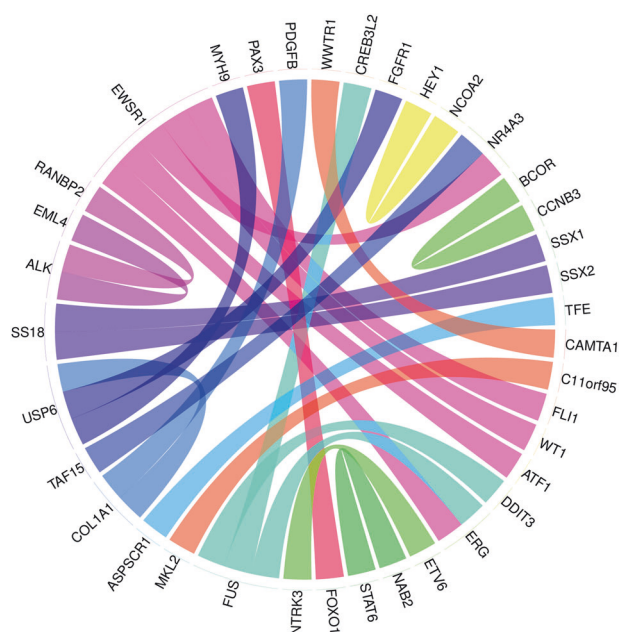


Fig. 1 Chord diagram displaying the 25 fusion gene variants identified in 113 patients.

the exact breakpoint and chromosomal location of both genes involved in the translocation limiting the possibility of a false positive result.

The targeted gene panel can simultaneously screen multiple genes in multiple samples in a single test, alleviating the need for multiple RT-PCR or FISH tests per sample thereby reducing the price, turnaround time as well as the amount of tissue required to achieve a definitive sarcoma subtype diagnosis. Sequencing with the targeted gene panel can be completed from library preparation to data analysis within 5 days for manual preparation or 3 days using automation. Analysis time and cost is also significantly reduced based on the automatic detection of approximately 80% of translocations (based on this cohort). A key advantage of this NGS panel is that it can be performed alongside immunohistochemical techniques instead of at the end of the histological diagnosis, as it does not require prior knowledge of the alterations to look for, potentially reducing the time to arrive at a comprehensive diagnosis. Nonetheless, FISH or RT-PCR may be the preferred method for confirming suspected translocations in urgent cases such as those requiring rapid, integral chemotherapy treatment or in cases with a strongly favored diagnosis.

Our results show that gene fusions can be accurately identified with as little as 15 ng input DNA from FFPE, making it suitable for both resection and core biopsy specimens. Assay sensitivity for FISH, RT-PCR, and particularly for WGS, decreases when using FFPE tissue in comparison to fresh tissue or in samples with significant heterogeneity, whereas the NGS targeted panel maintains its

high sensitivity with lower FFPE DNA input [54–56]. The NGS panel showed an overall sensitivity of 97% with no false positives and 0% failure rate, indicating that it can be implemented in a clinical setting without compromising on performance against FISH and RT-PCR. Furthermore, since the panel is DNA-based it can be applied to circulating cell-free DNA, opening opportunities to non-invasive sarcoma diagnosis and monitoring, as demonstrated recently in a small group of *EWSR1*-rearranged sarcomas [57].

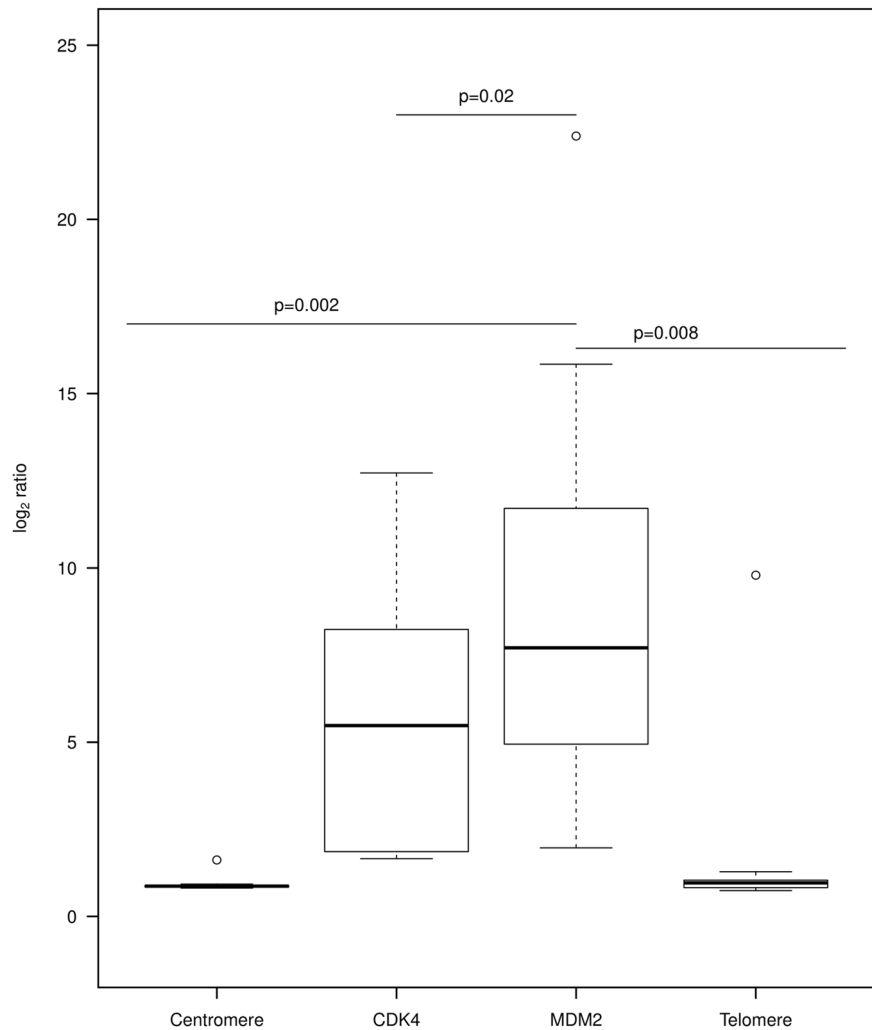
An obvious limitation of targeted NGS is the ability to detect fusion genes only in targeted regions included in the design. This panel targets intronic regions which have been known to result in low sequencing coverage in highly repetitive GC-rich regions [58]. Whilst low coverage is observed in some target regions, the percentage coverage in this panel is high on average (96%, range 35.6–100%). At least one gene involved in a fusion gene must be targeted for translocation detection. The unidentified *WWTR1*-*CAMTA1* fusion, confirmed by RT-PCR, can be explained by the omission of intron 7–8 of the *CAMTA1* gene and intron 4–5 of the *WWTR1* gene from panel design, due to the rarity of epithelioid hemangioendothelioma (prevalence of less than 1 in a million), and more specifically the rarity of the particular breakpoints in question [59]. This however could be easily amended by expanding the size of the gene panel and including these intronic regions in an updated version.

The second discrepant case was a round cell sarcoma with a *CIC* rearrangement confirmed by FISH that was not detected by the NGS assay. This could be due to the rearrangement having a breakpoint not previously reported in the literature and thus not targeted by the panel but could also be due to poor coverage in the *CIC* and *DUX4* regions on the panel, containing highly repetitive regions with poor mappability and, as such, shows lower coverage than most other regions in the NGS panel. More cases with this fusion are required to fully assess the limitations of the NGS panel.

Although this NGS panel is extremely comprehensive, it can be expanded to incorporate new fusion sarcomas and new fusion variants in known tumors. The approach to bone and soft tissue tumor diagnosis is well established and incorporates the clinical scenario, histopathological findings including IHC and the radiological impression which is particularly pertinent for bone tumors and molecular analysis. As a case may be encountered in which a diagnosis is strongly favored, it is important to consider all of these diagnostic parameters before offering a definitive diagnosis. In select cases, recourse may be taken to FISH or RT-PCR, however, given the breadth of this NGS panel and the option for its further expansion, this may be an uncommon scenario.

The *MDM2* gene was considered amplified if the \log_2 -ratio was greater than 1.5 as our cohort contained an intimal

Fig. 2 Boxplot showing \log_2 ratio for *CDK4*, *MDM2*, centromere and telomere control regions for the 10 *MDM2*-amplified well/dedifferentiated liposarcoma and intimal sarcoma cases. *p*-value was calculated using a student's *t*-test.



sarcoma with low level *MDM2* amplification. The average \log_2 ratio for well/dedifferentiated liposarcomas was 5.6. The NGS assay failed to detect *MDM2* amplification above the 1.5 \log_2 ratio threshold in a case previously confirmed using a *MDM2-CEP12* FISH probe. This particular case showed *MDM2* amplification in only a fraction of nuclei as well as polysomy (Supplementary File 4: Fig. S2), which may have affected the detection by NGS.

Based on sensitivity and specificity, lower input FFPE material required and cost effectiveness as well as its ability to detect both translocations and gene amplifications all in the one test, the targeted NGS panel is therefore a useful ancillary tool to be used alongside immunohistochemistry to improve sarcoma diagnosis.

Acknowledgements We would like to thank all the patients and their families, the Northern Ireland Biobank and the RNOH biobank, Dr. Louise Harewood, Elisa Izquierdo and Neil Mc Cafferty for their advice. This study was supported by Sarcoma UK.

Compliance with ethical standards

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens F, editors. WHO classification of tumours of soft tissue and bone. 4th ed. Lyon: IARC; 2013.
2. Ducimetière F, Lurkin A, Ranchère-Vince D, Decouvelaere A, Péoc'h M, Istier L, et al. Incidence of sarcoma histotypes and molecular subtypes in a prospective epidemiological study with central pathology review and molecular testing. Vol 6. San Francisco: Public Library of Science; 2011.
3. Bleloch JS, Ballim RD, Kimani S, Parkes J, Panieri E, Willmer T, et al. Managing sarcoma: Where have we come from and where are we going?. London: SAGE Publications; 2017. p. 637–659.

4. Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. *Nat Rev Cancer*. 2015;15:371–81.
5. Barr FG, Zhang PJ. The impact of genetics on sarcoma diagnosis: an evolving science. *Clin Cancer Res*. 2006;12:5256–7.
6. Mercado GE & Barr FG. Chromosomal translocations in sarcomas: new perspectives. 2006. <http://sarcomahelp.org/articles/chromosomal-translocations.html>.
7. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer*. 2007;7:233–45.
8. Parker BC, Zhang W. Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment. *Chin J cancer*. 2013;32:594–603.
9. Thway K, Wang J, Swansbury J, Min T, Fisher C. Fluorescence in situ hybridization for MDM2 amplification as a routine ancillary diagnostic tool for suspected well-differentiated and dedifferentiated liposarcomas: experience at a tertiary center. *Sarcoma*. 2015;2015:812089.
10. Weaver J, Downs-Kelly E, Goldblum JR, Turner S, Kulkarni S, Tubbs RR, et al. Fluorescence in situ hybridization for MDM2 gene amplification as a diagnostic tool in lipomatous neoplasms. *Mod Pathol*. 2008;21:943–9.
11. Ware PL, Snow AN, Gvalani M, Pettenati MJ, Qasem SA. MDM2 copy numbers in well-differentiated and dedifferentiated liposarcoma characterizing progression to high-grade tumors. *Am J Clin Pathol*. 2014;141:334–41.
12. Martín DH, Boro A, Schäfer BW. Cell-based small-molecule compound screen identifies fenretinide as potential therapeutic for translocation-positive rhabdomyosarcoma. *Public Libr Sci*. 2013;8:e55072.
13. Gupta S, Gouw L, Wright J, Chawla S, Pitt D, Wade M, et al. Phase II study of amrubicin (SM-5887), a synthetic 9-aminoanthracycline, as first line treatment in patients with metastatic or unresectable soft tissue sarcoma: durable response in myxoid liposarcoma with TLS-CHOP translocation. *Invest New Drugs*. 2016;34:243–52.
14. Brenner JC, Feng FY, Han S, Patel S, Goyal SV, Bou-Maroun LM, et al. PARP-1 inhibition as a targeted strategy to treat ewing's sarcoma. *Cancer Res*. 2012; 72:1608–13.
15. Xiao X, Garbutt CC, Hornicek F, Guo Z, Duan Z. Advances in chromosomal translocations and fusion genes in sarcomas and potential therapeutic applications. *Cancer Treat Rev*. 2018;63:61–70.
16. Pappo AS, Vassal G, Crowley JJ, Bolejack V, Hogendoorn PCW, Chugh R, et al. A phase 2 trial of R1507, a monoclonal antibody to the insulin-like growth factor-1 receptor (IGF-1R), in patients with recurrent or refractory rhabdomyosarcoma, osteosarcoma, synovial sarcoma, and other soft tissue sarcomas: results of a sarcoma alliance for research through collaboration study. *Cancer*. 2014;120:2448–56.
17. Tatai T, Gomi D, Fukushima T, Kobayashi T, Sekiguchi N, Sakamoto A, et al. Effectiveness of imatinib mesylate treatment in a patient with dermatofibrosarcoma protuberans with pulmonary and pancreatic metastases. *Intern Med*. 2016; 55:2507–11.
18. McGovern Y, Zhou CD, Jones RL. Systemic therapy in metastatic or unresectable well-differentiated/dedifferentiated liposarcoma. Vol 7. Lausanne: Frontiers Research Foundation; 2017. p. 292.
19. Italiano A, Di Mauro I, Rapp J, Pierron G, Auger N, Alberti L, et al. Clinical effect of molecular methods in sarcoma diagnosis (GENSARC): a prospective, multicentre, observational study. *Lancet Oncol*. 2016;17:532–8.
20. Smolle MA, Andreou D, Tunn P, Szkandera J, Liegl-Atzwanger B, Leithner A. Diagnosis and treatment of soft-tissue sarcomas of the extremities and trunk. *EFORT Open Rev*. 2017;2:421–31.
21. Iwasaki H, Nabeshima K, Nishio J, Jimi S, Aoki M, Koga K, et al. Pathology of soft-tissue tumors: daily diagnosis, molecular cytogenetics and experimental approach. *Pathol Int*. 2009;59:501–21.
22. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract*. 2013;98:236–8.
23. Dillio AA, Farhan SMK, Ghani M, Sato C, Liang E, Zhang M, et al. Targeted next-generation sequencing and bioinformatics pipeline to evaluate genetic determinants of constitutional disease. *J Vis Exp* 2018:e57266.
24. Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. *Clin Biochem Rev*. 2011;32:177.
25. Lam SW, Cleton-Jansen A, Cleven AHG, Ruano D, van Wezel T, Szuhai K, et al. Molecular analysis of gene fusions in bone and soft tissue tumors by anchored multiplex PCR-based targeted next-generation sequencing. *J Mol Diagn*. 2018;20:653–63.
26. NeoGenomics. NGS comprehensive sarcoma fusion profile. 2018; <https://neogenomics.com/test-menu/ngs-comprehensive-sarcoma-fusion-profile>. Accessed 20 August 2019.
27. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. *Nucleic Acids Res*. 2018;46:D75–D761.
28. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14:178–92.
29. Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: ultra-fast whole-genome secondary analysis on illumina sequencing platforms. *Bioinformatics*. 2013;29:2041–3.
30. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32:1220–2.
31. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. 2011; <https://www.nature.com/articles/nbt.1754>. Accessed 23 Jan 2018.
32. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25:1754–60.
33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
34. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
35. Kuilman T. CopywriteR: Copy number information from targeted sequencing using off-target reads. R package version 2.16.0. 2019; <https://github.com/PeeperLab/CopywriteR>. Accessed 16 August 2019.
36. Oliviera C, Wolf T. CNVPanelizer: Reliable CNV detection in targeted sequencing applications. R Package version 1.4.0. 2016.
37. Chang KTE, Goytain A, Tucker T, Karsan A, Lee C, Nielsen TO, et al. Development and evaluation of a pan-sarcoma fusion gene detection assay using the NanoString nCounter platform. *J Mol Diagn*. 2018;20:63–77.
38. Gruber K, Horn H, Kalla J, Fritz P, Rosenwald A, Kohlhäuff M, et al. Detection of rearrangements and transcriptional up-regulation of ALK in FFPE lung cancer specimens using a novel, sensitive, quantitative reverse transcription polymerase chain reaction assay. *J Thorac Oncol*. 2014;9:307–15.
39. Liu C, Pepper K, Hendrickson H, Cagle PT, Portier BP. Clinical validation of a novel commercial reverse transcription-quantitative polymerase chain reaction screening assay for detection of ALK translocations and amplifications in non-small cell lung carcinomas. *Arch Pathol Lab Med*. 2016;140:690–3.
40. Weber B, Liu M, Sobkin P, Morris SW, Hout D, van der Westhuizen N, et al. Successful treatment of hepatic oligometastases

- with stereotactic ablative radiotherapy and radiofrequency ablation in an anaplastic lymphoma kinase fusion-positive lung cancer patient. *J Med Radiat Sci.* 2016;63:67–70.
41. Noujaim J, Jones RL, Swansbury J, Gonzalez D, Benson C, Judson I, et al. The spectrum of EWSR1-rearranged neoplasms at a tertiary sarcoma centre; assessing 772 tumour specimens and the value of current ancillary molecular diagnostic modalities. *Br J Cancer.* 2017;116:669–78.
 42. Ten Heuvel SE, Hoekstra HJ, Suurmeijer AJH. Diagnostic accuracy of FISH and RT-PCR in 50 routinely processed synovial sarcomas. *Appl Immunohistochem Mol Morphol.* 2008;16:246–50.
 43. Williams HL, Walsh K, Diamond A, Oniscu A, Deans ZC. Validation of the oncoPrint™ focus panel for next-generation sequencing of clinical tumour samples. *Virchows Arch.* 2018;473:489–503.
 44. Lewis TB, Coffin CM, Bernard PS. Differentiating ewing's sarcoma from other round blue cell tumors using a RT-PCR translocation panel on formalin-fixed paraffin-embedded tissues. *Mod Pathol.* 2007;20:397–404.
 45. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, et al. Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagn.* 2015;17:251–64.
 46. Salto-Tellez M, Gonzalez de Castro D. Next-generation sequencing: A change of paradigm in molecular diagnostic validation. *J Pathol.* 2014;234:5–10.
 47. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Vol 31. Nature Publishing Group, a division of Macmillan Publishers Limited. 2013; p. 1023–31.
 48. Cottrell CE, Al-Kateb H, Bredemeyer AJ, Duncavage EJ, Spencer DH, Abel HJ, et al. Validation of a next-generation sequencing assay for clinical molecular oncology. *J Mol Diagn.* 2014;16:89–105.
 49. Gamberi G, Cocchi S, Benini S, Magagnoli G, Morandi L, Kreshak J, et al. Molecular diagnosis in ewing family tumors: The rizzoli Experience—222 consecutive cases in four years. Vol 13. American Society for Investigative Pathology; 2011. p. 313–24.
 50. Chen S, Deniz K, Sung Y, Zhang L, Dry S, Antonescu CR. Ewing sarcoma with ERG gene rearrangements: a molecular study focusing on the prevalence of FUS-ERG and common pitfalls in detecting EWSR1-ERG fusions by FISH. *Genes Chromosomes Cancer.* 2016;55:340–9.
 51. Ladanyi M, Antonescu CR, Leung DH, Woodruff JM, Kawai A, Healey JH, et al. Impact of SYT-SSX fusion type on the clinical behavior of synovial sarcoma: a multi-institutional retrospective study of 243 patients. *Cancer Res.* 2002;62:135–40.
 52. Sorensen PH, Lynch JC, Qualman SJ, Tirabosco R, Lim JF, Maurer HM, et al. PAX3-FKHR and PAX7-FKHR gene fusions are prognostic indicators in alveolar rhabdomyosarcoma: a report from the children's oncology group. *J Clin Oncol.* 2002;20:2672–9.
 53. Tafe LJ, Arcila ME. Genomic medicine: a practical guide. New York: Springer; 2019.
 54. Chen Q, Vansant G, Oades K, Pickering M, Wei JS, Song YK, et al. Diagnosis of the small round blue cell tumors using multiplex polymerase chain reaction. *J Mol Diagn.* 2007;9:80–8.
 55. Barr FG. Fusion genes in solid tumors: the possibilities and the pitfalls. *Expert Rev Mol Diagn.* 2016;16:921–3.
 56. Hastings RJ, Bown N, Tibiletti MG, Debiec-Rychter M, Vanni R, Espinet B, et al. Guidelines for cytogenetic investigations in tumours. *Eur J Hum Genet.* 2016;24:6–13.
 57. Shukla NN, Patel JA, Magnan H, Zehir A, You D, Tang J, et al. Plasma DNA-based molecular diagnosis, prognostication, and monitoring of patients with EWSR1 fusion-positive sarcomas. *JCO Precis Oncol* 2017;1:1–11.
 58. Abel HJ, Al-Kateb H, Cottrell CE, Bredemeyer AJ, Pritchard CC, Grossmann AH, et al. Detection of gene rearrangements in targeted clinical next-generation sequencing. Vol 16. American Society for Investigative Pathology; 2014. p. 405–17.
 59. Sardaro A, Bardoscia L, Petruzzelli MF, Portaluri M. Epithelioid hemangioendothelioma: an overview and update on a rare vascular tumor. Pavia: PAGEPress Publications; 2014. p. 259.